Topological Classification of Insulators without and with Time-Reversal Symmetry

Jeroen Grimbergen

Supervisors:
Prof. dr. G.R. Cavalcanti
Prof. dr. ir. H.T.C. Stoof

A thesis presented for the degree of
Master of Science

Theoretical Physics and Mathematical Sciences
Utrecht University
July 2023

Topological Classification of Insulators without and with Time-Reversal Symmetry

Jeroen Grimbergen

July 10, 2023

Abstract

Crystalline band insulators are described in the low temperature regime by a tight-binding model. A mathematically precise notion of topological equivalence of tight-binding models for insulators is formulated in terms of homotopy classes of maps. For insulators without any symmetry requirements, we review the equivalence of the homotopy classification of insulators in d spatial dimensions and the classification of complex vector bundles over a d-dimensional sphere or torus. A proof is presented of the fact that, in two spatial dimensions, the first Chern number is a $\mathbb{Z}$-valued topological invariant that fully classifies the two-dimensional insulators. We also discuss physical interpretations of the first Chern number in the context of the Hall conductivity and the bulk-boundary correspondence. For insulators with time-reversal symmetry, there is no $\mathbb{Z}$ invariant, but a $\mathbb{Z}_2$ invariant as shown by Fu, Kane and Mele. The classification of time-reversal symmetric insulators has been shown to be equivalent to the classification of "Quaternionic" vector bundles. Building on work by DeNittis and Gomi, a proof that the Fu-Kane-Mele invariant fully classifies the "Quaternionic" vector bundles associated to the continuum models of time-reversal symmetric insulators is provided in the language of equivariant homotopy and equivariant Čech cohomology.

Contents

1 Introduction **4**

I Topological Insulators **6**

2 Insulators and topological equivalence **7**
2.1 Electronic band theory of crystalline materials 7
2.2 Conductors and insulators 11
2.3 The tight-binding model and the continuum limit 12
2.4 Homotopy classification of insulators 16

3 Vector bundles and Chern classes **18**
3.1 Vector bundles 18
3.2 DeRham cohomology 20
3.3 Connections and curvature 22
3.4 Chern classes 24

4 Berry curvature, valence bundle and Hall conductivity **26**
4.1 The adiabatic connection 26
4.2 The valence bundle 30
4.3 Hall conductivity of a two-dimensional insulator 31

5 Topological classification of insulators **35**
5.1 Homotopy classification of complex vector bundles 36
5.2 Integrality of the Chern classes 37
5.3 Classification of insulators in spatial dimensions $d = 1, 2, 3$ 39
5.4 The two-band model 41

6 Bulk-boundary correspondence **44**
6.1 Effective Chern-Simons theory 44
6.2 Restoring current conservation 46
6.3 Chiral fermions in Dirac model with domain wall 47

II Topological Insulators with time-reversal symmetry **52**

7 Topological insulators with time-reversal symmetry **53**
7.1 time-reversal symmetry 53
7.2 The first Chern number is zero 55
7.3 Fu-Kane-Mele invariant 56
7.4 Example: A low-energy model for Bi_2Se_3 61

8 Classification of "Quaternionic" vector bundles **64**
8.1 "Quaternionic" vector bundles 65
8.2 Homotopy formulation of the FKMM invariant 68
8.3 The FKMM invariant for time reversal spheres 70
8.4 The FKMM invariant for time reversal tori 73

9 Classification of "Real" line bundles **75**
9.1 Čech cohomology 76
9.2 Sheaf cohomology 80
9.3 Equivariant Čech and sheaf cohomology 83
9.4 Classification of "Real" line bundles 89
9.5 Cohomology formulation of the FKMM invariant 91

10 Outlook **94**

A Gauge invariance of the polarization tensor **95**

B Equivalence of Čech and sheaf cohomology **97**
B.1 Ordinary setting 98
B.2 Equivariant setting 101

C Equivariant sheaf cohomology of $\widetilde{S}^d$ **102**
C.1 Proof of $H^2(\widetilde{S}^2;\mathbb{Z}_2,\widetilde{\mathbb{Z}})=0$ 102
C.2 Proof of $H^2(\widetilde{S}^d;\mathbb{Z}_2,\widetilde{\mathbb{Z}})=0$ for $d\geq 2$ 104
C.3 Proof of $H^2(\widetilde{S}^d,(\widetilde{S}^d)^\tau;\mathbb{Z}_2,\widetilde{\mathbb{Z}})\cong\mathbb{Z}_2$ for $d\geq 2$ 106

1 Introduction

Heuristically, an insulator is a material that does not conduct electricity. We restrict ourselves to the study of crystalline materials whose constituent atoms are arranged in a periodic fashion in space. To describe the electrons inside the material we assume the non-interacting electron approximation is valid. This means that all interactions of the electrons with each other will be ignored. The electronic properties of such materials can be understood in this approximation in terms of so-called *band theory* [1, 2], which provides a systematic way of representing the single-electron states. The low-energy properties of a crystalline material are captured within the framework of band theory by a matrix Hamiltonian that is continuously parameterized by a momentum variable which takes values in the *Brillouin zone*. The periodicity in real space is reflected in the fact that the Brillouin zone is a d-dimensional torus, where d is the spatial dimension of the material. At low temperature, the precise definition of an insulator can be formulated as being a material for which the *Fermi level* lies within a band gap. Mathematically, this condition translates into the requirement that the eigenvalues of the matrix Hamiltonian never vanish as function of momentum. This is called the *gap condition*.

For an insulator then, there is a splitting of the single-electron states into bands above the Fermi level, called the *conduction* bands, and bands below the Fermi level, called the *valence* bands. In the ground state, the valence bands are occupied, but the conduction bands are empty. For two-dimensional insulators a $\mathbb{Z}$-valued topological invariant called the *first Chern number* can be associated to the valence bands [3, 4]. Physically speaking, the topological nature of this invariant means that it cannot be changed by *adiabatic deformations* of the insulator that preserve the gap [5]. Mathematically, these deformations correspond to a notion of homotopy. The first Chern number has a physically measurable interpretation as the Hall conductivity of a two-dimensional insulator [6]. Moreover, the *bulk-boundary correspondence* states that an insulator with a non-trivial first Chern number necessarily has gapless, conducting edge states [7, 8, 9].

A natural question that arises is whether there are more topological invariants for insulators in less than or equal to three dimensions than just the first Chern number for two-dimensional insulators. It turns out that, in the absence of additional symmetries, the answer to this question is negative; the first Chern number for two-dimensional insulators is the *only* topological invariant possible. In the first part of this work, we will establish this fact using basic tools from algebraic topology including homotopy, homology and cohomology. First, we formulate the topological equivalence of insulators in terms of homotopy classes of maps from the Brillouin zone into a Grassmannian manifold [10]. From this, it can already be concluded that the only topological invariant for insulators in dimensions less than or equal to three is a $\mathbb{Z}$ invariant in two dimensions. However, it is not obvious that this invariant coincides with the first Chern number. In order to establish that this is indeed the case, we formulate the topological classification of insulators a second time, but now in terms of the classification of *complex vector bundles* over the Brillouin zone [11]. Because the Brillouin zone is a low dimensional torus, a full classification of the complex vector bundles over it is readily obtained, and seen to be given precisely by the first Chern number.

In the presence of time-reversal symmetry this story changes. First of all, it can be shown that the first Chern number of a two-dimensional insulators with time-reversal symmetry always vanishes [11]. Therefore, one would be tempted to conclude that there are no topologically nontrivial insulators with time-reversal symmetry. This conclusion would be correct with respect to the notion of topological equivalence that has been used

so far. However, in the context of time-reversal symmetric insulators, a new notion of topological equivalence that not only demands the preservation of the gap, but also of the time-reversal symmetry itself has to be considered. Doing so, a $\mathbb{Z}_2$-valued topological invariant for time-reversal symmetric insulators in both two, as well as three dimensions has been discovered and developed by Fu, Kane and Mele in [12, 13, 14, 15], and is therefore called the *FKM invariant*. We review the construction of this invariant and its topological nature in detail in the so-called *continuum limit*. Physically, this limit captures the response of the insulator to electric fields that vary slowly in space with respect to the inter-atomic spacing. Mathematically, it has the effect of replacing the Brillouin torus by a sphere [16], which simplifies the analysis.

Similar to the case without symmetries, one may ask whether there are more topological invariants for time-reversal symmetric insulators aside from the FKM invariant. Again, the answer turns out to be negative. In the symmetry less case, the analogous question could be answered by associating a complex vector bundle to an insulator and studying the classification of complex vector bundles. It was realized by DeNittis and Gomi [17, 18] that to treat the time-reversal symmetric case, one has to consider *"Quaternionic" vector bundles*, which are complex vector bundles with additional structure that precisely captures the action of the time-reversal symmetry on the valence bands, and is similar to Atiyah's KR-theory [19]. By studying the classification of "Quaternionic" vector bundles over low dimensional spaces using equivariant Borel cohomology, DeNittis and Gomi established the fact that the FKM invariant is the only topological invariant for time-reversal symmetric topological insulators. The main result of the second part of the present work is a reformulation of this result specifically for spheres in terms of equivariant homotopy and equivariant Čech cohomology.

The thesis is divided into two parts. The first part is dedicated to the study of topological insulators not constrained by any symmetries. The main subject of the first part is the topological classification of insulators in one, two and three dimensions in terms of the first Chern number. In Section 2, the mathematical formalism for insulators is introduced in the language of band theory and the tight binding model, and a first classification in terms of homotopy classes of maps is provided. In Section 3, we review the basic theory of complex vector bundles and Chern classes. Then, in Section 4, we show how to associate a complex vector bundle with a connection to an insulator. What is more, we show that the first Chern number of this vector bundle has a physical interpretation as the Hall conductivity of a two-dimensional insulator. In Section 5, the topological classification of insulators is shown to be equivalent to the classification of complex vector bundles. Subsequently, the classification of complex vector bundles over base spaces of dimension less than or equal to three is shown to be realized by the first Chern number. We also identify the two-dimensional, two-band model as the protoypical model for a topological insulator, and study it in detail. The first part is brought to an end in Section 6, where we provide an interpretation of the bulk-boundary correspondence based on a chiral-anomaly argument.

In the second part, we treat the classification of topological insulators with time-reversal symmetry, where we focus on continuum models. We start in Section 7 with the construction of the FKM invariant for these insulators. In Section 8, we introduce "Quaternionic" vector bundles, argue how they are related to time-reversal invariant insulators, and show their classification in terms of a $\mathbb{Z}_2$ invariant, which can be constructed only under certain special assumptions. Finally, in Section 9, we generalize Čech and sheaf cohomology to an equivariant setting in order to prove that the assumptions needed in Section 8 are met for continuum models of insulators.

Part I
Topological Insulators

2 Insulators and topological equivalence

The goal of this section is to establish a topological classification of insulators in terms of homotopy classes of maps. In order to do this, it is necessary to restrict ourselves to one specific choice of mathematical formalism for the description of insulators. The formalism of choice will be that of electronic band theory. More precisely, it will be the so-called *tight-binding model*, which is an approximation to band theory that is particularly suited for insulators [1]. The notion of topological equivalence of insulators has been physically formulated in terms of adiabatic deformations [5]. For tight-binding models this notion can be reformulated in terms of homotopy, which yields the desired topological classification of insulators [10].

We begin this section with a review of band theory for crystalline materials in terms of the Zak and Bloch transforms along the lines of [11] and provide a precise definition of what an insulator is in this context. Then, the tight-binding model is introduced as an approximation to the general band theory and we establish our working definition of a mathematical model for insulators. Finally, we define what it means for two insulators to be topologically equivalent and the topological classification of insulators is exhibited in terms of homotopy classes of maps from the Brillouin zone to a Grassmannian manifold.

2.1 Electronic band theory of crystalline materials

A crystalline material consists of positively charged atomic cores arranged in periodic fashion in space and negatively charged electrons moving around these cores. In principle, all the constituents interact with each other via the electromagnetic force. However, describing all these interactions is not feasible and also not necessary for our purposes. What we will do instead is work with approximations. In particular, we assume that the interactions of a single electron with the atomic cores and the other electrons can be approximated by an effective periodic potential. This is a standard approximation in solid state physics called the "non-interacting electron" or "independent electron" approximation. The non-interacting electron approximation reduces the description of the electrons in the material to a quantum mechanical single-particle problem. low-energy electronic properties of a crystalline material can then be obtained from the so-called band structure, which is a particularly useful representation of all the single-particle eigenstates as a function of a momentum variable.

A block of crystalline material sits in three dimensional space. To study the bulk properties of a block of crystalline material we often model the spatial extent of the block by $\mathbb{R}^3$. The state of an electron inside such a three dimensional block has two parts: a spatial part and a spin part. The spatial part is a complex valued function $\Psi : \mathbb{R}^3 \to \mathbb{C}$ called the *wavefunction*. The absolute value squared of the wavefunction gives the probability distribution for finding the electron in a certain location. In order for this to make sense as a probability distribution, one typically requires the wavefunction to be a normalized element of $L^2(\mathbb{R}^3, \mathbb{C})$. The spin part is a two component vector in $\mathbb{C}^2$, which encodes the orientation of the spin magnetic moment of the electron . The full single electron Hilbert space for an electron in three spatial dimensions is the tensor product $L^2(\mathbb{R}^3, \mathbb{C}) \otimes \mathbb{C}^2$. Sometimes, however, we want to model electrons restricted to move in a two-dimensional plane, which changes the $L^2(\mathbb{R}^3, \mathbb{C})$ component of the Hilbert space into $L^2(\mathbb{R}^2, \mathbb{C})$. In other situations, we want to model electrons whose spin is polarized by some external magnetic field, which changes the $\mathbb{C}^2$ component of the Hilbert space into $\mathbb{C}$. In what follows, we will ignore the spin part of the electron states and consider an electron in d spatial dimensions modeled by a single-particle Hilbert space $\mathcal{H} = L^2(\mathbb{R}^d, \mathbb{C})$.

Having specified the single-particle Hilbert space, we will now specify the single-particle Hamiltonian for the spinless fermions in a periodic potential arising from the atomic cores in the crystal. The periodicity of the crystal is encoded in a *lattice* Γ, which is an additive subgroup of $\mathbb{R}^d$ that is isomorphic to $\mathbb{Z}^d$. The lattice Γ is generated by d linearly independent vectors $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_d \in \mathbb{R}^d$ such that every element $\boldsymbol{\gamma} \in \Gamma$ can be written uniquely as an integer linear combination $\boldsymbol{\gamma} = \sum_{i=1}^d a_i \boldsymbol{\gamma}_i$ with $a_i \in \mathbb{Z}$. For later use we also define the *dual lattice*, or *reciprocal lattice*, Γ^* to be the lattice generated by the dual vectors $\boldsymbol{\gamma}_1^*, \ldots, \boldsymbol{\gamma}_d^* \in \mathbb{R}^d$ that are defined such that $\boldsymbol{\gamma}_i^* \cdot \boldsymbol{\gamma}_j = 2\pi\delta_{ij}$. Moreover, for a lattice Γ generated by $\{\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_d\}$ we define the unit cell Y to be

$$Y = \left\{ \boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} = \sum_{i=1}^d b_i \boldsymbol{\gamma}_i \text{ for } b_i \in [-1/2, 1/2] \right\} .$$

Similarly, we define a corresponding unit cell Y^* for the dual lattice Γ^* . The unit cell Y^* of the dual lattice is called the *Brillouin zone*. We also define the tori $\mathbb{T}_Y := \mathbb{R}^d/\Gamma$ and $\mathbb{T}_{Y^*} := \mathbb{R}^d/\Gamma^*$. The dual space torus $\mathbb{T}_{Y^*}$ is called the *Brillouin torus*. The periodic potential generated by the atomic cores in the crystal is described by a function $V_\Gamma : \mathbb{R}^d \to \mathbb{R}$ that satisfies $V_\Gamma(\boldsymbol{x} + \boldsymbol{\gamma}) = V_\Gamma(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^d$ and all $\boldsymbol{\gamma} \in \Gamma$. For a single fermion of mass m in the periodic potential V_Γ the Hamiltonian is

$$H = -\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2 + V_\Gamma \, . \tag{2.1}$$

The Hamiltonian is an operator on the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^d, \mathbb{C})$. The V_Γ part acts by multiplication on a wavefunction $\Psi \in \mathcal{H}$ through $(V_\Gamma \Psi)(\boldsymbol{x}) = V_\Gamma(\boldsymbol{x})\Psi(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^d$. Strictly speaking, the Hamiltonian H cannot act on $L^2(\mathbb{R}^d, \mathbb{C})$ because a function $\Psi \in L^2(\mathbb{R}^d, \mathbb{C})$ need not be differentiable. The proper Hilbert space to work with would be the Sobolev space $W^{2,2}(\mathbb{R}^d, \mathbb{C})$, consisting of functions whose derivatives up to order two are $L^2(\mathbb{R}^d, \mathbb{C})$. We will leave this distinction implicit. We are thus faced with having to solve the eigenproblem

$$H\Psi = \epsilon\Psi \, , \tag{2.2}$$

for all single-particle eigenstates of H. Just solving the eigenproblem in this form will not yield a band structure though. In order to get the band structure, we exploit the periodicity of the problem by performing the so-called *Bloch-Floquet-Zak transform* [11], or just *Zak transform*, for short.

Morally, the Zak transform is an isometry from the single-particle Hilbert space of arbitrary wavefunctions to a "Hilbert space of periodic wavefunctions". Initially, one can define the Zak transform of a rapidly decreasing function $\Psi \in \mathcal{S}(\mathbb{R}^d, \mathbb{C})$, with $\mathcal{S}(\mathbb{R}^d, \mathbb{C})$ the Schwartz space,[1] to be the function $\mathcal{U}_Z\Psi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ given by

$$(\mathcal{U}_Z\Psi)(\boldsymbol{k}, \boldsymbol{x}) = \sum_{\boldsymbol{\gamma}\in\Gamma} e^{-i\boldsymbol{k}\cdot(\boldsymbol{x}+\boldsymbol{\gamma})}\Psi(\boldsymbol{x} + \boldsymbol{\gamma}) \, , \tag{2.3}$$

where $\boldsymbol{x}$ indicates a variable in real space and $\boldsymbol{k}$ indicates a variable in reciprocal space. The function $\mathcal{U}_Z\Psi$ has the following periodicity properties: In the real space variable

$$(\mathcal{U}_Z\Psi)(\boldsymbol{k}, \boldsymbol{x} + \boldsymbol{\gamma}) = \sum_{\boldsymbol{\gamma}'\in\Gamma} e^{-i\boldsymbol{k}\cdot(\boldsymbol{x}+\boldsymbol{\gamma}+\boldsymbol{\gamma}')}\Psi(\boldsymbol{x} + \boldsymbol{\gamma} + \boldsymbol{\gamma}') = (\mathcal{U}_Z\Psi)(\boldsymbol{k}, \boldsymbol{x}) \, ,$$

[1] The Schwartz space consists of functions all of whose derivatives vanish faster than any polynomial at infinity.

for all $\boldsymbol{\gamma} \in \Gamma$, and in the reciprocal space variable

$$(\mathcal{U}_Z\Psi)(\boldsymbol{k}+\boldsymbol{\lambda},\boldsymbol{x}) = \sum_{\boldsymbol{\gamma}\in\Gamma} e^{-i(\boldsymbol{k}+\boldsymbol{\lambda})\cdot(\boldsymbol{x}+\boldsymbol{\gamma})}\Psi(\boldsymbol{x}+\boldsymbol{\gamma}) = e^{-i\boldsymbol{\lambda}\cdot\boldsymbol{x}}(\mathcal{U}_Z\Psi)(\boldsymbol{k},\boldsymbol{x})\,,$$

for all $\boldsymbol{\lambda} \in \Gamma^*$. The first property tells us that for fixed $\boldsymbol{k} \in \mathbb{R}^d$, the function $(\mathcal{U}_Z\Psi)(\boldsymbol{k},\cdot)$ is periodic in the real space lattice Γ. In other words, $(\mathcal{U}_Z\Psi)(\boldsymbol{k},\cdot) \in L^2(\mathbb{T}_Y,\mathbb{C})$ is a function on the torus $\mathbb{T}_Y$. If we define the *fiber Hilbert space* $\mathcal{H}_f := L^2(\mathbb{T}_Y,\mathbb{C})$ and a representation of the dual lattice

$$\tau : \Gamma^* \to \mathcal{U}(\mathcal{H}_f), \qquad (\tau(\boldsymbol{\lambda})\psi)(\boldsymbol{x}) = e^{i\boldsymbol{\lambda}\cdot\boldsymbol{x}}\psi(\boldsymbol{x})\,,$$

then the second periodicity property implies that $\mathcal{U}_Z\Psi$ is an element of the Hilbert space

$$\mathcal{H}_\tau = \{\psi \in L^2_{\mathrm{loc}}(\mathbb{R}^d,\mathcal{H}_f) : \psi(\boldsymbol{k}-\boldsymbol{\lambda}) = \tau(\boldsymbol{\lambda})\psi(\boldsymbol{k}) \text{ for all } \boldsymbol{\lambda}\in\Gamma^*\}\,,$$

where $L^2_{\mathrm{loc}}(\mathbb{R}^d,\mathcal{H}_f)$ are those functions that are square integrable on compact subsets of $\mathbb{R}^d$. The Hilbert space $\mathcal{H}_\tau$ is equipped with the inner product

$$\langle\varphi,\psi\rangle_{\mathcal{H}_\tau} = \int_{Y^*}\frac{d\boldsymbol{k}}{|Y^*|}\,\langle\varphi(\boldsymbol{k}),\psi(\boldsymbol{k})\rangle_{\mathcal{H}_f}\,,$$

where $|Y^*|$ is the volume of the reciprocal space unit cell. Note in particular that we integrate $\boldsymbol{k}$ only over the unit cell Y^* since all information of an element of $\mathcal{H}_\tau$ can be recovered from its behaviour on the Brillouin zone Y^*. This $\mathcal{H}_\tau$ is our Hilbert space of periodic functions. It is now easy to check that the Zak transform extends to a unitary operator

$$\mathcal{U}_Z : L^2(\mathbb{R}^d,\mathbb{C}) \to \mathcal{H}_\tau\,.$$

Indeed, for $\Phi,\Psi \in L^2(\mathbb{R}^d,\mathbb{C})$, we compute[2]

$$\begin{aligned}\langle\mathcal{U}_Z\Phi,\mathcal{U}_Z\Psi\rangle &= \int_{Y^*}\frac{d\boldsymbol{k}}{|Y^*|}\int_{\mathbb{T}_Y}d\boldsymbol{x}\sum_{\boldsymbol{\gamma},\boldsymbol{\gamma}'}e^{i\boldsymbol{k}\cdot(\boldsymbol{\gamma}-\boldsymbol{\gamma}')}\Phi^*(\boldsymbol{x}+\boldsymbol{\gamma})\Psi(\boldsymbol{x}+\boldsymbol{\gamma}')\\ &= \sum_{\boldsymbol{\gamma}}\int_{\mathbb{T}_Y}d\boldsymbol{x}\,\Phi^*(\boldsymbol{x}+\boldsymbol{\gamma})\Psi(\boldsymbol{x}+\boldsymbol{\gamma}) = \int_{\mathbb{R}^d}d\boldsymbol{x}\,\Phi^*(\boldsymbol{x})\Psi(\boldsymbol{x}) = \langle\Phi,\Psi\rangle\,.\end{aligned}$$

Moreover, the inverse Zak transform is given by

$$(\mathcal{U}_Z^{-1}\psi)(\boldsymbol{x}) = \int_{Y^*}\frac{d\boldsymbol{k}}{|Y^*|}\,e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\psi(\boldsymbol{k},[\boldsymbol{x}])\,,$$

where $[\boldsymbol{x}] \in \mathbb{T}_Y$ is the class represented by $\boldsymbol{x} \in \mathbb{R}^d$.

We use the Zak transform to transport the single-particle problem in $\mathcal{H}$ over to $\mathcal{H}_\tau$. In particular, we claim that, because of the periodicity of the Hamiltonian $H : \mathcal{H} \to \mathcal{H}$, the Zak transform of H decomposes as

$$\mathcal{U}_Z H\mathcal{U}_Z^{-1} = \int_{Y^*}^{\oplus}\frac{d\boldsymbol{k}}{|Y^*|}\,H_f(\boldsymbol{k})\,, \tag{2.4}$$

[2]We use the relation $\int_{Y^*}\frac{d\boldsymbol{k}}{|Y^*|}\,e^{i\boldsymbol{k}\cdot(\boldsymbol{\gamma}-\boldsymbol{\gamma}')} = \delta_{\boldsymbol{\gamma},\boldsymbol{\gamma}'}$. For later use, the relation with the roles of $\boldsymbol{k}$ and $\boldsymbol{\gamma}$ reversed is $\sum_{\boldsymbol{\gamma}\in\Gamma}e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{\gamma}} = \delta(\boldsymbol{k}-\boldsymbol{k}')$.

where H_f is a so-called *fiber Hamiltonian*. The $\int^{\oplus}$-notation means that $H_f(\boldsymbol{k})$ acts on the copy of $\mathcal{H}_f$ that is associated to the point $\boldsymbol{k} \in Y^*$. Explicitly doing the Zak transform for $H = -\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2 + V_\Gamma$ makes this clear. If $\varphi \in \mathcal{H}_\tau$, then

$$\begin{aligned}
(\mathcal{U}_Z H \mathcal{U}_Z^{-1})(\varphi)(\boldsymbol{k},\boldsymbol{x}) &= \sum_{\boldsymbol{\gamma}\in\Gamma} e^{-i\boldsymbol{k}\cdot(\boldsymbol{x}+\boldsymbol{\gamma})}(H\mathcal{U}_Z^{-1}\varphi)(\boldsymbol{x}+\boldsymbol{\gamma}) \\
&= \sum_{\boldsymbol{\gamma}\in\Gamma} e^{-i\boldsymbol{k}\cdot(\boldsymbol{x}+\boldsymbol{\gamma})}\left(-\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2 + V_\Gamma(\boldsymbol{x}+\boldsymbol{\gamma})\right)\int_{Y^*}\frac{d\boldsymbol{l}}{|Y^*|}\, e^{i\boldsymbol{l}\cdot(\boldsymbol{x}+\boldsymbol{\gamma})}\varphi(\boldsymbol{l},[\boldsymbol{x}]) \\
&= \int_{Y^*}\frac{d\boldsymbol{l}}{|Y^*|}\sum_{\boldsymbol{\gamma}\in\Gamma} e^{i(\boldsymbol{l}-\boldsymbol{k})\cdot(\boldsymbol{x}+\boldsymbol{\gamma})}\left(-\frac{\hbar^2}{2m}(\boldsymbol{\nabla}+i\boldsymbol{l})^2 + V_\Gamma(\boldsymbol{x})\right)\varphi(\boldsymbol{l},[\boldsymbol{x}]) \\
&= \int_{Y^*}\frac{d\boldsymbol{l}}{|Y^*|}\,\delta(\boldsymbol{l}-\boldsymbol{k})e^{i(\boldsymbol{l}-\boldsymbol{k})\cdot\boldsymbol{x}}\left(-\frac{\hbar^2}{2m}(\boldsymbol{\nabla}+i\boldsymbol{l})^2 + V_\Gamma(\boldsymbol{x})\right)\varphi(\boldsymbol{l},[\boldsymbol{x}]) \\
&= \frac{1}{|Y^*|}\left(-\frac{\hbar^2}{2m}(\boldsymbol{\nabla}+i\boldsymbol{k})^2 + V_\Gamma(\boldsymbol{x})\right)\varphi(\boldsymbol{k},[\boldsymbol{x}])\,.
\end{aligned}$$

This computation shows that indeed, the output of $\mathcal{U}_Z H \mathcal{U}_Z^{-1}$ at momentum $\boldsymbol{k}$ only depends on the input at $\boldsymbol{k}$, and it yields the decomposition in Eq. (2.4) when we set

$$H_f(\boldsymbol{k}) = -\frac{\hbar^2}{2m}(\boldsymbol{\nabla}+i\boldsymbol{k})^2 + V_\Gamma(\boldsymbol{x})\,. \tag{2.5}$$

The periodicity of the Hamiltonian crucially came in to allow us to perform the sum over the $\boldsymbol{\gamma}$'s yielding the $\delta(\boldsymbol{l}-\boldsymbol{k})$. One can think of $\boldsymbol{k}$ as a momentum, so the above computation is a variant of the familiar statement that translationally invariant Hamiltonians are diagonal in momentum space. The single-particle eigenproblem Eq. (2.2) for a spinless fermion in a crystal can thus be transformed via the Zak transform to the family of eigenproblems

$$H_f(\boldsymbol{k})u(\boldsymbol{k}) = \epsilon(\boldsymbol{k})u(\boldsymbol{k})\,, \tag{2.6}$$

for $\boldsymbol{k} \in Y^*$ and $u(\boldsymbol{k}) \in \mathcal{H}_f$.

We are now in a position to say what a band structure is. The eigenproblem Eq. (2.6) has multiple solutions $\epsilon_a(\boldsymbol{k})$ labelled by an index $a \in \mathcal{A}$. Each solution $\epsilon_a(\boldsymbol{k})$ is called an *energy band* and the index a is called the *band index*. The collection of all the $\{\epsilon_a\}_{a\in\mathcal{A}}$ is the *band structure*. The band structure can be represented schematically in a diagram like Fig. 1, where for simplicity we have assumed a to be discrete index. In fact, we will argue later that for the low-energy physics it always suffices to consider a as a discrete index and ignore any continuous portions of the spectrum.

As a final comment, the Zak transform is closely related to the so-called *Bloch transform*. Whereas the Zak transform comes with periodicity in real space and "quasi-periodicity" in reciprocal space, the Bloch transform makes the opposite tradeoff. The Bloch transform of $\Psi \in \mathcal{H}$ can be defined to be

$$(\mathcal{U}_B\Psi)(\boldsymbol{k},\boldsymbol{x}) = e^{i\boldsymbol{k}\cdot\boldsymbol{x}}(\mathcal{U}_Z\Psi)(\boldsymbol{k},\boldsymbol{x}) = \sum_{\boldsymbol{\gamma}\in\Gamma} e^{-i\boldsymbol{k}\cdot\boldsymbol{\gamma}}\Psi(\boldsymbol{x}+\boldsymbol{\gamma})\,.$$

For $\Psi \in \mathcal{H}$, the function $\mathcal{U}_B\Psi$ is periodic in $\boldsymbol{k}$, but for fixed $\boldsymbol{k}$ it is an element of

$$\mathcal{H}_{\boldsymbol{k}} = \{\psi \in L^2_{\mathrm{loc}}(\mathbb{R}^d,\mathbb{C}) : \psi(\boldsymbol{x}+\boldsymbol{\gamma}) = e^{i\boldsymbol{k}\cdot\boldsymbol{\gamma}}\psi(\boldsymbol{x}) \text{ for all } \boldsymbol{\gamma}\in\Gamma\}\,.$$

If $u(\boldsymbol{k})$ is a Zak function that solves Eq. (2.6), then the *Bloch function* $\psi_{\boldsymbol{k}}(\boldsymbol{x}) = e^{i\boldsymbol{k}\boldsymbol{x}}u(\boldsymbol{k},\boldsymbol{x})$ solves $H\psi_{\boldsymbol{k}} = \epsilon(\boldsymbol{k})\psi_{\boldsymbol{k}}$, with H the original single-particle Hamiltonian. The above discussion on Zak and Bloch transforms thus gives a result similar to the consequence of

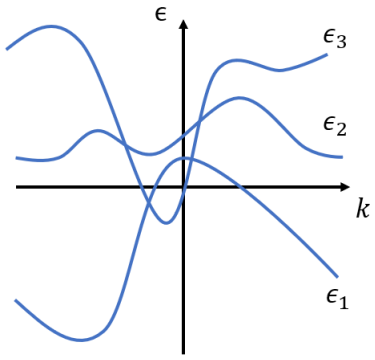

Figure 1: Schematic example of a band structure with three energy bands labelled by ϵ_1, ϵ_2 and ϵ_3 .

the Bloch theorem that there is a basis of $\mathcal{H}$ consisting of eigenfunctions of H of Bloch functions $\psi_{\boldsymbol{k}}(\boldsymbol{x})$ [1] .

2.2 Conductors and insulators

The difference between conductors and insulators can be discerned by considering the representation of the ground state in the band structure. Up to now, we have been considering a crystal that extends out to infinity. However, for the arguments that follow it is easiest to restrict to the physical situation of a finite size crystal. Let us say the crystal is a cube made up of N unit cells, where N is some very large number. A consequence of the finiteness of the crystal is that the momentum $\boldsymbol{k}$ becomes discretized. In particular, each energy band now consists of a finite number of states.

To obtain the ground state, note that each atom in the crystal comes with a fixed number of electrons so that there is in total a finite number of electrons in the crystal. The ground state is the lowest possible energy configuration that can be obtained by distributing the electrons of the crystal over the single-particle states in accordance with the Pauli principle. The band structure tells us what the states are that individual electrons can occupy. By the Pauli exclusion principle all the electrons have to be in different states. To obtain the ground state, we should thus start filling the single-particle states from the lowest energy state up, until we are out of electrons.

The *Fermi level* is then roughly defined to be the highest occupied energy level in the ground state. A crystal is a conductor when the Fermi level lies within a band. A crystal is an insulator when the Fermi level lies at the top of a band. A more precise definition of the Fermi level is that it coincides with the chemical potential at zero temperature. The chemical potential μ as a function of temperature T is determined by particle conservation. The number of particles in the system is $N_{\text{particles}} = \int d\epsilon\, N_{\text{FD}}(\epsilon - \mu(T), T)\rho(\epsilon)$, where N_{FD} is the Fermi-Dirac distribution and ρ is the density of states. If T changes, the Fermi-Dirac distribution changes and hence μ has to change to preserve the number of particles. The limit $\mu(T \to 0)$ defines the Fermi level uniquely, and for an insulator it will be somewhere inside a band gap. We will always redefine the energy levels so that the Fermi level lies at $\epsilon = 0$. See Fig. 2 for schematic examples of band structures for conductors and insulators. As a note on terminology, the bands with negative energy are referred to as *valence bands*, and the bands with positive energy are referred to as *conduction bands*.

Let us now reconcile this band theoretic definition of conductors and insulators with the

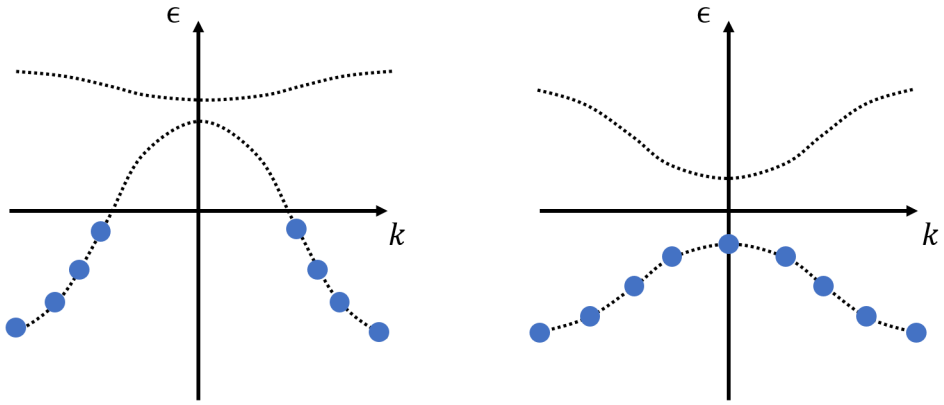

Figure 2: (Left) Band structure of a conductor. (Right) Band structure of an insulator. Energy bands are indicated with dotted line and occupied states with blue dots. The Fermi level lies at $\epsilon = 0$ by convention.

heuristic definition in terms of conduction of electricity. At zero temperature and without any externally applied fields, the crystal will be in the ground state. The difference between conductors and insulators is their response to a small external electric field. For simplicity, consider a one-dimensional crystal with N unit cells lying along the x-axis. Suppose we apply an electric field E in the x-direction. A vector potential that realizes this electric field is $A_t(t,x) = 0$ and $A_x(t,x) = Et$. The electrons couple to the vector potential by the minimal substitution

$$k \to k - \frac{e}{\hbar} Et\,.$$

Classically, this coupling reflects the fact the electric field acts as a force on the electrons thereby accelerating them, or in other words, changing their momentum. Quantum mechanically, if the electric field is small, then by the adiabatic theorem the electrons will actually stay within their respective eigenstates, but the eigenstates themselves start flowing along the bands according to the change in momentum. After some time, all single-particle states will have shifted by one place as in Fig. 3.[3] In this picture we have properly taken into account that the energy bands are periodic functions on the Brillouin torus. For a conductor, this shift results in a net balance of more right moving than left moving electrons. Hence, there is a current. For an insulator, however, the occupation of the states is invariant under the flow of the eigenstates, so there is no current when a small electric field is applied to an insulator.

2.3 The tight-binding model and the continuum limit

The problem of finding the eigenstates of the single-particle Hamiltonian $H = -\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2 + V_\Gamma$ in $L^2(\mathbb{R}^d, \mathbb{C})$ has been reformulated to finding the eigenstates of the family of Zak Hamiltonians $H_f(\boldsymbol{k}) = -\frac{\hbar^2}{2m}(\boldsymbol{\nabla} + i\boldsymbol{k})^2 + V_\Gamma$ in $L^2(\mathbb{T}_Y, \mathbb{C})$. Conceptually this reformulation is desirable because it allows the solution to be interpreted in terms of a band structure. Computationally, however, no progress has been made. It is not easier to find eigenstates of $H_f(\boldsymbol{k})$ than it is finding eigenstates of H. Approximate methods are needed.

First of all, we are not interested in all eigenstates. Our objective lies in capturing the low-energy physics, which is dominated by the eigenstates close to the Fermi level. For all

[3] We repeat this argument in more detail in Section 6.

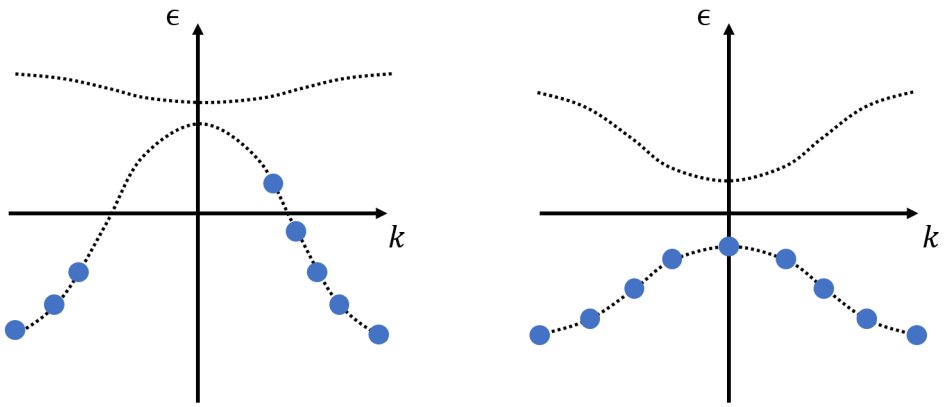

Figure 3: (Left) Occupation of states after turning on electric field for some time in conductor. There is a balance of more right moving and left moving electrons is disturbed, hence there is a in current. (Right) Occupation of states after turning on electric field for some time in insulator. There is no change, so no current.

$\boldsymbol{k}$ in the Brillouin zone Y^* the spectrum of $H_f(\boldsymbol{k})$ contains both a discrete and a continuous part. The continuous part corresponds to high energy scattering states associated to the continuum outside the crystal. We will ignore these scattering states. The only states we are interested in are those corresponding to a subset of the discrete part of the spectrum of $H_f(\boldsymbol{k})$ which are close the Fermi level. Recall that there is a one-to-one correspondence between the eigenfunctions $H_f(\boldsymbol{k})$ and the Bloch functions, which are eigenfunctions of H, given by $\psi(\boldsymbol{k}, \boldsymbol{x}) = e^{i\boldsymbol{k}\cdot\boldsymbol{x}} u(\boldsymbol{k}, \boldsymbol{x})$. The *tight-binding model* gives an approximate approach to finding the Bloch functions around the Fermi level. We review the tight-binding model here along the lines of [1] and [20].

The main assumption in the tight-binding model is that the electrons in the crystal are "tightly bound" to the individual atomic cores that make up the crystal, in the sense that the atomic orbitals of the isolated atoms are a good zeroth order approximation to the Bloch functions. To be more precise, consider an isolated atom located at the origin in $\mathbb{R}^d$. The atom has a spherically symmetric potential $V_{at}(\boldsymbol{x})$ and the Hamiltonian describing non-relativistic electrons orbiting around the static core is

$$H_{at} = -\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2 + V_{at}(\boldsymbol{x})\,.$$

The eigenvectors $\phi_a(\boldsymbol{x})$ corresponding to the discrete part of the spectrum of H_{at} are the atomic orbitals. To model the low-energy physics, let us consider a collection of atomic orbitals labelled by $a = 1, \ldots, M$. Suppose we construct a crystal by placing copies of the atom on a lattice Γ. The Hamiltonian describing an electron in the crystal is

$$H = -\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2 + \sum_{\boldsymbol{\gamma}} V_{at}(\boldsymbol{x} - \boldsymbol{\gamma}) = H_{at} + \Delta V(\boldsymbol{x})\,, \tag{2.7}$$

where $\Delta V(\boldsymbol{x}) = \sum_{\boldsymbol{\gamma}\neq 0} V_{at}(\boldsymbol{x} - \boldsymbol{\gamma})$. If the lattice spacing were macroscopic, say one meter, then the atoms would not notice each others presence. The overlap of their orbitals would be astronomically small and effectively vanishes. In this limiting case, the orbitals of the isolated atoms are extremely good approximations to the solutions of the eigenstates of the crystal. Now imagine that we shrink the lattice down to the actual lattice spacing.

The atomic orbitals will overlap nontrivially, however, the assumption of the tight-binding model is that the overlaps are sufficiently small such that the orbitals of the isolated atoms are still good zeroth order approximations to the eigenstates of the crystal.

The atomic orbitals can then be used to approximate the Bloch functions. In order to regularize the computations, we restrict ourselves to a finite portion of the crystal of $N+1$ atoms corresponding to points $\{\boldsymbol{\gamma}_j\}_{j=0}^N$ in the lattice, where $\boldsymbol{\gamma}_0 = 0$. Recall that a Bloch function $\psi(\boldsymbol{k})$ has to be part of the Hilbert space $\mathcal{H}_{\boldsymbol{k}}$, which means that it has to be periodic in momentum and quasi-periodic in space according to

$$\psi(\boldsymbol{k}, \boldsymbol{x} + \boldsymbol{\gamma}) = e^{i\boldsymbol{k}\cdot\boldsymbol{\gamma}}\psi(\boldsymbol{k}, \boldsymbol{x})\,.$$

The ansatz for a basis of Bloch functions is the *linear combination of atomic orbitals* (LCAO),

$$\psi_a(\boldsymbol{k}, \boldsymbol{x}) = \frac{1}{\sqrt{N+1}} \sum_{j=0}^{N} e^{i\boldsymbol{k}\cdot\boldsymbol{\gamma}_j}\phi_a(\boldsymbol{x} - \boldsymbol{\gamma}_j)\,, \tag{2.8}$$

for $a = 1, \ldots, M$. An easy check shows that $\psi_a(\boldsymbol{k})$ indeed satisfies the right periodicity properties.

We assume that the Bloch functions $\{\psi_a\}_{a=1}^M$ of the LCAO ansatz are independent so that for each $\boldsymbol{k}$ they form a basis for an M-dimensional subspace of $\mathcal{H}_{\boldsymbol{k}}$. Restricted to these subspaces, the Hamiltonian H in Eq. (2.7) can be expressed as a $\boldsymbol{k}$-dependent matrix $h(\boldsymbol{k})$ with respect to the basis provided by the LCAO ansatz, whose matrix elements are

$$\begin{aligned}
h_{ab}(\boldsymbol{k}) &= \langle \psi_a(\boldsymbol{k}), H\psi_b(\boldsymbol{k})\rangle \\
&= \frac{1}{N+1} \sum_{j,j'=0}^{N} e^{i\boldsymbol{k}\cdot(\boldsymbol{\gamma}_{j'} - \boldsymbol{\gamma}_j)} \int d\boldsymbol{x}\, \phi_a^*(\boldsymbol{x} - \boldsymbol{\gamma}_j)\left[H_{at} + \Delta V(\boldsymbol{x})\right]\phi_b(\boldsymbol{x} - \boldsymbol{\gamma}_{j'}) \\
&= \sum_{j=0}^{N} e^{-i\boldsymbol{k}\cdot\boldsymbol{\gamma}_j} \int d\boldsymbol{x}\, \phi_a^*(\boldsymbol{x} - \boldsymbol{\gamma}_j)\left[H_{at} + \Delta V(\boldsymbol{x})\right]\phi_b(\boldsymbol{x}) \\
&= \delta_{ab}E_b + \sum_{j=1}^{N} e^{-i\boldsymbol{k}\cdot\boldsymbol{\gamma}_j} 1_{ab}(\boldsymbol{\gamma}_j) + \sum_{j=0}^{N} e^{-i\boldsymbol{k}\cdot\boldsymbol{\gamma}_j}\Delta V_{ab}(\boldsymbol{\gamma}_j)\,,
\end{aligned}$$

where we have defined the matrix elements

$$1_{ab}(\boldsymbol{\gamma}) = \int d\boldsymbol{x}\, \phi_a^*(\boldsymbol{x} - \boldsymbol{\gamma})\phi_b(\boldsymbol{x}) \quad \text{and} \quad \Delta V_{ab}(\boldsymbol{\gamma}) = \int d\boldsymbol{x}\, \phi_a^*(\boldsymbol{x} - \boldsymbol{\gamma})\Delta V(\boldsymbol{x})\phi_b(\boldsymbol{x})\,.$$

If the tight-binding model is appropriate, then the overlap elements $1_{ab}(\boldsymbol{\gamma})$ and $\Delta V_{ab}(\boldsymbol{\gamma})$ fall of quickly when $\boldsymbol{\gamma}$ goes further away from 0. A good approximation to the Hamiltonian can then be obtained by taking only the first couple of terms from the sums over the lattice vectors. For example, one might take only the nearest neighbours, or next-nearest neighbours terms. Our discussion of the tight-binding model so far has been restricted to a model for a collection M orbitals of one atom in a unit cell. However, the same approach can be applied to the case of multiple atoms in a unit cell after introducing sublattice indices next to the usual lattice indices, see [20]. It is important to note that the Bloch functions of the LCAO ansatz do not diagonalize the Hamiltonian H. This means that they are not the eigenstates corresponding to a single energy band. To obtain the energy bands, one should find the eigenvectors of the matrices $h(\boldsymbol{k})$. It is these eigenvectors that correspond to the energy bands.

From now on the tight-binding model will be our model of choice for any insulator. In particular, we adopt the mathematical formalism that the single-particle Hamiltonian of a crystal is modelled by a $\boldsymbol{k}$-dependent matrix $h(\boldsymbol{k})$. In order for $h(\boldsymbol{k})$ to describe an insulator, the Fermi level must be inside the gap. By redefining the energies we can always assume the Fermi level to be at zero energy. Hence, $h(\boldsymbol{k})$ describes an insulator if and only if its eigenvalues never vanish. We will refer to this nonvanishing of the eigenvalues when $\boldsymbol{k}$ ranges over the Brillouin zone as the *gap condition*. Abstractly speaking, a model for insulator in d spatial dimensions that takes into account M bands thus takes the form of a map

$$h : \mathbb{T}^d \to \mathrm{Herm}(M, \mathbb{C}) \,, \tag{2.9}$$

that satisfies the gap condition, where $\mathrm{Herm}(M, \mathbb{C})$ denotes the $M \times M$ Hermitian matrices. The periodicity of the Bloch functions in momentum space has been used to reduce the domain of h from $\mathbb{R}^d$ to the Brillouin torus $\mathbb{T}^d$. The fact that insulators can exhibit non-trivial topological features can be traced back to the fact that the torus appearing as the domain of h is topologically non-trivial.

The abstract tight-binding model Eq. (2.9) contains the low-energy physics of an insulator. It can be used to describe the response of an insulator to an electric field that slowly varies in time. An additional limit of interest is the limit in which the electric field varies slowly in space as well. This is the so-called long-wavelength limit. A wavelenght is "long" when it is large with respect to the interatomic spacing of the crystal. Or conversely, if the interatomic spacing is small with respect to the wavelength of the external field. From the latter point of view of the external field the crystal then approximates a continuum. Hence, the long-wavelength limit is also called the *continuum limit*.

Given a periodic model $h : \mathbb{T}^d \to \mathrm{Herm}(M, \mathbb{C})$ the continuum limit can be obtained by expanding the matrix entries of $h(\boldsymbol{k})$ to first or second order in $\boldsymbol{k}$, since long wavelengths correspond to small values of $\boldsymbol{k}$. For example, in two dimensions one can write down the periodic model

$$h(\boldsymbol{k}) = \sin(k_x)\sigma_x + \sin(k_y)\sigma_y + (m + \cos(k_x) + \cos(k_y))\sigma_z \,,$$

where σ_i for $i = x, y, z$ are the Pauli matrices. Expanding to second order in $\boldsymbol{k}$ yields the continuum model[4]

$$h_{cont}(\boldsymbol{k}) = k_x\sigma_x + k_y\sigma_y + \left(m + 2 - \frac{k_x^2 + k_y^2}{2}\right)\sigma_z \,.$$

Clearly, h_{cont} is no longer defined on the torus. It looks like h_{cont} is defined on $\mathbb{R}^2$ and although this is true, it cannot be the whole story. Indeed, $\mathbb{R}^2$ is topologically trivial which would mean that the continuum model does not contain any topological information. This is however not the case. There is non-trivial topology also in the continuum model. The reason is that the behaviour of the eigenfunctions of $h_{cont}(\boldsymbol{k})$ as $k \to \infty$ serves as a boundary condition. The eigenfunctions of $h_{cont}(\boldsymbol{k})$ become independent of $\boldsymbol{k}$ as $k \to \infty$. Therefore, we can compactify $\mathbb{R}^2$ to S^2 by adding a point at infinity. As a general definition, a *continuum model* for an insulator is given by a map

$$h : S^d \to \mathrm{Herm}(M, \mathbb{C}) \,, \tag{2.10}$$

that satisfies the gap condition.

[4]This model can actually be experimentally realized [21] .

2.4 Homotopy classification of insulators

In the tight binding formalism, an insulator in d spatial dimensions is modelled by a smooth map

$$h : X \to \mathrm{Herm}(M, \mathbb{C})$$

that satisfies the gap condition. Here, X is a d-dimensional smooth manifold that is taken to be either the Brillouin torus $\mathbb{T}^d$ or the sphere S^d, depending on whether we consider the continuum limit or not. Without loss of generality, we will assume the Fermi level to be at zero energy so that the gap condition can equivalently be stated as the requirement that the eigenvalues of $h(p)$ are nonvanishing for all $p \in X$. The solutions to the characteristic equation

$$\det(h - \epsilon) = 0$$

are smooth functions $\epsilon_{\tilde{a}} : X \to \mathbb{R}$ labelled by band indices $\tilde{a} = 1, \ldots, M$.

Physically speaking, two insulators are said to be topologically equivalent if they can be adiabatically deformed into each other [5]. A deformation of a quantum mechanical system is adiabatic if the system, whenever it is prepared in an eigenspace of the Hamiltonian corresponding to an eigenvalue separated from the rest of the spectrum by a finite gap and the deformation occurs slowly enough, stays within that eigenspace during the deformation. In an insulator there is always a gap between the negative energy states and the positive energy states. Therefore, adiabatic deformations of insulators correspond to *continuous* deformations that preserve the insulating gap. Indeed, whenever a continuous deformation is possible one can then perform the deformation sufficiently slow so that, if the gap is maintained, valence states, i.e. negative energy states, will stay valence states. A continuous deformation of a map is called a *homotopy*. The definition of topological equivalence of insulators we adopt in this work is the following:

Definition 2.1. Two insulators $h, h' : X \to \mathrm{Herm}(M, \mathbb{C})$ are said to be *topologically equivalent as insulators* if there exists a homotopy from h to h' that preserves the insulating gap.

An important consequence of this definition is that for an insulator h the precise values of the eigenvalues ϵ_a are topologically irrelevant. The only topological information they contain is the distinction between valence and conduction bands. To see this, consider an insulator $h : X \to \mathrm{Herm}(M, \mathbb{C})$ with $m < M$ valence bands. Denote by $\epsilon_1, \ldots, \epsilon_m : X \to (-\infty, 0)$ the negative eigenvalues of h and by $\epsilon_{m+1}, \ldots, \epsilon_M : X \to (0, \infty)$ the positive ones. Since the Hamiltonian takes values in the Hermitian matrices, it can be diagonalized pointwise by finding its eigenvectors. However, globally continuous eigenvectors over the whole momentum space might not exist. The best we can do in general is to cover X with contractible open sets $\{U_\alpha\}_{\alpha \in \mathcal{A}}$ and find smooth orthonormal eigenvectors $\psi_{\alpha,a} : U_\alpha \to \mathbb{C}^M$ corresponding to ϵ_a for $a = 1, \ldots, M$ over each open U_α. The Hamiltonian may then locally be represented as

$$h|_{U_\alpha} = \sum_{a=1}^{m} \epsilon_a \left|\psi_{\alpha,a}\right\rangle \left\langle\psi_{\alpha,a}\right| + \sum_{a=m+1}^{M} \epsilon_a \left|\psi_{\alpha,a}\right\rangle \left\langle\psi_{\alpha,a}\right|,$$

where $\left|\psi_{\alpha,a}\right\rangle \left\langle\psi_{\alpha,a}\right|$ is the orthogonal projection onto the eigenspace of ϵ_a and we have suggestively split the Hamiltonian into its negative and positive parts. If we change frame by going from U_α to U_β, then on the overlap $U_{\alpha\beta} := U_\alpha \cap U_\beta$ the negative eigenvectors transform amongst themselves by a gauge transformation $g^-_{\alpha\beta} : U_{\alpha\beta} \to SU(m)$, while the positive eigenvectors transform amongst themselves by a gauge transformation $g^+_{\alpha\beta}$:

$U_{\alpha\beta} \to SU(M-m)$. It is then easy to see that the two projectors $\sum_{a=1}^{m} |\psi_{\alpha,a}\rangle \langle\psi_{\alpha,a}|$ and $\sum_{a=m+1}^{M} |\psi_{\alpha,a}\rangle \langle\psi_{\alpha,a}|$ that appear in the Hamiltonian are invariant under this change of frame from U_α to U_β. We can thus unambiguously write

$$h = \sum_{a=1}^{m} \epsilon_a |\psi_a\rangle \langle\psi_a| + \sum_{a=m+1}^{M} \epsilon_a |\psi_a\rangle \langle\psi_a| ,$$

where to evaluate h at a point $p \in X$ we have to pick an open U_α that contains p and use the frame $\{|\psi_{\alpha,a}\rangle\}_{a=1}^{M}$. At this stage, the Hamiltonian can be adiabatically connected to the so-called "flat-band Hamiltonian"

$$h_{flat} = -\sum_{a=1}^{m} |\psi_a\rangle \langle\psi_a| + \sum_{a=m+1}^{M} |\psi_a\rangle \langle\psi_a| , \tag{2.11}$$

by deforming ϵ_a via negative functions to the constant function -1 for $a = 1, \ldots, m$ and via positive functions to the constant function 1 for $a = m+1, \ldots, M$.

The expression (2.11) for the flat band Hamiltonian is independent of the choice of eigenvectors, and this independence can be made manifest using the notion of spectral projectors. Recall that for a linear map $A : \mathbb{C}^M \to \mathbb{C}^M$ with eigenvalue $\lambda \in \mathbb{C}$, the spectral projector $P_\lambda : \mathbb{C}^M \to \mathbb{C}^M$ onto the eigenspace of λ can be represented as

$$P_\lambda = \frac{1}{2\pi i} \int_{C_\lambda} dz \, [A - zI]^{-1} ,$$

where C_λ is a curve in $\mathbb{C}$ that encircles λ once in the counterclockwise direction and such that the only eigenvalue of A contained in the interior of the curve $\mathbb{C}$ is λ. If we let C be a fixed curve in $\mathbb{C}$ enclosing the negative part of the spectrum of h once counterclockwisely, then the spectral projector onto the negative energy eigenspaces

$$P : X \to \mathrm{Mat}(M, \mathbb{C}) , \quad P(p) = \frac{1}{2\pi i} \int_C dz \, [h(p) - zI]^{-1}$$

is a smooth map. The flat band Hamiltonian can be expressed in terms of P as

$$h_{flat} = (-1)P + (+1)(1 - P) = 1 - 2P ,$$

which indeed gives energy -1 for the valence states and energy 1 for the conduction states. There is thus a one-to-one correspondence between topological equivalence classes of insulators and homotopy classes of spectral projectors $P : X \to \mathrm{Mat}(M, \mathbb{C})$.

It will turn out to be useful to reformulate this result as follows. The spectral projector $P(p) : \mathbb{C}^M \to \mathbb{C}^M$ can be identified with the subspace it projects onto. In this way, the family of spectral projectors $P : X \to \mathrm{Mat}(M, \mathbb{C})$ gives rise to a smooth map $X \to G_m(\mathbb{C}^M)$, where $G_m(\mathbb{C}^M)$ is the Grassmannian of m-dimensional subspaces of $\mathbb{C}^M$. Letting $[X, G_m(\mathbb{C}^M)]$ denote the homotopy classes of maps $X \to G_m(\mathbb{C}^M)$, the final form in which we would like to present the topological classification of insulators with M total bands and m valence bands is

$$\{\text{Topological equivalence classes of insulators}\} \longleftrightarrow [X, G_m(\mathbb{C}^M)] .$$

For continuum models in d dimensions, $X = S^d$ so that $[X, G_m(\mathbb{C}^M)] = \pi_d(G_m(\mathbb{C}^M))$, the d-th homotopy group. If $M - m$ and m are sufficiently large with respect to d, then

the homotopy groups of the Grassmannian are known to be[5]

$$\pi_d(G_m(\mathbb{C}^M)) = \begin{cases} \mathbb{Z} & \text{if } d \text{ is even} \\ 0 & \text{if } d \text{ is odd} \end{cases}.$$

This implies that in two spatial dimensions there is an ensemble of topological equivalence classes of continuum models of insulators indexed by $\mathbb{Z}$. Moreover, in one and three spatial dimensions there are only trivial continuum models.

In Section 5, we study this classification more carefully. One issue with the formulation above is that it is in general hard to determine whether two maps are homotopic or not. In other words, given an explicit model for an insulator it is a difficult task to determine what equivalence class it belongs to. Another problem is that the advanced tools used to compute the homotopy classes $[S^d, G_m(\mathbb{C}^M)]$ cannot be used to compute $[\mathbb{T}^d, G_m(\mathbb{C}^M)]$. A more systematic computational method for distinguishing insulators is by computing a topological invariant called the Chern number. We will prove that the Chern number distinguishes all topological equivalence classes of an insulator in $d = 1, 2, 3$. To properly introduce this topological invariant, we will review the theory of Chern classes for vector bundles in the next section.

3 Vector bundles and Chern classes

In this section we introduce the definition of Chern classes for complex vector bundles in the Chern-Weil formalism. In this formalism the Chern classes are defined as deRham cohomology classes represented by invariant polynomials of the curvature 2-form. In order to make sense of this definition, we will first briefly review three preliminaries. First, we review the basic definitions and examples of vector bundles. Second, we review the definition of and try to give some intuition for the notion of deRham cohomology of a smooth manifold. Finally, we review the notion of connections and curvature of a smooth vector bundle as a generalization of the usual deRham complex associated to the exterior derivative on a smooth manifold. More thorough introductions to these concepts can be found in [24, 25, 26].

Although Chern classes are introduced in this section for *smooth* vector bundles, they are actually topological invariants of the underlying *topological* vector bundles. Explicit models of insulators are usually smooth so that the Chern classes can be computed by the methods laid out in this section. The topological nature of the Chern classes will be further explained in Section 5.

3.1 Vector bundles

Intuitively, a vector bundle is a family of vector spaces continuously parameterized by a topological space X. The precise definition is as follows. Let $\mathbb{K} = \mathbb{R}, \mathbb{C}$ be the field of real or complex numbers. A $\mathbb{K}$-*vector bundle of rank* m is a triple (E, X, π), where E and X are topological spaces and $\pi : E \to X$ is a continuous map such that for every point $p \in X$,

1. $\pi^{-1}(p)$ is an m-dimensional $\mathbb{K}$-vector space, called the *fiber* of E,

[5]For $d \leq 2(M-m)$ it can be shown that $\pi_d(G_m(\mathbb{C}^M)) \cong \pi_{d-1}(U(m))$ using the fibration sequence $0 \to U(m) \to V_m(\mathbb{C}^M) \to G_m(\mathbb{C}^M) \to 0$, where $V_m(\mathbb{C}^M)$ is the so-called Stieffel manifold, which is $2(M-m)$-connected. For $d \leq 2m$ it is known that $\pi_{d-1}(U(m))$ is isomorphic to $\mathbb{Z}$ for even d and 0 for odd d. This result is obtained from Bott periodicity and so-called "stable range" considerations [22, 23].

2. there is an open neighbourhood U of p and a homeomorphism $\varphi : \pi^{-1}(U) \to U \times \mathbb{K}^m$ covering the identity on U, meaning $\mathrm{pr}_U \circ \varphi = \pi$ for $\mathrm{pr}_U : U \times \mathbb{K}^m \to U$ the projection onto U, which fiberwise restricts to a linear isomorphism $\varphi_q : \pi^{-1}(q) \to \{q\} \times \mathbb{K}^m$ for all $q \in U$.

The homeomorphism $\varphi : \pi^{-1}(U) \to U \times \mathbb{K}^m$ in the second point of the definition is called a *local trivialization* over $U \subset X$. A vector bundle is said to be *trivial* if there is a local trivialization over X itself, which we then call a *global trivialization*.

If X is a smooth manifold, then a smooth vector bundle over X can be defined by only minor alterations to the above definition of vector bundle. A smooth vector bundle is a triple (E, X, π) with E and X smooth manifolds and $\pi : E \to X$ a smooth map such that properties 1 and 2 above hold, with the alteration that the local trivialization must now be a diffeomorphism instead of just a homeomorphism.

Given two vector bundles $\pi : E \to X$ and $\pi' : E' \to X$, both over the same X, a map $\varphi : E \to E'$ is a *morphism of vector bundles* if $\pi' \circ \varphi = \pi$ and $\varphi_p : E_p \to E'_p$ is a linear map for every $p \in X$. A morphism of vector bundles is an isomorphism already if φ_p is a linear isomorphism for all $p \in X$. For a fixed base space X we let $\mathrm{Vect}^m_{\mathbb{K}}(X)$ denote the set of all $\mathbb{K}$-vector bundles of rank m over X, up to isomorphism.

A basic question when one encounters a vector bundle is whether or not it is isomorphic to the trivial vector bundle. In other words, whether or not it can globally be written as a product $X \times \mathbb{K}^m$. An example of a trivial real line bundle, i.e. a rank 1 vector bundle, over S^1 is the cylinder $S^1 \times \mathbb{R}$. In terms of the precise definition the cylinder as a vector bundle is the triple $(S^1 \times \mathbb{R}, S^1, \mathrm{pr}_{S^1})$. For $p \in S^1$, the fiber $\mathrm{pr}^{-1}_{S^1}(p)$ is equal to $\mathbb{R}$, which is a one-dimensional vector space, and the identity $\mathrm{id} : \mathrm{pr}^{-1}_{S^1}(S^1) \to S^1 \times \mathbb{R}$ is a global trivialization. Henceforth, we will often refer to a vector bundle (E, X, π) as just E and let the base space X and the projection π be implicit.

An example of a non-trivial real line bundle over S^1 is the Möbius bundle. The Möbius bundle can be defined by taking the product space $[0, 1] \times \mathbb{R}$ and gluing the copies of $\mathbb{R}$ at $\{0\}$ and $\{1\}$ by identifying $(0, t)$ with $(1, -t)$. It can be shown that the Möbius bundle indeed does not admit a global trivialization. This is especially easy to see by studying the behaviour of the so-called sections of the Möbius bundle.

For a general $\mathbb{K}$-vector bundle $\pi : E \to X$, a *section* of E is a map $s : X \to E$ such that $\pi \circ s = \mathrm{id}_X$. The set of all sections of E is denoted by $\Gamma(E)$, or sometimes as $\Gamma(X; E)$ if we want to emphasize the base space. Actually, the set of all sections enjoys more structure than just that of a set. Indeed, we can add sections fiberwise to get new sections, and we can also multiply a section by a $\mathbb{K}$-valued function. Hence, the sections $\Gamma(E)$ have the structure of a $C^\infty(X; \mathbb{K})$-module. A *local section* over an open set $U \subset X$ is a section of the restricted vector bundle $E|_U := \pi^{-1}(U)$. A *local frame* of E over an open set $U \subset X$ is a collection of local sections $s_1, \ldots, s_m : U \to E|_U$ such that at each point $p \in U$, the vectors $\{s_1(p), \ldots, s_m(p)\}$ form a basis for the vector space $\pi^{-1}(p)$. A *global frame* is a local frame over X. It is straightforward to show that the existence of a local frame over U is equivalent to the existence of a local trivialization over U. Another criterion for a vector bundle $\pi : E \to X$ to be trivial is thus the existence of a global frame. In the case of a line bundle, a global frame is the same thing as a global non-vanishing section. The cylinder $S^1 \times \mathbb{R}$ has a global non-vanishing section. For example, one can take $s(x) = (x, 1)$. The Möbius bundle, on the other hand, does not have a global non-vanishing section. Indeed, global sections of the Möbius bundle correspond to continuous maps $f : [0, 1] \to \mathbb{R}$ such that $f(0) = -f(1)$. If $f(0) > 0$, then $f(1) < 0$ and so f has to vanish for some $x \in (0, 1)$. This proves that the Möbius bundle is non-trivial.

An important aspect to the study of vector bundles are the local representations of

global objects. For example, one can study global sections of vector bundles by comparing representations of the sections with respect to different choices of local frame. For any vector bundle $\pi : E \to X$ an open cover $\{U_\alpha\}_{\alpha\in\mathcal{A}}$ together with local trivializations $\varphi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{K}^m$ can be found. On the overlaps $U_{\alpha\beta} := U_\alpha \cap U_\beta$ one can compare the trivializations φ_α and φ_β. Consider the diagram

$$\begin{array}{ccc} & \pi^{-1}(U_{\alpha\beta}) & \\ {\scriptstyle\varphi_\alpha}\swarrow & & \searrow{\scriptstyle\varphi_\beta} \\ U_{\alpha\beta}\times\mathbb{K}^m & \xrightarrow{\varphi_\beta\circ\varphi_\alpha^{-1}} & U_{\alpha\beta}\times\mathbb{K}^m \end{array} .$$

On the fibers, the trivialization are given by linear isomorphisms. Therefore, we can write

$$(\varphi_\beta \circ \varphi_\alpha^{-1})(x, v) = (x, g_{\alpha\beta}(x)v) ,$$

where the map $g_{\alpha\beta} : U_{\alpha\beta} \to GL(m, \mathbb{K})$ is called a *transition function*. With respect to a local frame $\{e_1^\alpha, \ldots, e_n^\alpha\}$ over U_α a section s can be written as

$$s = \sum_a s_\alpha^a e_a^\alpha ,$$

where $s_\alpha^a : U_\alpha \to \mathbb{K}$ are the *component functions* of the section s. If $\{e_1^\beta, \ldots, e_n^\beta\}$ is any other frame over U_β such that $U_{\alpha\beta} \neq \emptyset$, then it holds that

$$s = \sum_a s_\alpha^a e_a^\alpha = \sum_b s_\beta^b e_b^\beta .$$

Therefore, the component functions s_α^a and s_β^b are related by the transition function $g_{\alpha\beta}$ as

$$s_\beta^b = \sum_a (g_{\alpha\beta})_a^b s_\alpha^a . \tag{3.1}$$

Conversely, if we are given functions $\{s_\alpha^a : a = 1, \ldots, m\}$ for each U_α of the open cover, then these are the components of a section of E if they satisfy Eq. (3.1).

We close this section with some more examples of vector bundles. Recall that the complex projective plane $\mathbb{C}P^1$ is defined as the space of all complex lines through the origin in $\mathbb{C}^2$. The *tautological line bundle* $\tau \to \mathbb{C}P^1$ is defined as

$$\tau = \{(\ell, v) \in \mathbb{C}P^1 \times \mathbb{C}^2 : v \in \ell\} .$$

In other words, the fiber over the line $\ell \in \mathbb{C}P^1$ is taken to be precisely the line ℓ, viewed as a subspace of $\mathbb{C}^2$. The tautological line bundle plays a central role in the classification of topological insulators as we will see in Section 5. All our examples so far have been of line bundles. Examples of higher rank real vector bundles of course also exist. Some notable examples are the tangent bundle TX and the cotangent bundle T^*X of a smooth manifold X. The rank of TX and T^*X is equal to the dimension of X. Sections of the tangent bundle are vector fields and sections of the cotangent bundle are covector fields, also called differential one-forms.

3.2 DeRham cohomology

The deRham cohomology of a smooth manifold recovers topological information about the manifold from its differential forms. We first illustrate by example the mechanism

the deRham cohomology uses to obtain topological information from differential objects. After the example, the technical definition is discussed.

Consider a rotation free vector field $\boldsymbol{E}$ on $\mathbb{R}^2$. It is a well-known fact that we can write $\boldsymbol{E}$ as the gradient of some scalar potential V. To see this, fix a reference point $\boldsymbol{x}_0 \in \mathbb{R}^2$ at which we set $V(\boldsymbol{x}_0) = 0$, and then define the potential at any other point $\boldsymbol{x} \in \mathbb{R}^2$ to be

$$V(\boldsymbol{x}) = -\int_{\boldsymbol{x}_0}^{\boldsymbol{x}} d\boldsymbol{l} \cdot \boldsymbol{E}\,,$$

where the integral is taken over the straight line segment from $\boldsymbol{x}_0$ to $\boldsymbol{x}$. Using $\frac{\partial E_x}{\partial y} = \frac{\partial E_y}{\partial x}$ one can check that $\boldsymbol{E} = -\boldsymbol{\nabla} V$.

However, on $\mathbb{R}^2 \setminus \{0\}$ some rotation free vector fields *cannot* be written as the gradient of a potential. Indeed, consider the vector field $\boldsymbol{E}$ on $\mathbb{R}^2 \setminus \{0\}$ given by

$$\boldsymbol{E}(x, y) = \frac{-y}{\sqrt{x^2 + y^2}}\hat{\boldsymbol{x}} + \frac{x}{\sqrt{x^2 + y^2}}\hat{\boldsymbol{y}}\,.$$

It is easy to check that $\boldsymbol{E}$ is rotation free by computing $\boldsymbol{\nabla} \times \boldsymbol{E} = 0$. Suppose then that there is a potential $V : \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}$ such that $\boldsymbol{E} = -\boldsymbol{\nabla} V$. In polar coordinates $\boldsymbol{E} = \hat{\boldsymbol{\theta}}$. Recall that the gradient in polar coordinates is given by

$$\boldsymbol{\nabla} V = \partial_r V\, \hat{\boldsymbol{r}} + \frac{1}{r}\partial_\theta V\, \hat{\boldsymbol{\theta}}\,.$$

The equation $\boldsymbol{E} = -\boldsymbol{\nabla} V$ requires $\partial_r V = 0$ and $\partial_\theta V = -r$. The latter implies $V(r, \theta) = -r\theta + c(r)$, where $c(r)$ is an integration constant that may still depend on r, but not on θ. The equation $\partial_r V = 0$, however, implies that $c'(r) = \theta$. This is a contradiction. Therefore, the vector field $\boldsymbol{E}$, although it is rotation free, is not the gradient of a scalar potential.

What this example shows is that puncturing $\mathbb{R}^2$ at the origin destroys the property that "rotation free vector fields are the gradients of scalar potentials". The difference between the plane $\mathbb{R}^2$ and the punctured plane $\mathbb{R}^2 \setminus \{0\}$ is purely topological. Therefore, the property that "rotation free vector fields are the gradients of scalar potentials" must be a topological property.

The deRham cohomology of a manifold in some way measures the existence of holes in the manifold by asking whether every closed differential form has a primitive. In the example above, rotation free vector fields correspond to closed differential forms and the potential would be its primitive. The precise construction is as follows. Let $\Omega^k(X)$ denote the real valued differential k-forms on X. The deRham complex is obtained by arranging the differential forms in a cochain complex

$$0 \longrightarrow \Omega^0(X) \xrightarrow{d} \Omega^1(X) \xrightarrow{d} \Omega^2(X) \longrightarrow \dots\,,$$

where the exterior derivative d maps from k-forms to $k+1$-forms and $d^2 = 0$.[6] A form $\omega \in \Omega^k(X)$ is called *closed* if $d\omega = 0$ and it is called *exact* if there is $\eta \in \Omega^{k-1}(X)$ such that $\omega = d\eta$. The relation $d^2 = 0$ implies that every exact form is closed. Hence, $\mathrm{im}(d : \Omega^{k-1}(X) \to \Omega^k(X))$ is a subvector space of $\ker(d : \Omega^k(X) \to \Omega^{k+1}(X))$. We define the *degree k deRham cohomology group of X with coefficients in* $\mathbb{R}$ to be the vector space quotient

$$H^k_{dR}(X; \mathbb{R}) = \frac{\ker(d : \Omega^k(X) \to \Omega^{k+1}(X))}{\mathrm{im}(d : \Omega^{k-1}(X) \to \Omega^k(X))}\,.$$

[6] The reason we call this a complex is because $d^2 = 0$. It is a cochain complex because d raises degrees by one. A chain complex would be a sequence of spaces and maps that square to zero, but that lower the degree by one.

In the example above, the plane and the punctured plane have different degree one deRham cohomology groups: $H^1_{dR}(\mathbb{R}^2;\mathbb{R}) = 0$, while $H^1_{dR}(\mathbb{R}^2 \setminus \{0\};\mathbb{R}) \neq 0$. To compute degree one deRham cohomology, we compared the closed one-forms, i.e. rotation free vector fields, to the exact one forms, i.e. those vector fields that are the gradient of some potential. On $\mathbb{R}^2$, these are the same things, hence $H^1_{dR}(\mathbb{R}^2;\mathbb{R}) = 0$. However, on $\mathbb{R}^2 \setminus \{0\}$, there are rotation free vector fields that are not the gradient of a potential, hence $H^1_{dR}(\mathbb{R}^2 \setminus \{0\};\mathbb{R}) \neq 0$.

The deRham cohomology can thus measure topological features of the underlying manifold, but it is by no means clear from the definition that deRham cohomology is a topological invariant, in the sense that it is the same for homotopy equivalent manifolds. This is the content of the deRham Theorem, stated for example in [27]. The deRham Theorem states that deRham cohomology is the same as so-called singular cohomology. Singular cohomology is defined for any topological space, independent of any smooth structures or differential forms. The equivalence of deRham and singular cohomology thus means that, although the construction of the deRham cohomology heavily uses the smooth structure, the result is only dependent on the underlying topology.

3.3 Connections and curvature

For a real valued function on a smooth manifold there is a canonical notion of "directional derivative" in terms of the limit of a difference quotient. However, for sections of a vector bundle no such canonical notion of directional derivative exists. The problem is that one cannot subtract two vectors that live in fibers over different points, because even though each fiber is isomorphic to $\mathbb{R}^n$ or $\mathbb{C}^n$, this isomorphism is not canonical; it requires the choice of a trivialization. A connection is a choice of directional derivative for the sections of a vector bundle that is independent of the choice of trivialization. Nevertheless, connections themselves are not unique. A vector bundle admits many connections, all giving a different, but well-defined notion of directional derivative.

Let E be a real vector bundle over X, then a *connection* ∇ on E is a map

$$\nabla : \Gamma(TX) \times \Gamma(E) \to \Gamma(E), \quad (V, s) \mapsto \nabla_V(s)$$

that is $C^\infty(X)$-linear in $\Gamma(TX)$, $\mathbb{R}$-linear in $\Gamma(E)$ and satisfies the Leibniz rule: if $f \in C^\infty(X)$, $V \in \Gamma(TX)$ and $s \in \Gamma(E)$, then

$$\nabla_V(fs) = df(V) + f\nabla_V(s)\,.$$

The derivational nature of the connection is precisely encoded in the Leibniz rule. It is easy to see that if ∇ and ∇' are two connections on E, then the difference $\nabla - \nabla'$ is $C^\infty(X)$-linear in $\Gamma(E)$ since the correction terms coming from the Leibniz rule cancel each other. The difference between two connections can thus be interpreted as a tensor, and more precisely, as a so-called endomorphism valued one-form.

Let us briefly explain this terminology. An *endomorphism* is a map from a space to itself that preserves the relevant structures. In the case of vector spaces, an endomorphism is simply a linear map from a vector space to itself. We denote the space of endomorphisms of a vector space W by $\operatorname{End} W$. The space of endomorphism of a vector space is again a vector space, since one can add linear maps and multiply them by scalars. Given a vector bundle $E \to X$ it thus make sense to define the *endomorphism bundle* $\operatorname{End} E \to X$ to be the vector bundle whose fiber over a point $p \in X$ is $\operatorname{End} E_p$. Sections of the endomorphism bundle correspond to $C^\infty(X)$-linear maps $\Gamma(E) \to \Gamma(E)$. An *endomorphism valued one-form* is a $C^\infty(X)$-linear map from vector fields to sections of the endomorphism bundle.

We denote the endomorpism valued one-forms by $\Omega^1(X, \operatorname{End} E)$. The difference between two connections is in $\Omega^1(X, \operatorname{End} E)$ because plugging in a vector field V into $\nabla - \nabla'$ yields a map $\nabla_V - \nabla'_V : \Gamma(E) \to \Gamma(E)$ that is $C^\infty(X)$-linear. In fact, it is true that given a fixed connection ∇ any other connection ∇' can be obtained as $\nabla' = \nabla + A$, where $A \in \Omega^1(X, \operatorname{End} E)$ is an endomorphism valued one-form.

Endomorphism valued one-forms also come up when considering the local representation of a single connection. Consider a general vector bundle E with connection ∇ and let $\{U_\alpha\}$ be a trivializing open cover with local frames $\{e_a^\alpha\}$. The connection ∇ acts on a section e_a^α to yield a new section over U_α, which can be expressed as

$$\nabla(e_a^\alpha) = \sum_b (A_\alpha)_a^b e_b^\alpha \,, \tag{3.2}$$

where A_α is an endomorphism valued one-form, which due to the choice of basis might as well be called a *matrix valued one-form*. For a general section $s = \sum_a s_\alpha^a e_a^\alpha$ over U_α we use the Leibniz rule and Eq. (3.2) to see that the connection ∇ acts on s according to

$$\nabla(s) = \nabla\left(\sum_a s_\alpha^a e_a^\alpha\right) = \sum_b ds_\alpha^b e_b^\alpha + \sum_a s_\alpha^a \nabla(e_a^\alpha) = \sum_b \left(ds_\alpha^b + \sum_a (A_\alpha)_a^b s_\alpha^a\right) e_b^\alpha \,.$$

The connection ∇ thus acts as $d + A_\alpha$ on the local coefficient functions. We call A_α a *local connection one-form*. If $U_{\alpha\beta} \neq \emptyset$, then the local connection one-forms A_α and A_β are related through the transition function $g_{\alpha\beta}$ by

$$A_\beta = g_{\alpha\beta} A_\alpha g_{\alpha\beta}^{-1} + g_{\alpha\beta} dg_{\alpha\beta}^{-1} \,. \tag{3.3}$$

Conversely, any collection A_α of matrix valued one-forms related by Eq. (3.3) defines a connection on E that locally acts as $d + A_\alpha$.

In the above discussion on the local representation of a connection we did not explicitly plug in any vector fields into the connection. The point of view implicit when doing this is that a connection $\nabla : \Gamma(TX) \times \Gamma(E) \to \Gamma(E)$, by defining $\Omega^1(X; E) := \Gamma(T^*X \otimes E)$, can be interpreted as a map

$$\nabla : \Gamma(E) \to \Omega^1(X; E)$$

that satisfies the Leibniz rule $\nabla(fs) = df \otimes s + f\nabla(s)$. The elements of $\Omega^1(X; E)$ are called vector valued one-forms. In general, $\Omega^n(X; E) := \Gamma(\bigwedge^n T^*X \otimes E)$ is the space of vector valued n-forms. An analogue of the deRham complex can be defined using vector valued forms. The differential $\nabla : \Omega^n(X; E) \to \Omega^{n+1}(X; E)$ in degree n can be defined using the Koszul formula [26]. A crucial difference between the ordinary deRham complex of real valued form and the generalized deRham complex of vector valued forms is that in the deRham complex $d^2 = 0$, but $\nabla^2 \neq 0$ in the generalized setting. In fact, the failure of ∇ to square to zero is what we define to be the *curvature*, $F = \nabla^2$. If the connection is locally represented as $d + A_\alpha$ then the curvature is locally represented as the matrix valued two-form F_α that acts on a section s of E as

$$\begin{aligned} F_\alpha(s) &= (d + A_\alpha)(d + A_\alpha)(s) \\ &= d^2 s + A_\alpha \wedge ds + d(A_\alpha s) + (A_\alpha \wedge A_\alpha)(s) \\ &= (dA_\alpha + A_\alpha \wedge A_\alpha)(s) \,, \end{aligned}$$

where we have used that $d^2 = 0$ and $d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^{\deg \omega} \omega \wedge d\eta$. The above computation recovers the familiar local formula $F_\alpha = dA_\alpha + A_\alpha \wedge A_\alpha$ for the curvature.

Using Eq. (3.3) one can check that the local curvature forms over U_α and U_β are related by the transition function $g_{\alpha\beta}$ through

$$F_\beta = g_{\alpha\beta} F_\alpha g_{\alpha\beta}^{-1} \,. \tag{3.4}$$

Note that the local curvature two-forms transform as sections of the endomorphism bundle $\operatorname{End} E$ so that the local curvature two-forms define a global object $F \in \Omega^2(X, \operatorname{End} E)$.

3.4 Chern classes

With the preliminaries out of the way we can now introduce the Chern classes. Chern classes are topological invariants of a complex vector bundles that measure the nontriviality of the vector bundle. For a fixed smooth manifold X, the Chern classes are natural maps from the isomorphism classes of complex vector bundles over X into the deRham cohomology of X. We provide a construction of the Chern classes that uses the curvature two-form F reviewed above.

Fundamental to the construction of the Chern classes from the curvature two-form are so-called invariant polynomials. An *invariant polynomial* P is a polynomial function $P : \operatorname{Mat}(m, \mathbb{C}) \to \mathbb{C}$ such that $P(gMg^{-1}) = P(M)$ for any $M \in \operatorname{Mat}(m, \mathbb{C})$ and $g \in GL(m, \mathbb{C})$. Two common examples of invariant polynomials are the trace and the determinant of a matrix.

Let $E \to X$ be a vector bundle of rank m with a connection ∇. We have seen that the local curvature two-form F_α changes as $F_\alpha \to F_\beta = g_{\alpha\beta} F_\alpha g_{\alpha\beta}^{-1}$ under a change of trivialization by the transition function $g_{\alpha\beta} : U_{\alpha\beta} \to GL(m, \mathbb{C})$. Hence, if P is an invariant polynomial, then $P(F_\alpha)$ is actually independent of any choice of trivialization over U_α. The collection $\{P(F_\alpha)\}_{\alpha \in \mathcal{A}}$ thus defines a global, complex valued two-form $P(F) \in \Omega^2(X, \mathbb{C})$. The complex valued forms are defined as $\Omega^k(X; \mathbb{C}) := \Omega^k(X) \otimes \mathbb{C}$. Concretely, a complex valued form is of the type $\omega + i\eta$, where both ω and η are real forms. The key result in the construction of the Chern classes as invariant polynomials of the curvature is the Chern-Weil theorem.

Theorem 3.1 (Chern-Weil)**.** If $F \in \Omega^2(X; \mathbb{C})$ is the curvature of a connection ∇ on a complex vector bundle E, then for any invariant polynomial P the following hold:

1. $dP(F) = 0$,
2. if F' is the curvature corresponding to some other connection ∇' on E, then $P(F) - P(F') \in H^*_{dR}(X; \mathbb{C})$ is exact.

The first part of the Chern-Weil Theorem implies that $P(F)$ defines a cohomology class in $H^*_{dR}(X; \mathbb{C})$ and the second part shows that this cohomology class is independent of the choice of connection. The independence of the choice of connection means that the cohomology class $P(F) \in H^*_{dR}(X; \mathbb{C})$ is a topological invariant of the vector bundle E.

Suppose now that E is a rank m complex vector bundle and fix a connection ∇ on E. The *total Chern class* $c(E)$ is defined to be the deRham cohomology class of the invariant polynomial $M \mapsto \det(I + \frac{i}{2\pi} M)$,

$$c(E) = \left[\det\left(I + \frac{i}{2\pi} F \right) \right] ,$$

where F is the curvature. Since F is of degree 2, the total Chern class can be written as

$$c(E) = I + c_1(E) + \cdots + c_m(E) \,,$$

where each $c_i(E)$ is a degree $2i$ cohomology class. The class $c_i(E)$ called the *i-th Chern class*. Note that since the determinant of an $m \times m$ matrix is a degree m polynomial, $c_i(E) = 0$ for $i > m$. Explicit formulas for the Chern classes can be obtained by first diagonalizing the curvature and then expanding the determinant, see [24] for details. The results for the first and second Chern classes are

$$c_1(E) = \frac{i}{2\pi}\,\mathrm{tr}[F]$$
$$c_2(E) = \frac{1}{2}\left(\frac{i}{2\pi}\right)^2 \left(\mathrm{tr}[F]^2 - \mathrm{tr}[F^2]\right) .$$

The above definition of the Chern classes establishes them as maps from the isomorphism classes of vector bundles to the cohomology ring $H^*_{dR}(X;\mathbb{C})$. That the Chern classes only depend on the isomorphism type of a vector bundle can be seen as follows. The claim is that if $\varphi : E \to E'$ is an isomorphism between vector bundles E and E' over X, then $c(E) = c(E')$. Indeed, one can always find an open cover $\{U_\alpha\}_{\alpha\in\mathcal{A}}$ over which E and E' can be simultaneously trivialized. If A_α are the local connection one-forms of a connection ∇ on E, then the local one-forms $A'_\alpha := \varphi_\alpha A_\alpha \varphi_\alpha^{-1} + \varphi_\alpha d\varphi_\alpha^{-1}$, where $\varphi_\alpha : U_\alpha \to GL(m,\mathbb{C})$ is the representation of the isomorphism with respect the trivialization, define a connection on E'. The curvature of this connection is determined by the local curvuture forms $F'_\alpha = \varphi_\alpha F_\alpha \varphi_\alpha^{-1}$. Hence,

$$c(E') = \left[\det\left(I + \frac{i}{2\pi}F'\right)\right] = \left[\det\left(I + \frac{i}{2\pi}F\right)\right] = c(E) .$$

The Chern classes are also *natural* in the sense that if $f : X' \to X$ is a map between smooth manifolds and $E \to X$ is a complex vector bundle, then $f^*(c(E)) = c(f^*E)$. This follows from the following observation. Let A_α be the local connection one-forms of a connection ∇ on E with respect to some trivialization over $\{U_\alpha\}_{\alpha\in\mathcal{A}}$. The pullbacks of the trivializations over $\{U_\alpha\}$ of E along f give a trivialization over $\{f^{-1}(U_\alpha)\}$ of f^*E. Moreover, the pullbacks f^*A_α of the local connection one forms of the connection ∇ define a connection ∇' on f^*E, which is called the pullback connection. The curvature of this pullback connection is the pullback of the curvature. Indeed, since $F_\alpha = dA_\alpha + A_\alpha \wedge A_\alpha$ and the pullback commutes with d and respects the wedge product, it follows that

$$F'_\alpha = d(f^*A_\alpha) + f^*A_\alpha \wedge f^*A_\alpha = f^*(dA_\alpha + A_\alpha \wedge A_\alpha) = f^*F_\alpha .$$

From this the naturality of the Chern classes follows immediately.

Using the naturality property of the Chern classes it can be shown that the Chern classes of a trivial vector bundle vanish. Indeed, if $E \to X$ is a trivial complex vector bundle of rank m, then $E \cong f^*(\{p_0\} \times \mathbb{C}^m)$, where $f : X \to p_0$ is the constant map to some point $p_0 \in X$. The isomorphism can be established by picking a global trivialization $\{e_1, \ldots, e_m\}$ for E, and then defining $\varphi : E \to f^*(\{p_0\} \times \mathbb{C}^m)$ by

$$\sum_a v^a e_a(p) \mapsto (v^1, \ldots, v^m) .$$

This is an isomorphism of vector bundles because it is a smooth map that is a linear isomorphism $E_p \to \mathbb{C}^m$ on each fiber. By the naturality property of the Chern classes, $c_i(E) = f^*(c_i(\{p_0\} \times \mathbb{C}^n)) = 0$ for $i > 0$, since all deRham cohomology groups of the point vanish in positive degree. We conclude that for a trivial vector bundle all the Chern classes vanish. The Chern classes can thus be used to determine the non-triviality of a vector bundle.

For example, it can be shown that the first Chern class of the tautological line bundle $\tau \to \mathbb{C}P^1$ is nonzero. Therefore, the tautological line bundle is a nontrivial bundle. In fact, in proving this one finds that

$$\int_{\mathbb{C}P^1} c_1(\tau) = \pm 1\,, \tag{3.5}$$

where the sign of the answer depends on the choice of orientation on $\mathbb{C}P^1$. In either case, $\int_{\mathbb{C}P^1} c_1(\tau)$ is an integer. The fact that this integral is an integer is not a coincidence. We will revisit the Chern classes in Section 5 and show that for vector bundles over a two-dimensional base space, the integral of the first Chern class is always an integer.

4 Berry curvature, valence bundle and Hall conductivity

As motivated in Section 2, the low-energy physics of a d-dimensional insulator can be modelled by a tight binding Hamiltonian $h : X \to \mathrm{Herm}(M, \mathbb{C})$ that satisfies the gap condition. The effective momentum space manifold X is either a torus, called the Brillouin torus, or a sphere, depending on whether we consider the continuum limit or not. The gap condition entails that the eigenvalues of $h(p)$ never vanish as a function of $p \in X$. As we will see in this section, the eigenvectors corresponding to the negative eigenvalues of h span a possibly non-trivial subvector bundle of $X \times \mathbb{C}^M$ which refer to as the *valence bundle* of the insulator. The valence bundle naturally comes equipped with the so-called *Berry connection*. It was realized back in the eighties by Thouless, Kohmoto, Nightingale and Den Nijs [6] that the Berry connection has physically measurable implications. They found that the integral of the curvature of the Berry connection over the Brillouin torus is equal to the Hall conductivity of a two-dimensional insulator, implying that the Hall conductivity is quantized.

In this section, we will first introduce the adiabatic connection in the general context of parameterized quantum systems due to Kato [28], which was later realized by Simon [29] to be exactly the same thing as the Berry connection [30]. Secondly, we construct the valence bundle of an insulator and show how the construction of the adiabatic/Berry connection applies to this. Finally, we show by a path integral computation that the Hall conductivity of a two-dimensional insulator is the integral of the Berry curvature over the momentum space X, reproducing the result of [6]. In fact, in Section 5, we will mathematically prove that the integral of the Berry curvature over X is always an integer, which then establishes that the Hall conductivity is indeed quantized.

4.1 The adiabatic connection

Consider a quantum mechanical system consisting of a separable Hilbert space $\mathcal{H}$ and a manifold X serving as a space of control parameters. Suppose $H : X \to \mathrm{Herm}(\mathcal{H})$ is a smooth map that assigns to each $p \in X$ a Hamiltonian $H(p)$. Assume there is a smooth function $\epsilon : X \to \mathbb{R}$ such that for each $p \in X$, $\epsilon(p)$ is an eigenvalue of $H(p)$ that is separated from the rest of the spectrum of $H(p)$ by a finite gap. More precisely,

$$\Delta := \inf_{p \in X} \mathrm{dist}(\epsilon(p), \sigma(H(p)) \setminus \{\epsilon(p)\}) > 0\,, \tag{4.1}$$

where $\sigma(H(p))$ is the spectrum of $H(p)$. Eq. (4.1) is referred to in this context as the gap condition. At a point $p \in X$ the spectral projector $P(p)$ onto the eigenspace of $\epsilon(p)$ may be expressed as

$$P(p) = \frac{1}{2\pi i} \int_{\Gamma_p} [H(p) - z]^{-1}\,,$$

where Γ_p is a curve encircling $\epsilon(p)$ once in the counterclockwise direction and whose interior contains no points of the spectrum except for $\epsilon(p)$. Because of the gap condition, the curves Γ_p can be chosen to vary smoothly with respect to changes in p.

Let $p_0 \in X$ be a fixed initial parameter configuration and assume the system is prepared in an eigenstate ψ_0 of $\epsilon(p_0)$. Let $\gamma : [0,1] \to X$ be a smooth curve in parameter space starting at p_0. We want to study the evolution of a state ψ_0 when the parameters of the system are slowly changed along the curve γ. Denote by $s \in [0,1]$ the parameter of the curve γ and introduce the short hand notations $H(s) := H(\gamma(s))$ and $P(s) := P(\gamma(s))$. We consider traversing the path γ in a large, physical time T, which will be referred to as the *adiabatic time scale*, and define the *physical time* $t \in [0,T]$ through $t = sT$. In terms of the physical time t, the Schrödinger equation is

$$i\hbar \frac{d}{dt}\psi(t) = H(t/T)\psi(t)\,, \tag{4.2}$$

In terms of the parameter $s \in [0,1]$ the Schrödinger equation thus becomes

$$i\hbar \frac{d}{ds}\psi_T(s) = TH(s)\psi_T(s)\,, \tag{4.3}$$

where $\psi_T(s) := \psi(sT)$. Changing our point of view from the Schrödinger picture to the Heisenberg picture we can define the *physical time evolution operator* $U_T(s)$ as the solution to

$$i\hbar \frac{d}{ds}U_T(s) = TH(s)U_T(s)\,. \tag{4.4}$$

For finite T the physical evolution $U_T(s)$ does not stay within the eigenspace of $\epsilon(s)$. More precisely, if at $t = 0$ the system is prepared in an eigenstate ψ_0 with energy $\epsilon(0)$, then at $t = sT$, the state $U_T(s)\psi_0$ might not be an eigenstate with energy $\epsilon(s)$. Physically speaking, the finiteness of the time interval in which the parameters are changed allows the change in parameters to induce excitations of the system out of the gapped states corresponding to $\epsilon(s)$. It turns out, however, that $\lim_{T\to\infty} U_T(s)$ does preserve the eigenspace of $\epsilon(s)$. We will show this by proving that in the limit $T \to \infty$ the physical time evolution $U_T(s)P(0)$ of the system starting in a state with energy $\epsilon(0)$ tends to the so-called *adiabatic time evolution* $U_A(s)P(0)$. The adiabatic evolution $U_A(s)$ is constructed precisely so that $P(s)U_A(s) = U_A(s)P(0)$, i.e. so that it preserves the eigenspace of $\epsilon(s)$. The original construction of the adiabatic evolution operator is due to Kato [28]. We follow the construction as described by Simon in [31].

The fundamental insight of Kato is the following Theorem.

Theorem 4.1 (Kato dynamics)**.** If $U_A : [0,1] \to L(\mathcal{H})$ is a solution to the initial value problem

$$\frac{d}{ds}U_A(s) = [\dot{P}(s), P(s)]U_A(s), \text{ subject to } U_A(0) = 1\,,$$

then $U_A(s)$ is unitary and $U_A(s)P(0) = P(s)U_A(s)$ for all $s \in [0,1]$.

Proof. Let $A(s) := [\dot{P}(s), P(s)]$ and assume $U_A(s)$ solves $\frac{d}{ds}U_A(s) = A(s)U_A(s)$ with $U_A(0) = 1$. The fact that $U_A(s)$ is unitary follows from standard existence and uniqueness theorem for ordinary differential equations in Banach spaces and the anti-Hermiticity of $A(s)$, see [31, Proposition 17.1]. The interesting part of the theorem is to prove

$$U_A(s)P(0) = P(s)U_A(s)\,. \tag{4.5}$$

This is done by showing that the left hand side and the right hand of Eq. (4.5) side solve the same initial value problem. The equality then follows from uniqueness of solutions to ordinary differential equations.

Since $U_A(0) = 1$ the left and right hand sides of Eq. (4.5) agree at $s = 0$. Furthermore, $W(s) := U_A(s)P(0)$ clearly solves

$$\frac{d}{ds}W(s) = A(s)W(s)\,.$$

The claim is that $P(s)U_A(s)$ is also a solution. Note that $P^2(s) = P(s)$ implies $\dot{P}(s) = \dot{P}(s)P(s) + P(s)\dot{P}(s)$, which in turn implies $P(s)\dot{P}(s)P(s) = 0$. Using these facts, we compute

$$\begin{aligned}\frac{d}{ds}(P(s)U_A(s)) &= \dot{P}(s)U_A(s) + P(s)\frac{d}{ds}U_A(s)\\ &= \dot{P}(s)U_A(s) + P(s)[\dot{P}(s), P(s)]U_A(s)\\ &= (\dot{P}(s)P(s) + P(s)\dot{P}(s))U_A(s) - P(s)\dot{P}(s)U_A(s)\\ &= [\dot{P}(s), P(s)]P(s)U_A(s)\,.\end{aligned}$$

We conclude that $U_A(s)P(0)$ and $P(s)U_A(s)$ both solve $\frac{d}{ds}W(s) = [\dot{P}(s), P(s)]W(s)$ with the same initial value, so $U_A(s)P(0) = P(s)U_A(s)$ for all $s \in [0, 1]$. □

By adding the term $\frac{i\hbar}{T}[\dot{P}(s), P(s)]$ from the Kato dynamics, the Hamiltonian $H(s)$ can be augmented to the *adiabatic Hamiltonian*

$$H_A(s) = H(s) + \frac{i\hbar}{T}[\dot{P}(s), P(s)]\,.$$

The time evolution generated by H_A is the *adiabatic time evolution*. If $U_A(s)$ is as in Theorem 4.1, then it can be shown that (see the proof of [31, Theorem 17.2])

$$||e^{-iT\int_0^s \epsilon(s)ds}U_A(s)P(0) - U_T(s)P(0)|| = O(1/T)\,. \tag{4.6}$$

The first term is the time evolution generated by the adiabatic Hamiltonian when starting in the eigenspace of $\epsilon(0)$. By the fact that $U_A(s)P(0) = P(s)U_A(s)$ it follows that this evolution preserves the eigenspaces of $\epsilon(s)$. The second term is the physical time evolution on the eigenspace of $\epsilon(0)$ generated by the usual Hamiltonian. Therefore, Eq. (4.6) implies that in the adiabatic limit $T \to \infty$, the physical evolution agrees with the adiabatic evolution on the eigenspaces of $\epsilon(s)$.

Let us now assume that $\epsilon \equiv 0$. This assumption gets rid of the standard dynamical phase factor $e^{-iT\int_0^s \epsilon(s)ds}$ in the adiabatic evolution and leaves only contributions due to U_A. If we let the system evolve adiabatically along a loop γ based at p_0, then the system eventually returns to the same eigenspace as that it started in. If the eigenspace of $\epsilon(p_0)$ is of dimension one and if $\psi_0 \in \text{im } P(p_0)$, then one would expect that after traversing the loop in a very large time T, the vector ψ_0 returns to itself, up to $O(1/T)$ terms. It was realized by Berry that this is not the full story. Under adiabatic evolution, the system picks up an additional phase which is purely geometrical. This is *Berry's phase*.

The mathematical interpretation of the Berry phase is as the holonomy of the so-called adiabatic connection (or Berry connection) in the bundle of eigenstates of $\epsilon(p)$. Consider the trivial bundle $X \times \mathcal{H}$ and define the subbundle

$$E = \{(p, \psi) \in X \times \mathcal{H} : P(p)\psi = \psi\} \tag{4.7}$$

of the eigenspaces of $\epsilon(p)$. A connection on E can be defined in such a way that parallel sections correspond to adiabatic evolutions. The connection that does the job is the *adiabatic connection*

$$\nabla = d - [dP, P], \tag{4.8}$$

where d is defined to take the exterior derivative of the components of sections with respect to some fixed basis of $\mathcal{H}$ and $[dP, P]$ is an $L(\mathcal{H})$-valued one-form on X. It is clear that ∇ satisfies the appropriate linearity conditions and the Leibniz rule making it at least a connection on $X \times \mathcal{H}$. What is not clear, is that it actually restricts to a connection on E. Let ψ be a section of E, then $P(p)\psi(p) = \psi(p)$ for all $p \in X$. Using $P\psi = \psi$ and $PdPP = 0$ we compute

$$P(\nabla\psi) = P(d\psi - [dP, P]\psi) = Pd\psi - P[dP, P]P\psi = Pd\psi$$

and

$$\nabla\psi = d\psi - [dP, P]\psi = d(P\psi) - (dP)\psi = Pd\psi,$$

where we have used the product rule to see that $d(P\psi) = (dP)\psi + Pd\psi$. Hence, $P(\nabla\psi) = \nabla\psi$ for $\psi \in \Gamma(E)$, so that ∇ indeed takes sections of E to sections of E. Moreover, we learn that the connection ∇ is really just Pd acting on sections of E.

The relation between parallel sections and adiabatic evolution is as follows. If $\gamma : [0, 1] \to X$ is a curve, then a section $\psi : X \to E$ is parallel along γ if

$$\nabla_{\dot\gamma(s)}(\psi)(s) = 0 \text{ for all } s \in [0, 1].$$

Writing out the definition of ∇, this is equivalent to

$$\nabla_{\dot\gamma(s)}(\psi)(s) = d\psi_{\gamma(s)}(\dot\gamma(s)) - [dP_{\gamma(s)}(\dot\gamma(s)), P(\gamma(s))]\psi(\gamma(s)).$$

In our shorthand notation $P(s) = P(\gamma(s))$, the condition to be parallel thus becomes

$$\frac{d}{ds}\psi(s) - [\dot P(s), P(s)]\psi(s) = 0,$$

which is precisely the adiabatic evolution of $\psi(0)$. In conclusion, adiabatic evolution is the same thing as parallel transport with respect to $\nabla = d - [dP, P]$.

We can also compute the curvature of the adiabatic connection. The curvature associated to the connection ∇ is $F = \nabla^2$. If $P\psi = \psi$, then we compute using $\nabla = Pd$ that

$$F\psi = \nabla^2\psi = (Pd)(Pd\psi) = PdP \wedge d\psi = PdP \wedge d(P\psi) = PdP \wedge dP\,\psi = PdP \wedge dPP\psi,$$

where we judiciously use the identities $dP = PdP + dPP$ and $PdPP = 0$. It follows that the curvature two-form corresponding to the adiabatic connection is

$$F = P(dP \wedge dP)P. \tag{4.9}$$

This expression is globally valid and does not depend on any choice of trivializations. As a final comment, it is standard terminology to refer to the adiabatic connection and curvature as the Berry connection and curvature. From now on, we adopt the standard terminology.

4.2 The valence bundle

Consider now an insulator modelled by a tight binding Hamiltonian $h : X \to \mathrm{Herm}(M, \mathbb{C})$, where M is the number of bands around the Fermi level that we take into account and the manifold X is a torus or a sphere depending on whether we consider a periodic or a continuum model. We can interpret X as a parameter manifold so that h corresponds to a family of Hamiltonians for a parameterized quantum system modelled by the trivial bundle $X \times \mathbb{C}^M$. The characteristic equation $\det(h - \epsilon) = 0$ may be solved to obtain M smooth functions $\epsilon_{\tilde{a}} : X \to \mathbb{R}$ for $\tilde{a} = 1, \ldots, M$, which at each $p \in X$ are the eigenvalues of $h(p)$. Suppose that the first m functions ϵ_a for $a = 1, \ldots, m$ are the negative energy solutions that correspond to the valence bands in the given model.

The eigenvectors corresponding to the negative energy eigenvalues span a possibly non-trivial subbundle of the trivial bundle $X \times \mathbb{C}^M$ called the valence bundle. Recall that $P(p) = \frac{1}{2\pi i} \int_C dz[h(p) - z]^{-1}$, where C is a curve in $\mathbb{C}$ enclosing the negative real axis once in the counterclockwise direction, denotes the spectral projection onto the negative energy eigenspaces. The *valence bundle* is defined to be

$$E_{val} = \{(p, \psi) \in X \times \mathbb{C}^M : P(p)\psi = \psi\}. \tag{4.10}$$

Note the similarity between the valence bundle and the subbundle of $X \times \mathcal{H}$ defined in Eq. (4.7). An equivalent definition of the valence bundle is that $E_{val} = \mathrm{im}\, P$, where P is viewed is a bundle map $P : X \times \mathbb{C}^M \to X \times \mathbb{C}^M$. Completely analogous to the constructions below Eq. (4.7), we can define the *Berry connection* $\nabla = Pd$ on the valence bundle E_{val}. By exactly the same computations as before, the curvature of the Berry connection, i.e. the *Berry curvature*, is given by $F = PdP \wedge dPP$.

The above expressions for the Berry connection and Berry curvature in terms of the abstract spectral projector P have the advantage that they are well-defined independent of any choice of trivializations or frames. However, quantum mechanical computations are often made more concrete using an explicit choice of "basis of wavefunctions", i.e. a frame. It is thus desirable to also obtain the local connection one-forms and curvature two-forms of the Berry connection and curvature with respect to a choice of local frame of eigenvectors.

Let $\psi_1, \ldots, \psi_m$ be a local orthonormal frame for E_{val} over an open set $U \subset X$, where orthonormality is required with respect to the inner product on the fibers of E_{val} induced from the standard inner product on $\mathbb{C}^M$. The spectral projector $P|_U$ onto the valence states over U can then be written as

$$P|_U = \sum_{a=1}^{m} \langle \psi_a, \cdot \rangle \psi_a .$$

The local connection one-form A_U of the Berry connection $\nabla = Pd$, with d the deRham derivative of components with respect to the standard global frame $e_1, \ldots, e_M$ of $X \times \mathbb{C}^M$, can be deduced from Eq. (3.2). We have

$$\nabla \psi_b = P|_U d\psi_b = \sum_{a=1}^{m} \langle \psi_a, d\psi_b \rangle \psi_a ,$$

which using Eq. (3.2) implies that the local connection one-form of the Berry connection is

$$(A_U)^a_b = \langle \psi_a, d\psi_b \rangle . \tag{4.11}$$

Eq. (4.11) is the local connection one-form of the Berry connection with respect to the frame $\{\psi_a\}$ for the valence bundle.[7] To emphasize how to interpret d in Eq. (4.11): The vector ψ_b can be expressed in terms of the standard orthonormal frame $e_1, \ldots, e_M$ of $\mathbb{C}^M$ as $\psi_b = \sum_{\tilde{a}} (\psi_b)^{\tilde{a}} e_{\tilde{a}}$ and the deRham derivative $d\psi_b$ yields the vector in $\mathbb{C}^M$ whose components are $d(\psi_b)^{\tilde{a}}$. A local expression for the curvature with respect to the frame $\{\psi_a\}$ can now be obtained from $F_U = dA_U + A_U \wedge A_U$ as

$$(F_U)^a_b = \langle d\psi_a, d\psi_b \rangle + \sum_{c=1}^{m} \langle \psi_a, d\psi_c \rangle \wedge \langle \psi_c, d\psi_b \rangle \,. \tag{4.12}$$

In the particular case of $m = 1$, the local frame of E_{val} consists of only one vector $\psi : U \to \mathbb{C}^M$. If we choose coordinates $\{k^i\}$ on the open set $U \subset X$, we can write the curvature two-form in components as

$$F_U = \frac{1}{2}(F_U)_{ij} dk^i \wedge dk^j = \frac{1}{2}(\langle \partial_i \psi, \partial_j \psi \rangle - \langle \partial_j \psi, \partial_i \psi \rangle) dk^i \wedge dk^j \,.$$

If X is two-dimensional, then the integral of F over U is given by

$$\int_U F = \int_{k(U)} dk^1 dk^2 (F_U)_{12} = \int_{k(U)} dk^1 dk^2 \left(\langle \partial_i \psi, \partial_j \psi \rangle - \langle \partial_j \psi, \partial_i \psi \rangle \right), \tag{4.13}$$

where $k(U) \subset \mathbb{R}^2$ is the image of U under the coordinate chart $k = (k^1, k^2)$. It is in the form of Eq. (4.13) that the Berry connection will come up in the Hall conductivity of a two-dimensional insulator.

4.3 Hall conductivity of a two-dimensional insulator

For a general two-dimensional material, the electric current $\boldsymbol{J}$ in the material due to an applied external electric field $\boldsymbol{E}$ is given in linear response theory by

$$J_i = \sum_j \sigma_{ij} E_j \,, \tag{4.14}$$

where σ_{ij} for $i, j = x, y$ are the components of the *conductivity tensor*. If the material is rotationally symmetric, which is generally a valid assumption in the bulk of a slab of material, then the components of the conductivity tensor satisfy $\sigma_{xx} = \sigma_{yy}$ and $\sigma_{xy} = -\sigma_{yx}$. For an insulator, the longitudinal component σ_{xx} of the conductivity vanishes. However, the transverse component σ_{xy} may be nonzero. The transverse component σ_{xy} is also called the *Hall conductivity* because of its relation to the Hall effect. We will show that in the low-energy, long-wavelength and low-temperature limit, the Hall conductivity of a two-band insulator is proportional to the integral of the Berry curvature of the valence bundle over the momentum space. In the next section we will prove that this integral of the Berry curvature is an integer topological invariant making the Hall conductivity into a *topological effect*. The next section will also show that the two-band model, although it seems like a very narrow model, is in some precise sense the most general case we have to consider in a study of topological effects of two-dimensional insulators.

Consider then a two-dimensional insulator $h : X \to \mathrm{Herm}(2, \mathbb{C})$ with $X = S^2$ or $\mathbb{T}^2$. To compute the conductivity tensor we have to subject the insulator to an external electric

[7] Usually, in the physics literature there is an additional factor of i in the Berry connection and/or a minus sign. These factors are remnants of how one wants to define the local connection one-forms. In physics, one usually writes $\nabla = d \pm iA$ instead of $\nabla = d + A$ for local expressions of the connection.

field that is constant in time and uniform in space. However, for the calculations it turns out to be useful to first consider a time varying electric field $E_j(t) = e^{-i\omega t}E_j$ and take the constant field limit, or DC-limit, at the end of the calculation. We work in a gauge in which $A_0 = 0$ and $A_j(t) = \frac{1}{-i\omega}e^{-i\omega t}E_j$ such that $E_j(t) = \partial_t A_j(t)$.[8]

In the absence of any external field the electrons in the insulator are described by the action

$$S_{\text{matter}}[\phi^*, \phi] = \sum_{n,a,a'} \int_X \frac{d\boldsymbol{k}}{(2\pi)^2}\, \phi_a^*(\boldsymbol{k}, i\omega_n)\left[-i\hbar\omega_n + h_{aa'}(\boldsymbol{k})\right]\phi_{a'}(\boldsymbol{k}, i\omega_n)\,,$$

where ϕ_a are complex Grassmann valued fields, $\sum_n$ is a sum over the Matsubara frequencies $i\omega_n$ and $\int_X \frac{d\boldsymbol{k}}{(2\pi)^2}$ is an integral over the two-dimensional momentum space. All momentum integrals will be over X, so we will drop the subscript X on the integral signs. Coupling to the electromagnetic field is obtained by the minimal substitution $\boldsymbol{k} \to \boldsymbol{k} + \frac{e}{\hbar}\boldsymbol{A}$ in the Hamiltonian. For reasons that will be clarified shortly, the linear response is contained in the terms of the action at most quadractic in $\boldsymbol{A}$. After incorporating conservation of energy and momentum the relevant part of the action is

$$\begin{aligned}
S[\phi^*, \phi; \boldsymbol{A}] &= \sum_n \int \frac{d\boldsymbol{k}}{(2\pi)^2}\, \phi_a^*(\boldsymbol{k}, i\omega_n)\left[-i\hbar\omega_n + h_{aa'}(\boldsymbol{k})\right]\phi_{a'}(\boldsymbol{k}, i\omega_n) \\
&+ \frac{e}{\hbar}\frac{1}{\sqrt{\hbar\beta}}\sum_{n,m}\int \frac{d\boldsymbol{k}d\boldsymbol{q}}{(2\pi)^4}\, \phi_a^*(\boldsymbol{k}+\boldsymbol{q}, i\omega_n + i\omega_m)\left[A_i(\boldsymbol{q}, i\omega_m)\frac{\partial h_{aa'}(\boldsymbol{k})}{\partial k^i}\right]\phi_{a'}(\boldsymbol{k}, i\omega_n) \\
&+ \frac{1}{2}\left(\frac{e}{\hbar}\right)^2\frac{1}{\hbar\beta}\sum_{n,m,r}\int \frac{d\boldsymbol{k}d\boldsymbol{q}d\boldsymbol{p}}{(2\pi)^6}\, \phi_a^*(\boldsymbol{k}+\boldsymbol{q}+\boldsymbol{p}, i\omega_n + i\omega_m + i\omega_r) \\
&\times \left[A_i(\boldsymbol{q}, i\omega_m)A_j(\boldsymbol{p}, i\omega_r)\frac{\partial^2 h_{aa'}(\boldsymbol{k})}{\partial k^i \partial k^j}\right]\phi_{a'}(\boldsymbol{k}, i\omega_n)\,, \qquad (4.15)
\end{aligned}$$

where the sum over repeated indices of type i (the directions in momentum space) and a (the band labels) is implicit. The action is quadratic in the fermionic fields and can be written suggestively in the form

$$S[\phi^*, \phi; \boldsymbol{A}] = -\hbar\sum_{n,n'}\int \frac{d\boldsymbol{k}d\boldsymbol{k}'}{(2\pi)^2}\, \phi_a^*(\boldsymbol{k}, i\omega_n)G^{-1}_{aa'}(\boldsymbol{k}, i\omega_n; \boldsymbol{k}', i\omega_{n'})\phi_{a'}(\boldsymbol{k}', i\omega_{n'})\,,$$

where the inverse Green function G^{-1} can be written as a sum $G^{-1} = G_0^{-1} + G_1^{-1} + G_2^{-1}$ so that G_0^{-1} contains the terms without $\boldsymbol{A}$, G_1^{-1} contains the terms linear in $\boldsymbol{A}$ and G_2^{-1} contains the terms quadratic in $\boldsymbol{A}$. Integrating out the fermions yields the effective action for the electromagnetic theory

$$-\frac{1}{\hbar}S_{\text{eff}}[\boldsymbol{A}] = \text{Tr}[\log(-G^{-1})] = \text{Tr}[\log(G_0^{-1})] + \text{Tr}[\log(1 + G_0(G_1^{-1} + G_2^{-1}))]\,. \qquad (4.16)$$

The current can be obtained from this effective action by computing the variational derivative

$$J^i(\boldsymbol{p}, i\omega_r) = \frac{\delta S_{\text{eff}}[\boldsymbol{A}]}{\delta A_i(\boldsymbol{p}, i\omega_r)}\,. \qquad (4.17)$$

From Eq. (4.17) we can in first instance derive the frequency and momentum dependent *polarization tensor* $\pi_{ij}(\boldsymbol{p}, i\omega_r)$ defined through $J^i(\boldsymbol{p}, i\omega_r) = \sum_j \pi_{ij}(\boldsymbol{p}, i\omega_r)A_j(-\boldsymbol{p}, -i\omega_r)$.

[8] We also work in units where $c = 1$.

Figure 4: (Left) The Kubo diagram corresponding to $-\frac{1}{2}\operatorname{Tr}[G_0 G_1^{-1} G_0 G_1^{-1}]$. (Right) The fish diagram corresponding to $\operatorname{Tr}[G_0 G_2^{-1}]$.

From the above definition, it is clear that the polarization tensor is completely determined by the terms in Eq. (4.16) that are quadratic in $\boldsymbol{A}$. After Taylor expanding the logarithm one finds that the only two terms quadratic in $\boldsymbol{A}$ are $-\frac{1}{2}\operatorname{Tr}[G_0 G_1^{-1} G_0 G_1^{-1}]$ and $\operatorname{Tr}[G_0 G_2^{-1}]$. These terms can be represented diagramatically, as done in Fig. 4. The first term corresponds to the "Kubo" diagram and the second term to the "fish" diagram. Since the Hamiltonian describes an insulator, we expect only an antisymmetric contribution to the conductivity tensor. The antisymmetric contribution must be contained in the Kubo diagram because the fish diagram is invariant under exchange of A_i and A_j. The contribution to the polarization tensor coming from the Kubo diagram is

$$-\hbar \frac{\delta}{\delta A_i(\boldsymbol{p}, i\omega_r)} \left(-\frac{1}{2} \operatorname{Tr}[G_0 G_1^{-1} G_0 G_1^{-1}] \right) = \pi_{ij}^{\text{Kubo}}(\boldsymbol{p}, i\omega_r) A_j(-\boldsymbol{p}, -i\omega_r)$$

with, in the uniform field limit $\boldsymbol{p} \to 0$,

$$\pi_{ij}^{\text{Kubo}}(i\omega_r) = \left(\frac{e}{\hbar}\right)^2 \frac{1}{\beta} \sum_n \int \frac{d\boldsymbol{k}}{(2\pi)^2} \Big\{ [-i\hbar\omega_n + h(\boldsymbol{k})]^{-1}_{aa'} \frac{\partial h_{a'a''}(\boldsymbol{k})}{\partial k^i} \times [-i\hbar\omega_n + i\hbar\omega_r + h(\boldsymbol{k})]^{-1}_{a''a'''} \frac{\partial h_{a'''a}(\boldsymbol{k})}{\partial k^j} \Big\}.$$

Further simplification of this expression can be realized by expressing the Hamiltonian with respect to a basis of eigenvectors. A global basis of eigenvector does not necessarily exist, however, whether the momentum space X is the sphere S^2 or the torus $\mathbb{T}^2$, it is in both cases possible to find a frame $\psi_- : U \to \mathbb{C}^2$ for the valence bundle over a chart $U \subset X$ with coordinates (k^1, k^2) that covers X up to a set of measure zero. Similarly, we can find a frame ψ_+ for the conduction states. Let us use the braket notation

$$\psi_\pm(\boldsymbol{k}) = |\boldsymbol{k}\pm\rangle \quad \text{and} \quad \langle \psi_\pm(\boldsymbol{k}), \cdot \rangle = \langle \boldsymbol{k}\pm|$$

to denote these frames. The Hamiltonian can then be written over U in diagonal form as

$$h(\boldsymbol{k}) = \epsilon_+(\boldsymbol{k}) \, |\boldsymbol{k}+\rangle \langle \boldsymbol{k}+| + \epsilon_-(\boldsymbol{k}) \, |\boldsymbol{k}-\rangle \langle \boldsymbol{k}-| \, .$$

In the $|\boldsymbol{k}\pm\rangle$-basis, the Green function G_0 takes the form

$$-\frac{1}{\hbar} G_0(\boldsymbol{k}, i\omega_n) = [-i\hbar\omega_n + h(\boldsymbol{k})]^{-1} = \frac{|\boldsymbol{k}+\rangle \langle \boldsymbol{k}+|}{-i\hbar\omega_n + \epsilon_+(\boldsymbol{k})} + \frac{|\boldsymbol{k}-\rangle \langle \boldsymbol{k}-|}{-i\hbar\omega_n + \epsilon_-(\boldsymbol{k})} \, ,$$

as can be checked by direct multiplication with $\hbar G_0^{-1}$. Hence, the Kubo diagram yields

$$\pi_{ij}^{\text{Kubo}}(i\omega_r) = \left(\frac{e}{\hbar}\right)^2 \frac{1}{\beta}\sum_n \int \frac{d\boldsymbol{k}}{(2\pi)^2}\ \text{Tr}\Bigg\{ \left[\frac{|\boldsymbol{k}+\rangle\langle\boldsymbol{k}+|}{-i\hbar\omega_n+\epsilon_+(\boldsymbol{k})} + \frac{|\boldsymbol{k}-\rangle\langle\boldsymbol{k}-|}{-i\hbar\omega_n+\epsilon_-(\boldsymbol{k})}\right]\frac{\partial h(\boldsymbol{k})}{\partial k^i} \\ \times \left[\frac{|\boldsymbol{k}+\rangle\langle\boldsymbol{k}+|}{-i\hbar\omega_n+i\hbar\omega_r+\epsilon_+(\boldsymbol{k})} + \frac{|\boldsymbol{k}-\rangle\langle\boldsymbol{k}-|}{-i\hbar\omega_n+i\hbar\omega_r+\epsilon_-(\boldsymbol{k})}\right]\frac{\partial h(\boldsymbol{k})}{\partial k^j}\Bigg\}. \tag{4.18}$$

Then, after working out the Matsubara sums, it can be shown that in the low-temperature limit the only relevant terms will be

$$\pi_{ij}^{\text{Kubo}}(i\omega_r) = -\left(\frac{e}{\hbar}\right)^2 \int \frac{d\boldsymbol{k}}{(2\pi)^2} M_{ij}(\boldsymbol{k}) \left(\frac{n_{\text{FD}}(\epsilon_+(\boldsymbol{k}))}{\epsilon_-(\boldsymbol{k})-\epsilon_+(\boldsymbol{k})+i\hbar\omega_r} + \frac{n_{\text{FD}}(\epsilon_-(\boldsymbol{k})+i\hbar\omega_r)}{\epsilon_+(\boldsymbol{k})-\epsilon_-(\boldsymbol{k})-i\hbar\omega_r}\right) \\ -\left(\frac{e}{\hbar}\right)^2 \int \frac{d\boldsymbol{k}}{(2\pi)^2} M_{ji}(\boldsymbol{k}) \left(\frac{n_{\text{FD}}(\epsilon_-(\boldsymbol{k}))}{\epsilon_+(\boldsymbol{k})-\epsilon_-(\boldsymbol{k})+i\hbar\omega_r} + \frac{n_{\text{FD}}(\epsilon_+(\boldsymbol{k})+i\hbar\omega_r)}{\epsilon_-(\boldsymbol{k})-\epsilon_+(\boldsymbol{k})-i\hbar\omega_r}\right), \tag{4.19}$$

where we have introduced the short hand

$$M_{ij} = \langle \boldsymbol{k}+|\frac{\partial h(\boldsymbol{k})}{\partial k^i}|\boldsymbol{k}-\rangle\langle \boldsymbol{k}-|\frac{\partial h(\boldsymbol{k})}{\partial k^j}|\boldsymbol{k}+\rangle\ .$$

In principle, there are two more terms in π_{ij}^{Kubo}, but these are proportional to derivatives of the Fermi-Dirac distribution evaluated, in the low-energy limit, away from 0 and thus they vanish in the low-energy and low-temperature limit .

So far, we have been working in imaginary time and Matsubara frequencies, but at this point we can go back to real time by analytically continuing $\pi_{ij}^{\text{Kubo}}(z)$ to the whole of $\mathbb{C}$. The real time response is $\pi_{ij}^{\text{Kubo}}(\omega+i\eta^+)$, where η^+ means that we let η approach 0 from above. This factor of η^+ is needed to get the proper time ordering in the path integral, see [32]. After the analytic continuation, the response is

$$\pi_{ij}^{\text{Kubo}}(\omega) = -\left(\frac{e}{\hbar}\right)^2 \int \frac{d\boldsymbol{k}}{(2\pi)^2} \frac{(\epsilon_+(\boldsymbol{k})-\epsilon_-(\boldsymbol{k}))\left(M_{ij}n_{\text{FD}}(\epsilon_-(\boldsymbol{k})+\hbar\omega)+M_{ji}n_{\text{FD}}(\epsilon_-(\boldsymbol{k}))\right)}{(\epsilon_+(\boldsymbol{k})-\epsilon_-(\boldsymbol{k}))^2-(\hbar\omega)^2} \\ -\left(\frac{e}{\hbar}\right)^2 \int \frac{d\boldsymbol{k}}{(2\pi)^2} \frac{\hbar\omega(M_{ij}n_{\text{FD}}(\epsilon_-(\boldsymbol{k})+\hbar\omega)-M_{ji}n_{\text{FD}}(\epsilon_-(\boldsymbol{k})))}{(\epsilon_+(\boldsymbol{k})-\epsilon_-(\boldsymbol{k}))^2-(\hbar\omega)^2}. \tag{4.20}$$

Here we have used that $\lim_{\eta\downarrow 0}\frac{1}{(\epsilon_+-\epsilon_-)-\omega-i\eta^+} = \frac{1}{((\epsilon_+-\epsilon_-)-\omega)^2}+i\pi\delta((\epsilon_+-\epsilon_-)-\omega)$ and that due to the gap condition $\epsilon_+-\epsilon_->0$, the argument of the δ-function will not become zero in the low-energy limit $\omega\to 0$.

To get to the conductivity tensor, recall that $A_j = \frac{1}{-i\omega}e^{-i\omega t}E_j$ so that the frequency dependent conductivity can be obtained from the polarization through $\sigma_{ij}(\omega) = \frac{1}{-i\omega}\pi_{ij}(\omega)$. The first term in π_{ij}^{Kubo} will thus yield a divergence in the conductivity in the limit $\omega\to 0$. However, in Appendix A it is shown that this divergence is precisely cancelled by the fish diagram. The cancellation of the divergences in the longitudinal conductivity reflects the fact that we consider an insulator. The conductivity tensor in the low-energy, long-wavelength and low-temperature limit is thus

$$\sigma_{ij} = -\frac{ie^2}{\hbar}\int \frac{d\boldsymbol{k}}{(2\pi)^2}\frac{M_{ij}-M_{ji}}{(\epsilon_+(\boldsymbol{k})-\epsilon_-(\boldsymbol{k}))^2}\,.$$

It remains to compute the matrix elements $M_{ij} = \langle \boldsymbol{k}+|\frac{\partial h(\boldsymbol{k})}{\partial k^i}|\boldsymbol{k}-\rangle\langle \boldsymbol{k}-|\frac{\partial h(\boldsymbol{k})}{\partial k^j}|\boldsymbol{k}+\rangle$.

Using $1 = |\boldsymbol{k}+\rangle\langle\boldsymbol{k}+| + |\boldsymbol{k}-\rangle\langle\boldsymbol{k}-|$, it follows that

$$\begin{aligned}\langle\boldsymbol{k}+|\frac{\partial h(\boldsymbol{k})}{\partial k^i}|\boldsymbol{k}-\rangle &= \langle\boldsymbol{k}+|\frac{\partial}{\partial k^i}\Big(\epsilon_+(\boldsymbol{k}) + (\epsilon_-(\boldsymbol{k})-\epsilon_+(\boldsymbol{k}))\,|\boldsymbol{k}-\rangle\langle\boldsymbol{k}-|\,\Big)|\boldsymbol{k}-\rangle \\ &= (\epsilon_-(\boldsymbol{k})-\epsilon_+(\boldsymbol{k}))\,\langle\boldsymbol{k}+|\frac{\partial}{\partial k^i}\Big(\,|\boldsymbol{k}-\rangle\langle\boldsymbol{k}-|\,\Big)|\boldsymbol{k}-\rangle \\ &= (\epsilon_-(\boldsymbol{k})-\epsilon_+(\boldsymbol{k}))\,\langle\boldsymbol{k}+|\,\frac{\partial}{\partial k^i}(|\boldsymbol{k}-\rangle)\end{aligned}$$

and, similarly,

$$\langle\boldsymbol{k}-|\frac{\partial h(\boldsymbol{k})}{\partial k^j}|\boldsymbol{k}+\rangle = (\epsilon_-(\boldsymbol{k})-\epsilon_+(\boldsymbol{k}))\frac{\partial}{\partial k^j}(\langle\boldsymbol{k}-|)\,|\boldsymbol{k}+\rangle\ .$$

Hence,

$$\begin{aligned}M_{ij} &= (\epsilon_-(\boldsymbol{k})-\epsilon_+(\boldsymbol{k}))^2\,\langle\boldsymbol{k}+|\,\frac{\partial}{\partial k^i}(|\boldsymbol{k}-\rangle)\frac{\partial}{\partial k^j}(\langle\boldsymbol{k}-|)\,|\boldsymbol{k}+\rangle \\ &= (\epsilon_+(\boldsymbol{k})-\epsilon_-(\boldsymbol{k}))^2\frac{\partial}{\partial k^j}(\langle\boldsymbol{k}-|)\,|\boldsymbol{k}+\rangle\,\langle\boldsymbol{k}+|\,\frac{\partial}{\partial k^i}(|\boldsymbol{k}-\rangle) \\ &= (\epsilon_+(\boldsymbol{k})-\epsilon_-(\boldsymbol{k}))^2\left(\frac{\partial}{\partial k^j}(\langle\boldsymbol{k}-|)\frac{\partial}{\partial k^i}(|\boldsymbol{k}-\rangle) - \frac{\partial}{\partial k^j}(\langle\boldsymbol{k}-|)\,|\boldsymbol{k}-\rangle\langle\boldsymbol{k}-|\,\frac{\partial}{\partial k^i}(|\boldsymbol{k}-\rangle)\right)\ .\end{aligned}$$

Note that the second term in the brackets is symmetric in i and j and, therefore, drops out when taking the difference of $M_{ij} - M_{ji}$ in the conductivity tensor. It follows that

$$\sigma_{ij} = -\frac{ie^2}{\hbar}\int_X \frac{d\boldsymbol{k}}{(2\pi)^2}\,\left[\frac{\partial}{\partial k^j}(\langle\boldsymbol{k}-|)\frac{\partial}{\partial k^i}(|\boldsymbol{k}-\rangle) - \frac{\partial}{\partial k^i}(\langle\boldsymbol{k}-|)\frac{\partial}{\partial k^j}(|\boldsymbol{k}-\rangle)\right]\ .$$

In the square brackets we now recognize the local expression of the Berry curvature Eq. (4.12). The conductivity tensor can thus be finally expressed as

$$\sigma_{ij} = \frac{i}{2\pi}\frac{e^2}{h}\int_X d\boldsymbol{k}\,F_{ij}(\boldsymbol{k})\,, \tag{4.21}$$

where F_{ij} is the Berry curvature of the valence bundle. Although the computations above have been performed in a gauge with $A_0 = 0$, we argue in Appendix A that the result Eq. (4.21) is independent of this choice of gauge.

In the next section we will prove that the quantity $C_1 = \frac{i}{2\pi}\int_X d\boldsymbol{k}\,F_{12}(\boldsymbol{k})$, called the *first Chern number*, is always an integer, and that this integer topologically classifies the two-dimensional insulators. We also argue that a two-band model is in some particular sense the most general model needed in the study of topological effects in two-dimensional insulators. Therefore, Eq. (4.21) shows that the Hall conductivity of any two-dimensional insulator is quantized in units of e^2/h.

5 Topological classification of insulators

Tight-binding models of d-dimensional insulators with M total bands and m valence bands are topologically classified by the homotopy classes $[X, G_m(\mathbb{C}^M)]$, where $X = \mathbb{T}^d, S^d$ depending on whether we study periodic or continuum models. On the other hand, to every such insulator we can associate its valence bundle, which is a rank m subbundle of $X\times\mathbb{C}^M$. The relation between these two facts is that the set of homotopy classes $[X, G_m(\mathbb{C}^M)]$ is in bijection with the set of isomorphism classes $\mathrm{Vect}^m_{\mathbb{C}}(X)$, provided that the number M

is large enough. Under this bijection, the topological equivalence class of an insulator is the same thing as the isomorphism class of its valence bundle. In one dimension and three dimensions, all complex vector bundles are trivial, whilst in two dimensions, complex vector bundles are fully classified by their first Chern class. A key fact in these low dimensional cases is that the non-triviality of a vector bundle of arbitrary rank over a low dimensional base space can only be contained in a low rank subbundle. In particular, the non-triviality over a two- or three-dimensional base can only be contained in a line bundle. In this way, the two-band model then becomes the universal model for topological effects in two and three dimensions. As a bonus, the Chern number for two-band models has an easy-to-compute interpretation in terms of a winding number.

5.1 Homotopy classification of complex vector bundles

Insulators with M total bands and m valence bands are topologically classified by the homotopy classes $[X, G_m(\mathbb{C}^M)]$, where $X = \mathbb{T}^d$ or S^d and $G_m(\mathbb{C}^M)$ is the Grassmannian. On the other hand, we will argue below that whenever M is sufficiently large with respect to m and d, complex vector bundles of rank m over a compact base space X of dimension d are also classified by $[X, G_m(\mathbb{C}^M)]$. Studying topological equivalence classes of insulators is thus equivalent to studying isomorphism classes of complex vector bundles. We show that the valence bundle construction of Eq. (4.10) connects the two points of view.

The main object in the topological classification of complex vector bundles is the so-called *tautological bundle* over a Grassmannian. Recall that the Grassmannian $G_m(\mathbb{C}^M)$ consists of the m-dimensional linear subspaces of $\mathbb{C}^M$, or the m-planes in $\mathbb{C}^M$. The tautological bundle $\pi : \tau_{m,M} \to G_m(\mathbb{C}^M)$ is defined as follows. The total space $\tau_{m,M}$ is realized as a subspace of $G_m(\mathbb{C}^M) \times \mathbb{C}^M$ via

$$\tau_{m,M} = \{(\ell, v) \in G_m(\mathbb{C}^M) \times \mathbb{C}^M : v \in \ell\}.$$

In words, the total space $\tau_{m,M}$ consists of all pairs (ℓ, v) where ℓ is an m-plane in $\mathbb{C}^M$ and v is a vector in the m-plane ℓ. The projection π is given by $\pi(\ell, v) = \ell$. If $\ell \in G_m(\mathbb{C}^M)$ is an m-plane, then a local trivialization around ℓ can be constructed over the open

$$U_\ell = \{\ell' \in G_m(\mathbb{C}^M) : \dim \pi_\ell(\ell') = m\},$$

where $\pi_\ell : \mathbb{C}^M \to \ell$ denotes the orthogonal projection, by defining[9]

$$\varphi : \pi^{-1}(U_\ell) \to U_\ell \times \ell \cong U_\ell \times \mathbb{C}^m, \quad (\ell', v) \mapsto (\ell', \pi_\ell(v)).$$

The importance of the tautological bundles is that vector bundles over compact base spaces X can be obtained as the pullback of a tautological bundle along a map $f : X \to G_m(\mathbb{C}^M)$. Recall that if $\pi : E \to Y$ is a vector bundle and $f : X \to Y$ a map, then the pullback bundle $f^*E \to X$ is defined to be the bundle with fibers $(f^*E)_x = E_{f(x)}$. The homotopy classification of complex vector bundles can be stated in the following theorem, for the proof of which we refer to [33].

Theorem 5.1. If X is a paracompact topological space, then

$$[X, G_m(\mathbb{C}^\infty)] \to \mathrm{Vect}^m_{\mathbb{C}}(X), \quad [f : X \to G_m(\mathbb{C}^\infty)] \mapsto [f^*\tau_{m,\infty}]$$

is a bijection.

[9] See [33, Lemma 1.15] for the details concerning continuity.

A few comments about the statement are in order. First of all, any compact space is also paracompact so the statement applies in particular to $X = \mathbb{T}^d$ or S^d, which are the spaces we care about for the physics of insulators. Secondly, $G_m(\mathbb{C}^\infty)$ and $\tau_{m,\infty}$ are the direct limits of $G_m(\mathbb{C}^M)$ and $\tau_{m,M}$ as $M \to \infty$. The advantage of formulating the theorem in terms of the direct limit is that we get a statement independent of M. We call $G_m(\mathbb{C}^\infty)$ the *classifying space* and $\tau_{m,\infty}$ the *universal bundle* for complex vector bundles of rank m.

The statement in Theorem 5.1 is designed so as to become independent of the base space X and the rank m. However, for specific combinations of X and m it is possible to discern an M such that $[f : X \to G_m(\mathbb{C}^M)] \mapsto [f^*\tau_{m,M}]$ already yields a bijection of $[X, G_m(\mathbb{C}^M)]$ with $\mathrm{Vect}^m_{\mathbb{C}}(X)$, without the need for the direct limit. Suppose then that P is the spectral projector of an insulator for a sufficiently large total number of bands M. The valence bundle of the insulator is defined as the image of P, when interpreted as a bundle map $P : X \times \mathbb{C}^M \to X \times \mathbb{C}^M$. On the other hand, we can also view P as a map $P : X \to G_m(\mathbb{C}^M)$ by identifying the projection $P(p) : \mathbb{C}^M \to \mathbb{C}^M$ with its image in $\mathbb{C}^M$. In this way, the fiber of $P^*\tau_{m,M}$ over the point $p \in X$ gets identified with the image of $P(p)$ in $\mathbb{C}^M$. It follows that $P^*\tau_{m,M}$ is isomorphic to the valence bundle. This isomorphism combined with the bijection $[X, G_m(\mathbb{C}^M)] \to \mathrm{Vect}^m_{\mathbb{C}}(X)$ establishes the correspondence between topological equivalence classes of insulators and the isomorphism classes of their valence bundles.

As we will argue later in this section, the case of interest for the physics of insulators in $d \leq 3$ is the classification of rank $m = 1$ complex vector bundles, i.e. complex line bundles. For the case of line bundles we can make precise how large M has to be in order for $G_1(\mathbb{C}^M)$ to classify line bundles over S^d or $\mathbb{T}^d$. It turns out that we need $2(M-1) \geq d$. To see this, we use some basic considerations about CW-complexes. First of all, by definition, $G_1(\mathbb{C}^M)$ is equal to the $2(M-1)$-dimensional CW-complex $\mathbb{C}P^{M-1}$. Recall that $\mathbb{C}P^M$ is obtained from $\mathbb{C}P^{M-1}$ by attaching a cell of dimension $2M$. If X is a CW-complex of dimension d, then it follows from the Cellular Approximation Theorem that the inclusion $\mathbb{C}P^{M-1} \to \mathbb{C}P^M$ induces a bijection $[X, \mathbb{C}P^{M-1}] \to [X, \mathbb{C}P^M]$ as long as $d \leq 2(M-1)$. Hence, by induction, the inclusion $\mathbb{C}P^{M-1} \to \mathbb{C}P^\infty$ induces a bijection $[X, \mathbb{C}P^{M-1}] \to [X, \mathbb{C}P^\infty]$. So, as long as we work with $d \leq 2(M-1)$ it does not matter whether we use $\mathbb{C}P^\infty$ or $\mathbb{C}P^{M-1}$ in Theorem 5.1. In particular, for $d = 2$ it suffices to consider maps $X \to \mathbb{C}P^1$ in the classification of line bundles over X. We will exploit this fact when developing the theory of the two-dimensional, two-band model.

However, the classification of complex line bundles in terms of $\mathbb{C}P^\infty$ contains crucial information at an abstract level. One of the special properties of $\mathbb{C}P^\infty$ is that it is the so-called Eilenberg-MacLane space $K(\mathbb{Z}, 2)$. If one is not familiar with Eilenberg-MacLane spaces, then this may seem like a rather arbitrary property. It has an important implication though. Namely, that for CW-complexes X, like $\mathbb{T}^d$ and S^d, there is an isomorphism

$$[X, \mathbb{C}P^\infty] \to H^2(X; \mathbb{Z})$$

given by the Representability of Cohomology [34]. Here, $H^2(X; \mathbb{Z})$ denotes the degree two integral cohomology of X. The isomorphism implies that the topological equivalence class of an insulator can be determined from a certain cohomology class. Our purpose is now to prove that this cohomology class can be determined from the Berry curvature.

5.2 Integrality of the Chern classes

Recall that the Chern classes have been introduced as topological invariants for smooth vector bundles over a manifold X in terms of complexified deRham cohomology classes in

$H^*_{dR}(X;\mathbb{C})$, computed from the curvature. However, from the purely topological point of view the Chern classes can be defined axiomatically as integer cohomology classes taking values in the singular cohomology ring $H^*(X;\mathbb{Z})$. The definition in terms of integral singular cohomology makes clear the integral nature of the Chern classes. Moreover, for smooth vector bundles the two different definitions agree. One of the consequences of this result is that the first Chern number as it appears below Eq. (4.21) is always an integer.

The axiomatic definition of the Chern classes proceeds as follows. The Chern classes are the maps $c_i : \mathrm{Vect}_{\mathbb{C}}(X) \to H^{2i}(X;\mathbb{Z})$ for $i = 1, 2, \ldots$ that are uniquely determined by the following four axioms:

1. (Naturality) $c_i(f^*E) = f^*c_i(E)$
2. (Whitney sum) $c(E_1 \oplus E_2) = c(E_1) \smile c(E_2)$, where $c = 1 + c_1 + c_2 + \cdots \in H^*(X;\mathbb{Z})$
3. $c_i(E) = 0$ if $i > \operatorname{rank} E$
4. $c_1(\tau_{1,\infty}) \in H^2(\mathbb{C}P^\infty;\mathbb{Z})$ is a generator.

The idea behind these axioms is that once we fix $c_1(\tau_{1,\infty}) \in H^2(\mathbb{C}P^\infty;\mathbb{Z})$, then all the Chern classes of all other complex vector bundles are determined. Axioms three and four together determine all the Chern classes of $\tau_{1,\infty} \to \mathbb{C}P^\infty$. Indeed, the first Chern class is taken to be one of the two generators of $H^2(\mathbb{C}P^\infty;\mathbb{Z}) \cong \mathbb{Z}$, and the higher order Chern classes vanish. As we have seen, every complex line bundle over X is isomorphic to the pullback of the tautological line bundle $\tau_{1,\infty} \to \mathbb{C}P^\infty$ via a map $f : X \to \mathbb{C}P^\infty$. The first axiom thus determines the Chern classes for all line bundles. Finally, "the splitting principle" [33, Propostion 3.3] states that for any vector bundle $E \to X$ there is a space $F(E)$ and a map $p : F(E) \to X$ such that $p^*(E) \to F(E)$ is isomorphic to a direct sum of line bundles, and $p^* : H^*(X;\mathbb{Z}) \to H^*(F(E);\mathbb{Z})$ is injective. The Chern classes of $p^*(E) \to F(E)$ in $H^*(F(E);\mathbb{Z})$ are determined by the second axiom and the fact that we know all Chern classes of line bundles. The Chern classes of $p^*(E)$ are the images of the Chern classes of E under p^*. Since p^* is injective, these images uniquely determine the Chern classes of E.

From the discussion in Section 3.4, it can be seen that the deRham cohomology formulation of the Chern classes satisfies all the four axioms. The first axiom, naturality, and the third axiom, that the Chern classes vanish when $i > \operatorname{rank} E$, have been verified explicitly. Moreover, the second axiom is easily seen to hold after replacing the cup product ($\smile$) of singular cohomology by the wedge product ($\wedge$) of deRham cohomology. Finally, Eq. (3.5) states something resembling the fourth axiom, be it formulated in terms of $\mathbb{C}P^1$ rather than $\mathbb{C}P^\infty$. This is not a problem, because the inclusion $\iota : \mathbb{C}P^1 \to \mathbb{C}P^\infty$ induces an isomorphism $\iota^* : H^2(\mathbb{C}P^\infty;\mathbb{Z}) \to H^2(\mathbb{C}P^1;\mathbb{Z})$. Moreover, the tautological line bundle $\tau_{1,1}$ over $\mathbb{C}P^1$ is the pullback of the tautological line bundle $\tau_{1,\infty}$ over $\mathbb{C}P^\infty$ along the inclusion ι. Fixing the first Chern class of $\tau_{1,\infty}$ to be a generator of $H^2(\mathbb{C}P^\infty;\mathbb{Z})$ is thus equivalent to fixing the first Chern class of $\tau_{1,1}$ to be a generator of $H^2(\mathbb{C}P^1;\mathbb{Z})$. The reason for a priori formulating the fourth axiom in terms of $\mathbb{C}P^\infty$ is that it is then immediately applicable in the determination of the first Chern class for all line bundles via $[X, \mathbb{C}P^\infty] \cong \mathrm{Vect}^1_{\mathbb{C}}(X)$. However, by the discussion above, knowing the first Chern class of $\tau_{1,1} \to \mathbb{C}P^1$ together with the first three axioms also allows one to determine the Chern classes for all vector bundles.

Consider now the composite

$$H^2(\mathbb{C}P^1;\mathbb{Z}) \xrightarrow{j_*} H^2(\mathbb{C}P^1;\mathbb{C}) \xrightarrow{\cong} H^2_{dR}(\mathbb{C}P^1;\mathbb{C}), \tag{5.1}$$

where the first map is induced by the inclusion $j : \mathbb{Z} \to \mathbb{C}$ and the second map is the deRham isomorphism [27]. To show the equivalence of the Chern classes in the deRham formulation to the Chern classes in the integral cohomology formulation we will argue that the composite Eq. (5.1) is injective and maps the integer first Chern class $c_1(\tau_{1,1})$ to the curvature class $\frac{i}{2\pi}[F]$ of $\tau_{1,1}$. The first step is to note that by the Universal Coefficient Theorem, the canonical homomorphism

$$j_* : H^2(\mathbb{C}P^1; \mathbb{Z}) \to H^2(\mathbb{C}P^1; \mathbb{C})$$

fits into a commutative diagram

$$\begin{array}{ccc} H^2(\mathbb{C}P^1; \mathbb{Z}) & \xrightarrow{j_*} & H^2(\mathbb{C}P^1; \mathbb{C}) \\ \downarrow\cong & & \downarrow\cong \\ \mathrm{Hom}(H_2(\mathbb{C}P^1); \mathbb{Z}) & \xrightarrow{j_*} & \mathrm{Hom}(H_2(\mathbb{C}P^1); \mathbb{C}) \end{array} ,$$

where the homology groups are with $\mathbb{Z}$ coefficients, the vertical arrows are the isomorphisms of the Universal Coefficient Theorem and the lower horizontal arrow is given by postcomposition with j, which is injective since j is injective. Therefore, the upper horizontal arrow $j_* : H^2(\mathbb{C}P^1; \mathbb{Z}) \to H^2(\mathbb{C}P^1; \mathbb{C})$ is injective.

Now, $H^2(\mathbb{C}P^1) \cong \mathbb{Z}$ and is generated by an orientation class $\mu_{\mathbb{C}P^1}$. Therefore, a map in $\mathrm{Hom}(H_2(\mathbb{C}P^1), R)$ for $R = \mathbb{Z}, \mathbb{C}$ is uniquely determined by where it sends $\mu_{\mathbb{C}P^1}$. The first Chern class $c_1(\tau_{1,1}) \in H^2(\mathbb{C}P^1; \mathbb{Z})$ maps under the isomorphism of the Universal Coefficient Theorem to a map $c_1(\tau_{1,1}) \frown -$ that sends $\mu_{\mathbb{C}P^1}$ to $\pm 1 \in \mathbb{Z}$, since an isomorphism sends generators to generators. The sign depends on the choice of orientation class $\mu_{\mathbb{C}P^1}$. For definiteness, let us assume that $c_1(\tau_{1,1}) \frown \mu_{\mathbb{C}P^1} = 1 \in \mathbb{Z}$. Postcomposing with the inclusion $j : \mathbb{Z} \to \mathbb{C}$, we get the map $j(c_1(\tau_{1,1}) \frown -)$ that maps $\mu_{\mathbb{C}P^1}$ to $1 \in \mathbb{C}$. From the commutative diagram, it then follows that $j_* c_1(\tau_{1,1}) \frown -$ is the map that sends $\mu_{\mathbb{C}P^1}$ to $1 \in \mathbb{C}$. We conclude from the fact that $\int_{\mathbb{C}P^1} \frac{i}{2\pi} F = 1$, cf. Eq. (3.5), that under the deRham isomorphism, $j_*(c_1(\tau_{1,1}))$ corresponds to $\frac{i}{2\pi}[F]$. This proves that $c_1(\tau_{1,1})$ maps to $\frac{i}{2\pi}[F]$ under the composite Eq. (5.1).

The first Chern class of the tautological line bundle in the deRham formulation is thus the image of the axiomatic first Chern class $c_1(\tau_{1,1})$ under j_*. Using the first three axioms, the Chern classes of all other vector bundles are determined in terms of $c_1(\tau_{1,1})$. Therefore, the axiomatic definition of the Chern classes coincides with the curvature definition. The fact that the Chern classes in $H^*_{dR}(X; \mathbb{C})$ correspond to the Chern classes in $H^*(X; \mathbb{Z})$ implies that the integral of the top Chern class over an even dimensional X is always an integer. In particular, for a two-dimensional base space X this implies that

$$C_1(E) = \frac{i}{2\pi} \int_X [F]$$

is always an integer, which proves that indeed the Hall conductivity discussed previously is quantized.

5.3 Classification of insulators in spatial dimensions $d = 1, 2, 3$

We show that complex line bundles over a compact base space are completely classified by their first Chern class. Furthermore, we show that the topological classification of rank m complex vector bundles over a base space X of dimension $d \leq 3$ is equivalent to the classification of complex line bundles because every rank m vector bundle over such an X can be decomposed into a direct sum of $m - 1$ trivial line bundles and one possibly

non-trivial line bundle. Combination of these results yields a complete classification of complex vector bundles, and thus of topological insulators, in $d \leq 3$ in terms of the first Chern class.

The classification of complex line bundles in terms of their first Chern class is summarized in the following theorem.

Theorem 5.2. The map $\mathrm{Vect}^1_{\mathbb{C}}(X) \to H^2(X;\mathbb{Z})$ given by $E \mapsto c_1(E)$ is an isomorphism.

Proof. There is a chain of isomorphisms

$$\mathrm{Vect}^1_{\mathbb{C}}(X) \xleftarrow{\cong} [X, \mathbb{C}P^\infty] \xrightarrow{\cong} H^2(X;\mathbb{Z}) \ .$$

The first map is the isomorphism $[f] \mapsto f^*\tau_{1,\infty}$ from the homotopy classification of line bundles. The second map is the representability of cohomology, which is explicitly given by $[f] \mapsto f^*c_1(\tau_{1,\infty})$. If $E \cong f^*\tau_{1,\infty}$, then by the first axiom of the Chern classes,

$$f^*c_1(\tau_{1,\infty}) = c_1(f^*\tau_{1,\infty}) = c_1(E)\,.$$

Hence, the assignment $E \mapsto c_1(E)$ is an isomorphism. □

This is a very special property of line bundles. Higher rank bundles are in general not fully classified by all their nonzero Chern classes. However, over base spaces of dimensions less than or equal to three, it is actually true that arbitrary rank complex vector bundles are classified by their *first* Chern class. The key point is that for a complex vector bundle $E \to X$ of rank m over a base space X of dimension $d \leq 3$, we can decompose E as

$$E \cong L \oplus \underline{\mathbb{C}}^{n-1}\,,$$

where L is a, possibly non-trivial, line bundle and $\underline{\mathbb{C}}$ denotes the trivial line bundle over X. The decomposition $E \cong L \oplus \underline{\mathbb{C}}^{m-1}$ is straightforward to derive in the smooth world from transversality considerations [26], and the result is collected in the Proposition below.

Proposition 5.3. Let $E \to X$ be a smooth real vector bundle such that $\operatorname{rank} E > \dim X$, then E has non-vanishing smooth section.

One might be worried about the smoothness assumption, since we are purporting to derive a *topological* classification. However, by the Whitney Approximation Theorem in combination with the homotopy classification of complex vector bundles, any isomorphism class of topological complex vector bundles may be represented by a smooth complex vector bundle. For the topological classification of complex vector bundles it thus suffices to consider smooth complex vector bundles.

Although stated for *real* vector bundles, Proposition 5.3 can be used to find a non-vanishing section of a *complex* vector bundle as well. Given a complex vector bundle $E \to X$, the underlying real vector bundle $E_{\mathbb{R}} \to X$ has a complex structure

$$J : E_{\mathbb{R}} \to E_{\mathbb{R}}\,.$$

We can apply the proposition to the real vector bundle $E_{\mathbb{R}}$ to obtain a non-vanishing section $s : X \to E_{\mathbb{R}}$. Using the complex structure, we define a second non-vanishing section $J \circ s : X \to E_{\mathbb{R}}$. Now, s and $J \circ s$ span a trivial complex line bundle L_s in $E_{\mathbb{R}}$. Going back to the actual complex bundle E, this yields $E \cong E' \oplus \underline{\mathbb{C}}$, where we take $\underline{\mathbb{C}}$ to be the trivial complex line bundle spanned by s, and E' to be the orthogonal complement of this trivial bundle in E. If $d = 1$, then we can repeat this procedure all the way down

until we have $E \cong \underline{\mathbb{C}}^m$. There are thus no non-trivial complex vector bundles over a one-dimensional base space, which implies that there are no topological insulators in $d = 1$. On the other hand, if $d = 2, 3$, then we can only repeat the procedure until we arrive at $E = L \oplus \underline{\mathbb{C}}^{m-1}$, where L is still a possibly non-trivial line bundle. Suppose then that we have a rank m complex vector bundle E over a $d = 2, 3$ dimensional base space X, and that we have decomposed $E \cong L \oplus \underline{\mathbb{C}}^{m-1}$. The first Chern class of E is equal to the first Chern class of L by the Whitney sum axiom. Moreover, all higher Chern classes are zero. For a different decomposition $E \cong L' \oplus \underline{\mathbb{C}}^{m-1}$, we have $c_1(L) = c_1(L')$ since both equal $c_1(E)$. Since line bundles are classified by their first Chern class, it follows that L' is isomorphic to L. So, rank m complex vector bundles over a $d = 2, 3$ base space are equivalent to complex line bundles, and are thus classified by the first Chern class.

The valence bundle is a vector bundle over $\mathbb{T}^d$ or S^d. The first Chern class thus takes values in $H^2(\mathbb{T}^d; \mathbb{Z})$ or $H^2(S^d; \mathbb{Z})$. These cohomology groups are known and have been collected in Table 1. The $\mathbb{Z}$ entries for $\mathbb{T}^2$ and S^2 mean that $H^2(\mathbb{T}^2; \mathbb{Z})$ and $H^2(S^2; \mathbb{Z})$ are

d	$H^2(\mathbb{T}^d; \mathbb{Z})$	$H^2(S^d; \mathbb{Z})$
2	$\mathbb{Z}$	$\mathbb{Z}$
3	$\mathbb{Z}^3$	0

Table 1: Second degree cohomology of $\mathbb{T}^d$ and S^d in low dimensions.

isomorphic to $\mathbb{Z}$. In both cases, the isomorphism is explicitly realized by Poincaré duality, i.e. by integrating the first Chern class, viewed as a two-form, over the base manifold. This integral the first Chern number. This shows that insulators in two dimensions, for both periodic and continuum models, are classified by their first Chern number.

The $\mathbb{Z}^3$ entry for $H^2(\mathbb{T}^3; \mathbb{Z})$ is due to the Künneth formula. More concretely, each copy of $\mathbb{Z}$ in the $\mathbb{Z}^3$ is actually just the $\mathbb{Z}$ invariant of a copy of $\mathbb{T}^2$ embedded in $\mathbb{T}^3$. The three copies of $\mathbb{Z}$ correspond to the three inequivalent ways of embbedding $\mathbb{T}^2$ into $\mathbb{T}^3$. In physical terms, each copy of $\mathbb{Z}$ in the $\mathbb{Z}^3$ invariant corresponds to a stacking together of two-dimensional insulators. It is thus not a truly a new three-dimensional invariant. That there is no truly three-dimensional invariant for topological insulators is reflected in the fact that $H^2(S^3; \mathbb{Z}) = 0$, which means that continuum models in three dimensions are always trivial.

5.4 The two-band model

Let us use the tools we have developed for the classification of complex line bundles to develop a detailed understanding of the two-dimensional, two-band model

$$h : X \to \mathrm{Herm}(2, \mathbb{C}), \quad h(p) = \boldsymbol{B}(p) \cdot \boldsymbol{\sigma}\,,$$

where $X = S^2$ or $\mathbb{T}^2$, $\boldsymbol{B} = (B_x, B_y, B_z) : X \to \mathbb{R}^3$ and $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli matrices. Writing h in this form gives the most general two-band Hamiltonian with Fermi level at 0. However, this specific representation of the two-band model as a dot product of some vector $\boldsymbol{B}$ with a vector of Pauli matrices induces some very nice coincidences. Most importantly, we can obtain an expression for the first Chern number as a simple integral over some derivatives of $\boldsymbol{B}$. This is of great use, because it means that to compute the first Chern number of the two-band model we do not even have to compute any Berry curvatures. We only have to plug in the coefficients $\boldsymbol{B}$ that appear in the Hamiltonian into a simple formula.

First note that the eigenvalues of the Hamiltonian $h(p) = \boldsymbol{B}(p) \cdot \boldsymbol{\sigma}$ are $\pm||\boldsymbol{B}(p)||$. Therefore, h describes an insulator if and only if $B(p) := ||\boldsymbol{B}(p)|| > 0$ for all $p \in X$. In other words, to model an insulator one has to require $\boldsymbol{B} : X \to \mathbb{R}^3 \setminus \{0\}$. The space $\mathbb{R}^3 \setminus \{0\}$ is homotopy equivalent to S^2. We are interested in homotopy invariant properties and so, without loss of generality, we assume $\boldsymbol{B}$ is normalized such that $B(p) = 1$ for all $p \in X$. The Hamiltonian is thus fully determined by a map $\boldsymbol{B} : X \to S^2$. We have noted before that for $\dim X = 2$, complex line bundles over X are classified by $[X, \mathbb{C}P^1]$. Since $\mathbb{C}P^1$ is homeomorphic to S^2, one might suspect that $\boldsymbol{B}$ is the classifying map of the valence bundle of $h(p) = \boldsymbol{B}(p) \cdot \boldsymbol{\sigma}$. This suspicion turns out to be correct.

The valence bundle of h can be realized as the pullback of the tautological line bundle $\tau_{1,1} \to \mathbb{C}P^1$. The tautological line bundle $\tau_{1,1}$ itself actually corresponds to a physical system. We will argue that $\tau_{1,1}$ is the bundle of negative energy eigenstates of a spin-1/2 particle in a magnetic field, described by the family of Hamiltonians $H(\boldsymbol{B}) = \boldsymbol{B} \cdot \boldsymbol{\sigma}$ for $\boldsymbol{B} \in S^2$. It is important to stress that in the case of the spin-1/2 particle in a magnetic field, $\boldsymbol{B}$ itself as a unit vector in $\mathbb{R}^3$ already plays the role of the parameters in the system. The full Hilbert space of the spin-1/2 particle is $\mathbb{C}^2$, and so the parameterized quantum system is described by the trivial bundle $S^2 \times \mathbb{C}^2$. We let $E^H \subset S^2 \times \mathbb{C}^2$ denote the bundle of negative energy eigenstates of $H(\boldsymbol{B}) = \boldsymbol{B} \cdot \boldsymbol{\sigma}$. We will see that E^H corresponds to $\tau_{1,1}$ under the identification of S^2 with $\mathbb{C}P^1$.

In matrix notation, the Hamiltonian for a spin in a magnetic field takes the form

$$H(\boldsymbol{B}) = \begin{pmatrix} B_z & B_x - iB_y \\ B_x + iB_y & -B_z \end{pmatrix}.$$

Let $U = \{\boldsymbol{B} \in S^2 : B_z \neq 1\}$ and let $V = \{\boldsymbol{B} \in S^2 : B_z \neq -1\}$ be an open cover for S^2. The normalized negative energy eigenvectors u_- and v_- over U and V, respectively, are

$$u_-(\boldsymbol{B}) = \frac{1}{\sqrt{2B(B - B_z)}} \begin{pmatrix} -B + B_z \\ B_x + iB_y \end{pmatrix} \quad \text{and} \quad v_-(\boldsymbol{B}) = \frac{1}{\sqrt{2B(B + B_z)}} \begin{pmatrix} -B_x + iB_y \\ B + B_z \end{pmatrix}.$$

The homeomorphism from $S^2 \to \mathbb{C}P^1$ can be realized via stereographic projection. The complex projective space $\mathbb{C}P^1$ can be coverd by two charts

$$\widetilde{U} = \{(-1 : w) : w \in \mathbb{C}\} \quad \text{and} \quad \widetilde{V} = \{(w : 1) : w \in \mathbb{C}\}.$$

These charts are chosen so that they are orientation preserving with respect to the inward pointing orientation on S^2 and the standard orientation on $\mathbb{C}^2$. Orientation preserving stereographic projection from the north pole gives the homeomorphism $U \to \widetilde{U}$ by

$$U \ni (B_x, B_y, B_z) \mapsto \left(-1 : \frac{B_x + iB_y}{B - B_z}\right) \in \widetilde{U}.$$

Hence, the vector $u_-(\boldsymbol{B})$ lies in the line corresponding to $\boldsymbol{B}$ under the identification of S^2 with $\mathbb{C}P^1$ via sterographic projection. Similarly, the orientation preserving stereographic from the south pole gives

$$V \ni (B_x, B_y, B_z) \mapsto \left(\frac{-B_x + iB_y}{B + B_z} : 1\right) \in \widetilde{V}.$$

Again, the vector $v_-(\boldsymbol{B})$ lies in the line corresponding to $\boldsymbol{B}$ under the identification of S^2 with $\mathbb{C}P^1$. It is straightforward to check that the above identifications of S^2 with $\mathbb{C}P^1$ agree whenever $B_z \neq \pm 1$. Therefore, the bundle over S^2 of the negative energy eigenstates of $H(\boldsymbol{B}) = \boldsymbol{B} \cdot \boldsymbol{\sigma}$ is isomorphic to the tautological line bundle $\tau_{1,1} \to \mathbb{C}P^1$.

Let us then turn to the Hamiltonian $h(p) = \boldsymbol{B}(p) \cdot \boldsymbol{\sigma}$ describing the two-dimensional, two-band insulator. The fiber of the valence bundle over $p \in X$ can be identified with the fiber of the tautological line bundle $\tau_{1,1} \to \mathbb{C}P^1$ over the image of the point $\boldsymbol{B}(p)$ in $\mathbb{C}P^1$ under the homeomorphism $S^2 \to \mathbb{C}P^1$. Hence, the valence bundle of h is isomorphic to pullback bundle $\boldsymbol{B}^*\tau_{1,1}$. The map $\boldsymbol{B} : X \to S^2$ can thus indeed be identified with the classifying map for the valence bundle for $h : X \to \mathrm{Herm}(2, \mathbb{C})$.

The only topological invariant of the two-dimensional topological insulator is thus the homotopy class of $\boldsymbol{B} : X \to S^2$. The homotopy class can be computed by the degree of $\boldsymbol{B}$. However, we have argued above that insulators in two dimensions are fully classified by their first Chern number. As a consistency check, we show that the degree of $\boldsymbol{B}$ and the first Chern number coincide.

The valence bundle $E^h \to X$ of the two-band model is the pullback of the bundle $E^H \to S^2$ of the negative energy eigenstates of the spin in a magnetic field along the map $\boldsymbol{B} : X \to S^2$. This is pretty much a tautology. A more lowbrow way of expressing this tautology is that the negative energy eigenstate of $h(p)$ at $p \in X$ is the negative energy eigenstate of a spin subject to the magnetic field $\boldsymbol{B}(p)$. Now, if p runs around a loop in the momentum space X, then this corresponds to a loop on the magnetic field space S^2 via the map $\boldsymbol{B}$. The Berry phase that corresponds to this loop on the momentum space is precisely the Berry phase of the loop on S^2. The picture one should have in mind is that there is a curved bundle $E^H \to S^2$ corresponding to the spin in a magnetic field and that $h(p) = \boldsymbol{B}(p) \cdot \boldsymbol{\sigma}$ "probes" this bundle as a function of p. According to Eq. (4.11), the Berry connection A^H on E^H over the chart U is

$$A^H_U(\boldsymbol{B}) = \langle u_-(\boldsymbol{B}), du_-(\boldsymbol{B}) \rangle \,,$$

where d is the deRham differential on S^2 and is taking derivatives with respect to $\boldsymbol{B}$. The Berry connection A^h on E^h over $\boldsymbol{B}^{-1}(U)$ is

$$A^h(p) = \langle (u_- \circ \boldsymbol{B})(p)), d(u_- \circ \boldsymbol{B})(p) \rangle \,,$$

where now d is the deRham differential on X and is taking derivatives with respect to p. Similar relations hold over V and $\boldsymbol{B}^{-1}(V)$. The upshot is that $A^h = \boldsymbol{B}^* A^H$, where we use the pullback operation on one-forms. This relation carries through to the Berry curvature since d commutes with $\boldsymbol{B}^*$, i.e.

$$F^h = dA^h = d(\boldsymbol{B}^* A^H) = \boldsymbol{B}^*(dA^H) = \boldsymbol{B}^* F^H \,.$$

This is actually just a restatement of the fact that the first Chern class of E^h is indeed the pullback of the first Chern class of E^H along $\boldsymbol{B}$, which could have already concluded from the fact that $\boldsymbol{B}$ is the classifying map of E^h.

Recall that the first Chern number of the insulator h is $\frac{i}{2\pi}\int_X F^h$, where F^h is the Berry curvature of the valence bundle $E^h \to X$. Since $F^h = \boldsymbol{B}^* F^H$, it follows by definition of the degree of a map that

$$\int_X F^h = \int_{S^2} \boldsymbol{B}^* F^H = \deg \boldsymbol{B} \int_{S^2} F^H \,,$$

where $\deg \boldsymbol{B}$ is the degree of the map $\boldsymbol{B}$. Since $\boldsymbol{B} : X \to S^2$, one can think of $\deg \boldsymbol{B}$ as the number of times that X wraps around S^2 under $\boldsymbol{B}$. A straightforward computation establishes that $\int_{S^2} F^H = -2\pi i$. Therefore, the degree of $\boldsymbol{B}$ coincides with the first Chern number

$$C_1 = \frac{i}{2\pi} \int_X F^h = \deg \boldsymbol{B} \,. \tag{5.2}$$

Finally, an explicit formula for the $\deg \boldsymbol{B}$ can be obtained because

$$\deg \boldsymbol{B} = \frac{\int_X \boldsymbol{B}^* \omega}{\int_{S^2} \omega}$$

for *any* 2-form ω, not only the Berry curvature. In particular, we can pick the volume form $\omega = xdy \wedge dz - ydx \wedge dz + zdx \wedge dy$ on S^2. The integral $\int_{S^2} \omega = 4\pi$ just gives the area of the unit sphere. If we let BZ denote the Brillouin zone with coordinates (k_x, k_y), then

$$\boldsymbol{B}^* \omega|_{BZ} = \boldsymbol{B} \cdot (\partial_{k_x} \boldsymbol{B} \times \partial_{k_y} \boldsymbol{B}) dk_x \wedge dk_y \,.$$

Since the Brillouin zone covers the full momentum space X up to a set of measure zero,

$$C_1 = \deg \boldsymbol{B} = \frac{1}{4\pi} \int_{BZ} \boldsymbol{B} \cdot (\partial_{k_x} \boldsymbol{B} \times \partial_{k_y} \boldsymbol{B}) dk_x \wedge dk_y \,. \tag{5.3}$$

In conclusion, the two-band model $h(p) = \boldsymbol{B}(p) \cdot \boldsymbol{\sigma}$ over a two-dimensional space is fully classified by the first Chern number, and Eq. (5.3) gives an expression for the first Chern number purely in terms of the parameters $\boldsymbol{B}(p)$ that appear in the Hamiltonian.

6 Bulk-boundary correspondence

On the interface between two topologically inequivalent insulators, there must exists gapless edge states. This is the *bulk-boundary correspondence*. Let us restrict ourselves to two-dimensional insulators. The interface between a pair of two-dimensional insulators is a one-dimensional system. The gapless edge states on the one-dimensional interface turn out to be massless chiral fermions. The quantitative version of the bulk-boundary correspondence states that the difference between the number of right and left chiral fermions on the interface equals the difference in Chern numbers between the two insulators. Proofs of the bulk-boundary correspondence using Green's functions are given in [7, 8].

In this section, we derive the bulk-boundary correspondence in the specific case of a two-band insulator in two dimensions using the so-called chiral anomaly. This argument is inspired by [9]. We first introduce the Chern-Simons theory as the effective low-energy theory for the linear response of the two-band model. The Chern-Simons theory is not gauge invariant in the presence of a spatial boundary, which indicates a non-conservation of electric current. A one-dimensional theory of massless chiral fermions also has a non-conserved electric current. It turns out that combining the bulk Chern-Simons theory with the massless chiral fermions on the boundary restores the conservation of electric current in the full system a topological insulator with boundary. The existence of the massless chiral edge states is also explicitly verified in a simple Dirac model with a spatially dependent mass term that models the interface between a trivial and a non-trivial insulator.

6.1 Effective Chern-Simons theory

In Section 4.3 we have derived the linear response of a two-dimensional, two-band bulk insulator to an external electric field, in the limit of low energy, low temperature and long wavelength, to be

$$J^i_{\text{bulk}} = \sigma_H \epsilon^{ij} E_j \,, \tag{6.1}$$

where $i, j = x, y$ and $\sigma_H = C_1 \frac{e^2}{h}$ is the Hall conductivity with C_1 denoting the first Chern number. This result has been obtained through a path integral calculation in a gauge with $A_0 = 0$. Nonetheless, the result is true in any gauge (see Appendix A). The continuity

equation for electric charge in the bulk leads to the following equation for the charge density ρ_{bulk} of the system

$$\partial_t \rho_{\text{bulk}} = -\partial_i J^i_{\text{bulk}} = -\sigma_H \epsilon^{ij} \partial_i E_j = \sigma_H \partial_t B \,,$$

where B is the magnetic field perpendicular to the insulator. This implies $\rho_{\text{bulk}}(B) - \rho_0 = \sigma_H B$, where ρ_0 is the charge density in the ground state for $B = 0$ [4]. Absorbing ρ_0 into $\rho_{\text{bulk}}(B)$ the two bulk response equations

$$\rho_{\text{bulk}} = \sigma_H B \quad \text{and} \quad J^i_{\text{bulk}} = \sigma_H \epsilon^{ij} E_j$$

together describe a bulk theory with charge conservation. They can be combined into one covariant equation

$$J^\mu_{\text{bulk}} = \sigma_H \epsilon^{\mu\nu\rho} \partial_\nu A_\rho \,, \tag{6.2}$$

where $\mu, \nu, \rho = t, x, y$. The effective action that describes the bulk electromagnetic response Eq. (6.2) of the insulator is the Chern-Simons action

$$S_{CS}[A] = \frac{\sigma_H}{2} \int_{\mathbb{R}\times\Sigma} dt d^2x \, \epsilon^{\mu\nu\rho} A_\mu \partial_\nu A_\rho \,, \tag{6.3}$$

where Σ denotes the spatial extent of the insulator in the plane.

The Chern-Simons action is not obviously gauge invariant because it depends directly on the potential A_μ. In fact, under a gauge transformation $A_\mu \to A_\mu + \partial_\mu \chi$, the action changes by a total derivative

$$\delta_\chi S_{CS}[A_\mu] = \frac{\sigma_H}{2} \int_{\mathbb{R}\times\Sigma} dt d^2x \, \epsilon^{\mu\nu\rho} \partial_\mu (\chi \partial_\nu A_\rho) \,.$$

To model the bulk of an insulator we take $\Sigma = \mathbb{R}^2$ and we assume that the fields vanish at spatial infinity. The bulk theory is thus gauge invariant. However, if the insulator is only extended in a finite region Σ with a nonempty boundary, then gauge invariance fails when χ is nonzero on the boundary. By Stokes' theorem,

$$\delta_\chi S_{CS}[A_\mu] = \frac{\sigma_H}{2} \int_{\mathbb{R}\times\Sigma} dt d^2x \, \epsilon^{\mu\nu\rho} \partial_\mu (\chi \partial_\nu A_\rho) = \frac{\sigma_H}{2} \int_{\mathbb{R}\times\partial\Sigma} dt d\tilde{x} \, \hat{n}_\mu \epsilon^{\mu\nu\rho} \chi \partial_\nu A_\rho \,,$$

where $\hat{n}_\mu$ is the normal vector pointing out of Σ and $\tilde{x}$ is the coordinate on the boundary $\partial\Sigma$ induced from the coordinates x on Σ.

The failure of the effective Chern-Simons theory to be gauge invariant at the boundary spurs us to reconsider the conservation of the current in Eq. (6.2). The key observation of [9] is that spatial dependence of the Hall conductivity σ_H spoils the conservation of the current, because then

$$\partial_\mu J^\mu_{\text{bulk}} = (\partial_\mu \sigma_H) \epsilon^{\mu\nu\rho} \partial_\nu A_\rho \,.$$

For a bulk theory, σ_H is constant and so the current is still conserved. However, at the interface between two insulators with different Chern number σ_H changes. Therefore, at the interface, the current is not conserved. For example, let us consider a situation where $\Sigma \subset \mathbb{R}^2$ corresponds to the spatial extent of an insulator with Chern number C_1 and $\mathbb{R}^2 \setminus \Sigma$ corresponds to the vacuum with Chern number 0. At the boundary $\partial\Sigma$, the Hall conductivity drops from $\sigma_H = C_1 \frac{e^2}{h}$ to zero, so that near the boundary the derivative $\partial_\mu \sigma_H$ takes the form of a δ-function times the unit normal vector to the boundary. The violation of current conservation is thus

$$\int_{\mathbb{R}\times\Sigma} dt d^2x \, \partial_\mu J^\mu = \int_{\mathbb{R}\times\partial\Sigma} dt d\tilde{x} \, \sigma_H \hat{n}_\mu \epsilon^{\mu\nu\rho} \partial_\nu A_\rho = \int_{\mathbb{R}\times\partial\Sigma} dt d\tilde{x} \, C_1 \frac{e^2}{h} \epsilon^{\tilde{\mu}\tilde{\nu}} \partial_{\tilde{\mu}} A_{\tilde{\nu}} \,. \tag{6.4}$$

In order for current conservation to be restored, we need a theory on the boundary that satisfies

$$\partial_{\tilde{\mu}} J^{\tilde{\mu}}_{\text{boundary}} = -C_1 \frac{e^2}{h} \epsilon^{\tilde{\mu}\tilde{\nu}} \partial_{\tilde{\mu}} A_{\tilde{\nu}} \,, \tag{6.5}$$

where the index $\tilde{\mu}$ runs over t and the boundary coordinate $\tilde{x}$, so as to cancel the integrand on the right hand side of Eq. (6.4). Note that $\epsilon^{\tilde{\mu}\tilde{\nu}} \partial_{\tilde{\mu}} A_{\tilde{\nu}}$ is just the component of an external electric field parallel to the boundary of Σ.

6.2 Restoring current conservation

When quantizing the theory of massless $1+1$-dimensional fermions there is a so-called *chiral anomaly*. Massless fermions in $1+1$ dimension come in two different chirality species: a left moving and a right moving species. The chiral anomaly is generally associated to a non-conservation of the chiral current, which is the difference between right and left moving currents. The form of the chiral anomaly is very similar to Eq. (6.5). It is given by

$$\partial_{\tilde{\mu}} J_5^{\tilde{\mu}} = \frac{2e^2}{h} \epsilon^{\tilde{\mu}\tilde{\nu}} \partial_{\tilde{\mu}} A_{\tilde{\nu}} \,,$$

where we let $\tilde{\mu}$ label the spacetime indices of the $1+1$-dimensional theory and J_5 denotes the chiral current. This does not look immediately useful, however, because we are looking for a non-conserved electric current, not a chiral current. But, when reducing to just one chirality species, say right chirality, it is actually the electric current which becomes non-conserved precisely following the requirement of Eq. (6.5) in the case $C_1 = 1$. Similarly, restricting to left chirality yields Eq. (6.5) in the case $C_1 = -1$. Appropriate combinations of copies of left and right massless chiral fermions allows for the restoration of current cosnservation for any difference in Chern numbers.

Heuristically, the picture is as follows. In one spatial dimension, massless chiral fermions are massless fermions that can only move in one direction. There are right chiral fermions, that only move to the right, and, similarly, there are left chiral fermions that only move to the left. Let us focus on right chiral fermions. The dispersion relation for a right chiral fermion is $\epsilon(k) = v_f k$, where $v_f > 0$ is the speed of the fermion inside the material. Imagine chiral fermions in a one-dimensional box of length L with periodic boundary conditions. The Hamiltonian is

$$H = -i\hbar v_f \partial_x \,.$$

The eigenstates of H are $|k\rangle = \frac{1}{\sqrt{L}} e^{ikx}$ for $k = \frac{2\pi}{L} n$ with $n \in \mathbb{Z}$. In order to study $\partial_{\tilde{\mu}} J^{\tilde{\mu}} = \partial_t \rho + \partial_x j$ for the chiral fermion, where ρ is the charge density and j is the current density, consider applying a constant external electric field E in the negative x-direction. Since the electric field is uniform in space, $\partial_x J$ will be zero. The interesting behaviour is contained in the time derivative of the charge density.

We choose a gauge such that $A_t = 0$ and $A_x = -Et$. After minimal substitution, the Hamiltonian becomes time dependent and takes the form

$$H(t) = -i\hbar v_f (\partial_x + i \frac{e}{\hbar} Et) \,.$$

Let us assume that E is small enough so that $H(t)$ can be viewed as an adiabatic deformation of H. The instantaneous eigenstates of $H(t)$ are the wavefunctions $|k\rangle$ for $k = \frac{2\pi}{L} n$ with $n \in \mathbb{Z}$ and the instantaneous spectrum of $H(t)$ is

$$\sigma(t) = \left\{ \hbar v_f \left(\frac{2\pi}{L} n + \frac{e}{\hbar} Et \right) : n \in \mathbb{Z} \right\} .$$

At $t = 0$, the instantaneous ground state corresponds to all states with nonpositive momentum being filled, see Fig. 5. The ground state is gapped out from the rest of the spectrum. By the adiabatic theorem, the system will thus remain in the instantaneous ground state as time progresses. After $t = \frac{2\pi}{L}\frac{\hbar}{eE}$, all the filled eigenstates $|k\rangle$ will have shifted to $|k + \frac{2\pi}{L}\rangle$. Within this model of a single one-dimensional system, one can interpret this as a single unit of charge coming from the "Dirac sea" that has entered the system, see Fig. 5 . A less confabulatory explanation of this effect is possible when we consider the one-dimensional theory as being just one of two edges of a rectangular piece of insulator along which the electric field is parallel. The unit charge appearing on one edge is compensated by a unit of charge disappearing from the opposite edge. Focusing

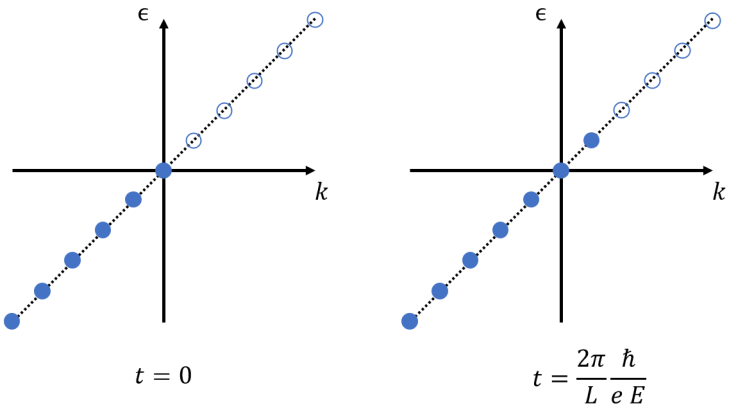

Figure 5: Instantaneous ground states of $H(t) = -iv_f(\partial_x + i\frac{e}{\hbar}Et)$ before and after adiabatic evolution from $t = 0$ to $t = \frac{2\pi}{L}\frac{\hbar}{eE}$.

again on one edge, the change after $\Delta t = \frac{2\pi}{L}\frac{\hbar}{eE}$ in the charge density is $\Delta\rho = -e/L$. Hence, in the limit $L \to \infty$, the time derivative of the charge density is

$$\partial_t \rho = \lim_{L\to\infty} \frac{\Delta\rho}{\Delta t} = -\frac{e^2}{h}E\,.$$

We conclude that the massless right chiral fermion satisfies

$$\partial_{\tilde{\mu}} J^{\tilde{\mu}} = \partial_t \rho = -\frac{e^2}{h}E = -\frac{e^2}{h}\epsilon^{\tilde{\mu}\tilde{\nu}}\partial_{\tilde{\mu}} A_{\tilde{\nu}}\,.$$

A right moving chiral fermion on the one-dimensional edge can thus compensate for the non-conserved charge on the edge of a two-dimensional Hall insulator with Chern number $C_1 = 1$.

The dispersion for a left moving chiral fermion is $\epsilon(k) = -v_f k$. The same calculation as above shows that a left chiral fermion has to appear on the edge of a Hall insulator with $C_1 = -1$. For general $C_1 \in \mathbb{Z}$, an appropriate number of species of left or right chiral fermions on the edge restores current conservation. For example, if $C_1 = -3$, then there have to be three left moving chiral fermion species on the edge.

6.3 Chiral fermions in Dirac model with domain wall

The current conservation argument layed out above shows that at the boundary of a non-trivial topological insulator in two spatial dimensions there should be massless chiral

fermions. We illustrate this principle for the interface between continuum two-band Dirac models with a mass parameter m, which are described by the Hamiltonian

$$h(k_x, k_y) = k_x\sigma_x + k_y\sigma_y + (m + \eta k^2)\sigma_z = \begin{pmatrix} m + \eta k^2 & k_x - ik_y \\ k_x + ik_y & -m - \eta k^2 \end{pmatrix}, \tag{6.6}$$

for $(k_x, k_y) \in \mathbb{R}^2$ and $\eta > 0$ a regularization constant. The regularization constant is needed to ensure that the eigenvectors at $\boldsymbol{k} = \infty$ agree so that the momentum space $\mathbb{R}^2$ can be compactified to S^2. For $m \neq 0$ Eq. (6.6) describes an insulator and the Chern number of h can be computed using the winding number formula Eq. (5.3) to be

$$C_1 = \begin{cases} 0 & \text{if } m < 0 \\ 1 & \text{if } m > 0 \end{cases}.$$

The interface between a trivial and a non-trivial insulator can be modelled by making the mass in the Dirac model spatially dependent. In particular, consider a "domain wall" structure for the mass given by $m(x) = \mathrm{sgn}(x)m$, where $m > 0$ is fixed. For $x < 0$, this models a trivial insulator and for $x > 0$ this models a non-trivial insulator with $C_1 = 1$. According to the bulk-boundary correspondence, there should thus be one right chiral fermion species moving in the positive y-direction on the interface at $x = 0$. We will argue its existence here directly from the solutions to the time-independent Schrödinger equation.

First of all, in the low-energy limit, the massless chiral edge states are to be found near $\boldsymbol{k} = 0$. The topological nature of the Chern number implies that the regularization parameter η in Eq. (6.6) can be made arbitrarily small, as long as it remains nonzero, without changing the Chern numbers. Near $\boldsymbol{k} = 0$, it then suffices to study the non-regularized Hamiltonian

$$h(k_x, k_y) = \begin{pmatrix} m & k_x - ik_y \\ k_x + ik_y & -m \end{pmatrix}, \tag{6.7}$$

since its eigenvectors differ from those of Eq. (6.6) at order $O(\eta k^2)$. Secondly, if we introduce a mass term that explicitly depends on x, then translational invariance in x-direction is broken. The k_x dependence of the bulk model Eq. (6.7) thus has to be reverted to $-i\partial_x$ as the Fourier transform in the x-direction is no longer valid.

The above considerations imply that we should study the spectrum of the one-dimensional domain wall Hamiltonian

$$h(x, k_y) = \begin{pmatrix} m(x) & -i\partial_x - ik_y \\ -i\partial_x + ik_y & -m(x) \end{pmatrix}, \tag{6.8}$$

where k_y, which is the momentum along the interface, is treated as a parameter. We will look for energy eigenstates of the form $\psi(x, y) = u(x, k_y)e^{ik_y y}$ that are plane waves in the y-direction. This yields a family of one-dimensional eigenproblems parameterized by k_y,

$$h(x, k_y)u(x, k_y) = \epsilon(k_y)u(x, k_y). \tag{6.9}$$

We are interested in finding solutions to Eq. (6.9) that are localized near $x = 0$ because those localized solutions are the edge states. Hence, we fix a k_y and look for solutions $u(x, k_y)$ of the one-dimensional problem that decay exponentially away from zero. The solutions $u(x, k_y)$ are required to be continuous, but there will be a discontinuity in the first derivative because $m(x)$ is discontinuous at $x = 0$ and $h(x, k_y)$ contains only first

derivatives in x. We shall thus look for exponentially decaying eigenvectors for $x < 0$ and $x > 0$ separately, and then restrict ourselves to solutions that can be glued continuously at $x = 0$. Many decaying solutions exist for $x < 0$ and $x > 0$ separately, but it is only possible to glue the solutions for $x < 0$ and $x > 0$ continuously at $x = 0$ for some special energy eigenvalues. By finding the energies for which decaying solutions glue, we end up with the spectrum of a chiral edge fermion. The details are worked out below.

For notational convenience, we drop the explicit k_y dependence from the notation for now. For $x < 0$ we take the ansatz $u_L(x) = e^{\lambda_L x} v_L$ for some $\lambda_L > 0$ and v_L a two-component spinor. Plugging the ansatz $u_L(x)$ into Eq. (6.9), we see that the spinor v_L has to satisfy

$$\begin{pmatrix} -m-\epsilon & -i\lambda_L - ik_y \\ -i\lambda_L + ik_y & m-\epsilon \end{pmatrix} v_L = 0\,. \tag{6.10}$$

A non-trivial solution for v_L exists if and only if the determinant of this matrix in Eq. (6.10) is zero. Hence, non-trivial solutions for v_L exist if and only if $\lambda_L = \sqrt{m^2 + k_y^2 - \epsilon^2}$. For this λ_L, we have that

$$v_L = \begin{pmatrix} -i\lambda_L - ik_y \\ m+\epsilon \end{pmatrix}$$

is a solution to Eq. (6.10). Similarly, for $x > 0$ we take the ansatz $u_R(x) = e^{-\lambda_R x} v_R$ with $\lambda_R > 0$. The spinor v_R has to satisfy

$$\begin{pmatrix} m-\epsilon & i\lambda_R - ik_y \\ i\lambda_R + ik_y & -m-\epsilon \end{pmatrix} v_R = 0\,. \tag{6.11}$$

Again, non-trivial solutions for v_R exists if and only if $\lambda_R = \sqrt{m^2 + k_y^2 - \epsilon^2}$. A solution to Eq. (6.11) is

$$v_R = \begin{pmatrix} m+\epsilon \\ i\lambda_R + ik_y \end{pmatrix}.$$

We now check for which values of ϵ we can combine u_L and u_R to obtain a wavefunction that is continuous at $x = 0$. First note that the solutions v_L and v_R above are not unique. Any nonzero complex multiple of v_L and v_R is also a valid solution. The norms of the v_L and v_R given above are the same, so we need not normalize these spinors further. However, we might have to alter the phase to be able to glue the solutions at $x = 0$. In other words, we want to find ϵ such that $v_L = e^{i\varphi} v_R$ for some $\varphi \in [0, 2\pi]$.

Note that $\lambda_L = \lambda_R$, so we can drop the subscripts and just write λ. We first need to know for which values of relative phase φ the equation

$$\begin{pmatrix} -i\lambda - ik_y \\ m+\epsilon \end{pmatrix} = v_L = e^{i\varphi} v_R = e^{i\varphi} \begin{pmatrix} m+\epsilon \\ i\lambda + ik_y \end{pmatrix}$$

can hold. The equality of the first components means $\lambda + k_y = ie^{i\varphi}(m+\epsilon)$. Using the equality $m + \epsilon = ie^{i\varphi}(\lambda + k_y)$ due to the second component, we conclude that

$$e^{i2\varphi} = -1\,.$$

Hence, $v_L = e^{i\varphi} v_R$ can only hold for $\varphi = \pi/2$ or $\varphi = 3\pi/2$. For $\varphi = \pi/2$, we need to solve

$$\begin{pmatrix} -i\lambda - ik_y \\ m+\epsilon \end{pmatrix} = \begin{pmatrix} im + i\epsilon \\ -\lambda - k_y \end{pmatrix}$$

for ϵ. Both components lead to the same equation $\lambda = -m-\epsilon-k_y$. Squaring this equality and recalling the definition of λ, we find

$$m^2 + k_y^2 - \epsilon^2 = m^2 + \epsilon^2 + k_y^2 + 2m\epsilon + 2mk_y + 2\epsilon k_y\,,$$

hence,

$$\epsilon^2 + \epsilon(m + k_y) + mk_y = 0\,.$$

The solutions for ϵ are thus

$$\begin{aligned}\epsilon_\pm(\varphi = \pi/2) &= \frac{1}{2}\left(-(m+k_y) \pm \sqrt{(m+k_y)^2 - 4mk_y}\right) \\ &= -\frac{1}{2}(m+k_y) \pm \frac{1}{2}|m-k_y| = \begin{cases} \frac{-m\pm m}{2} + \frac{-k_y \mp k_y}{2} & \text{if } k_y \le m \\ \frac{-m\mp m}{2} + \frac{-k_y \pm k_y}{2} & \text{if } k_y > m \end{cases}.\end{aligned}$$

A similar computation applies to $\varphi = 3\pi/2$. In this case, we need to solve

$$\begin{pmatrix} -i\lambda - ik_y \\ m + \epsilon \end{pmatrix} = \begin{pmatrix} -im - i\epsilon \\ \lambda + k_y \end{pmatrix},$$

leading to $\lambda = m + \epsilon - k_y$. Squaring the equation and substituting the definition of λ^2 yields

$$\epsilon_\pm(\varphi = 3\pi/2) = \begin{cases} \frac{-m\mp m}{2} + \frac{k_y \mp k_y}{2} & \text{if } k_y \le -m \\ \frac{-m\pm m}{2} + \frac{k_y \pm k_y}{2} & \text{if } k_y > -m \end{cases}.$$

The various branches of possible solutions for ϵ as a function of k_y for the two angles

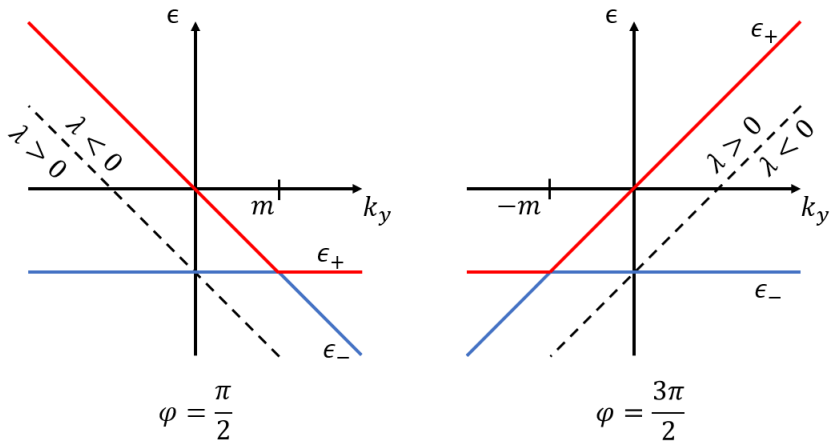

Figure 6: Energies ϵ for which v_L and v_R agree at $x = 0$.

$\varphi = \pi/2$ and $\varphi = 3\pi/2$ are plotted in Fig. 6. Not all branches constitute valid solutions though. Indeed, λ has to be positive and the spinors v_L and v_R should be continuous in k_y. The regions of positive and negative λ have been indicated in Fig. 6. For $\varphi = \pi/2$, there is no continuous solution for ϵ as a function of k_y that satisfies $\lambda > 0$. For $\varphi = 3\pi/2$, it is clear that at least for $k_y > -m$, the ϵ_+ branch consitutes a valid solution. We have to deal, however, with the peculiar behaviour at $k_y = -m$.

The issues that occur at $k_y = -m$ become apparent when we try to normalize the spinor $v_L(k_y)$. Normalizing and using $\lambda + k_y = m + \epsilon$ as required for $\varphi = 3\pi/2$, yields

$$\hat{v}_L(k_y) = [(\lambda + k_y)^2 + (m + \epsilon(k_y))^2]^{-1/2} \begin{pmatrix} -i\lambda - ik_y \\ m + \epsilon(k_y) \end{pmatrix} = \frac{1}{\sqrt{2}}\mathrm{sgn}(m + \epsilon(k_y)) \begin{pmatrix} -i \\ 1 \end{pmatrix}, \quad (6.12)$$

where $\mathrm{sgn}(x) = x/|x|$. There is a possible problem when $\epsilon(k_y) = -m$, because there the sign function is not well-defined. The solution branch $\epsilon_+(\varphi = 3\pi/2)$ for $k_y < -m$ is constantly $-m$. We cannot define $\mathrm{sgn}(m + \epsilon(k_y))$ unambiguously as a function of k_y on this branch. We could try to take the limit of $\epsilon(k_y) \to -m$, but depending on the direction from which we approach $-m$, this leads to different results for $\mathrm{sgn}(m+\epsilon(k_y))$. Hence, the phase of $v_L(k_y)$ is ill-defined on this branch and we thus discard it.

Our only possibility is then to take the $\epsilon_+(\varphi = 3\pi/2)$ solution for $k_y > -m$ and the $\epsilon_-(\varphi = 3\pi/2)$ solution for $k_y < -m$. The issue with the sign function when $\epsilon(k_y) = -m$ seemingly still persists when $k_y = -m$ though. However, well-definition of $v_L(k_y)$ at $k_y = -m$ can be ensured by continuously extending the solution of $v_L(k_y)$ from the ϵ_+ and ϵ_- branch. This is not completely obvious since in Eq. (6.12) it seems as though $v_L(k_y)$ has to flip sign when passing through $k_y = -m$. The problem is resolved by recalling that we have only fixed the *relative* phase of v_L and v_R, but never the *absolute* phases. Therefore, the ostensible phase flip in $v_L(k_y)$ when k_y goes through $-m$ can be remedied by altering the phase of our original, arbitrary, choice for the phase of $v_L(k_y)$. The final solution to Eq. (6.9) is then

$$u(x, k_y) = \begin{cases} \frac{1}{\sqrt{2}} \begin{pmatrix} -i \\ 1 \end{pmatrix} e^{-mx} & \text{if } x < 0 \\ \frac{1}{\sqrt{2}} \begin{pmatrix} -i \\ 1 \end{pmatrix} e^{mx} & \text{if } x > 0 \end{cases}$$

with dispersion $\epsilon(k_y) = k_y$. This corresponds precisely to the one right moving chiral fermion the bulk-boundary correspondence predicts.

Part II

Topological Insulators with time-reversal symmetry

7 Topological insulators with time-reversal symmetry

Two insulators have been defined to be topologically equivalent whenever their momentum space Hamiltonians can be continuously deformed into each other while preserving the insulating gap. There is a subset of insulator consisting of those insulators that are invariant under the operation of time reversal. These are the so-called *time reversal symmetric insulators*. In less than or equal to three spatial dimensions, a time reversal symmetric insulator can always be continuously deformed to a trivial insulator while preserving the gap. One would therefore think that these systems are of no topological interest. However, if one requires the deformation to preserve the time-reversal symmetry, then it turns out that some insulators cannot be deformed into the trivial insulator anymore. This behaviour is encoded in a $\mathbb{Z}_2$ invariant due to Fu-Kane-Mele (FKM) [12, 14, 15], which will be reviewed in this section.

7.1 time-reversal symmetry

Time reversal is a discrete operation that reverses the direction of the time parameter in a dynamical theory. In quantum mechanics, it turns out that time reversal is represented on the Hilbert space by an *anti*unitary operator $T : \mathcal{H} \to \mathcal{H}$. For the single-particle physics of condensed matter systems embedded in three-dimensional space, the time-reversal operator is explicitly realized as $T = e^{i\pi S_y/\hbar}K$, where S_y is the spin operator in the y direction and K is the complex conjugation operator. We briefly review these facts here following [35, 36], and interpret them in the context of insulators.

Consider first a spinless quantum-mechanical particle with Hilbert space $\mathcal{H}$ and time evolution described by a Hamiltonian H. Suppose that $|\psi(t)\rangle$ solves the Schrödinger equation

$$i\frac{d}{dt}|\psi(t)\rangle = H|\psi(t)\rangle \ . \tag{7.1}$$

Simply reversing the direction of time by replacing t by $-t$ yields

$$-i\frac{d}{dt}|\psi(-t)\rangle = H|\psi(-t)\rangle \ . \tag{7.2}$$

Now, Eq. (7.2) is not the Schrödinger equation for the Hamiltonian H because of the minus sign on the left-hand side. However, we can take the complex conjugate of Eq. (7.2) to obtain

$$i\frac{d}{dt}|\psi(-t)\rangle^* = H^*|\psi(-t)\rangle^* \ , \tag{7.3}$$

which is the Schrödinger equation for the complex conjugated Hamiltonian H^*. In general, a dynamical theory is called time-reversal symmetric if, for any solution to the equations of motion, traversing the solution in the negative time direction is also a solution to the equations of motion. Using this criterion, the quantum-mechanical theory is to be called time-reversal symmetric if $H^* = H$, since in this case Eq. (7.3) shows that the reversed solution $|\psi_R(t)\rangle := |\psi(-t)\rangle^*$ is a solution to the original Schrödinger equation Eq. (7.1). Note that although the wavefunction of the reversed solution does not retrace the original wavefunction due to the additional complex conjugation, it does retrace the physical state of the original solution. More precisely, if $|\psi(t)\rangle$ and $|\phi(t)\rangle$ are two physical evolutions, then their overlap at time $-t$ is the same as the overlap at time t of the reversed evolutions $|\psi_R(t)\rangle$ and $|\phi_R(t)\rangle$ according to

$$|\langle\psi_R(t)|\phi_R(t)\rangle|^2 = |\langle\psi(-t)|\phi(-t)\rangle^*|^2 = |\langle\psi(-t)|\phi(-t)\rangle|^2 .$$

The most general operation on the state $|\psi(t)\rangle$ that has the property that it preserves the overlaps when reversing the direction of time is $T = U_T K$, where $U_T \in U(\mathcal{H})$ is some unitary and K is the complex conjugation operator. The unitary U_T depends on the type of system under consideration. The reality condition $H^* = H$ in the general setting becomes the condition that $[H, T] = 0$, or equivalently, $H = U_T H^* U_T^\dagger$.

For a particle with spin s, a useful representation of T on the Hilbert space $\mathbb{C}^{2s+1}$ is

$$T = e^{i\pi S_y/\hbar} K\,,$$

where S_y is the spin in the y-direction. For a spin-1/2 particle, $S_y = \frac{\hbar}{2}\sigma_y$ so that $T = i\sigma_y K$. One can check explicitly that in this case T has the effect of flipping the direction of a spin, which is precisely what we would expect when we flip the direction of time. The reason for rotating around the y-axis is that S_y is by convention completely imaginary so that $e^{i\pi S_y/\hbar}$ commutes with K. It is then easy to check that for integer spins $T^2 = 1$ and for half integer spins $T^2 = -1$. Indeed, $T^2 = e^{i2\pi S_y/\hbar}$ is a 2π-radians rotation around the y-axis so that a particle of spin s picks up a phase $e^{i2\pi s}$ under T^2.

In a tight-binding model with M bands the time-reversal operator is represented on the effective fiber Hilbert space $\mathbb{C}^M$ as $T = U_T K$ for some unitary $U_T \in U(M)$, which depends on the model. Because in physical space the operation of time reversal flips the sign of time, it flips the sign of momentum in reciprocal space. If the momentum space is X, then this flipping of the sign of momentum can be encoded in a map $\tau : X \to X$ that satisfies $\tau^2 = \mathrm{id}$. In other words, when applying the time-reversal operation, we not only act with T on the Hilbert space, but we also act with τ on the momentum space. In periodic models, the momentum space is the torus. Viewing the torus as a square with its edges identified, the action of τ is given by $[\boldsymbol{k}] \mapsto [-\boldsymbol{k}]$, where the square brackets indicate equivalence classes due to the identification of the boundary. In the continuum model, the momentum space $\mathbb{R}^d$ comes equipped with the action $\tau : \boldsymbol{k} \mapsto -\boldsymbol{k}$. After compactification to S^d this becomes

$$\tau : (k_1, \ldots, k_d, k_{d+1}) \mapsto (-k_1, \ldots, -k_d, k_{d+1})\,,$$

where the compactification happens along the $d+1$-th direction. Properly taking into account the action of time reversal on the momentum space of a condensed matter system modelled by a tight binding Hamiltonian $h : X \to \mathrm{Herm}(M, \mathbb{C})$, the condition for h to be time-reversal invariant can be stated as

$$h(\tau(p)) = T h(p) T^{-1} = U_T h(p)^* U_T^\dagger \quad \text{for all } p \in X\,. \tag{7.4}$$

We end our preliminary discussion on time-reversal symmetry by deriving the so-called *Kramers' degeneracy*. Let $h : X \to \mathrm{Herm}(M, \mathbb{C})$ be a tight-binding model for an insulator, let $\tau : X \to X$ be the time reversal involution on the momentum space and let $T : \mathbb{C}^M \to \mathbb{C}^M$, given by $T = U_T K$, be the time-reversal operator on the effective Hilbert space. Assume that h is time reversal symmetric. If $\psi \in \mathbb{C}^M$ is an eigenstate of $h(p)$ with energy $\epsilon(p)$, i.e. if $h(p)\psi = \epsilon(p)\psi$, then by Eq. (7.4)

$$h(\tau(p)) T\psi = \epsilon(p) T\psi\,.$$

Hence, $T\psi$ is an eigenstate of $h(\tau(p))$ with energy $\epsilon(p)$. The spectrum of h as a function of p is thus symmetric under τ in the sense that for any solution $\epsilon : X \to \mathbb{R}$ of $\det(h - \epsilon)$, the function $\epsilon \circ \tau$ is also a solution. At a *time-reversal invariant momentum* $p \in X$, meaning $p = \tau(p)$, the spectrum of h thus to degenerates. Moreover, at a time-reversal invariant

momentum it makes sense to compute the inner product of ψ and $T\psi$ because they live in the same fiber. Using the property of the antiunitary operator T that $\langle T\psi, T\varphi\rangle = \langle \varphi, \psi\rangle$ and the fact that for fermions $T^2 = -1$, it follows that

$$\langle \psi, T\psi\rangle = \langle T^2\psi, T\psi\rangle = -\langle \psi, T\psi\rangle$$

so that $\langle \psi, T\psi\rangle = 0$. The spectrum of h is thus not only degenerate at the time reversal momentum, but any eigenvector ψ is orthogonal to $T\psi$. This is Kramers' degeneracy.

7.2 The first Chern number is zero

We have seen that two-dimensional insulators are classified by their first Chern number. Below, we prove that the first Chern number of a two-dimensional, time reversal symmetric insulator always vanishes.[10]

Let $h : X \to \mathrm{Herm}(M, \mathbb{C})$ be a time reversal symmetric insulator. Recall that the first Chern number can be computed as

$$C_1 = \frac{i}{2\pi}\int_X dk^1 dk^2 \,\mathrm{Tr}[F_{12}]\,, \tag{7.5}$$

where F_{12} is the $(1,2)$-component of the Berry curvature with respect to coordinates (k^1, k^2) on the momentum space X. The spectral projector P onto the valence states can be expressed pointwise through

$$P(p) = \frac{1}{2\pi i}\int_C dz\,[h(p) - z]^{-1}\,,$$

where the curve C encloses the negative real axis once in the counterclockwise direction. A coordinate free form of the Berry curvature is $F = PdP \wedge dPP$. Its $(1,2)$-component can thus be written as $F_{12} = P[\partial_1 P, \partial_2 P]P$. The time-reversal symmetry of h given in Eq. (7.4) carries through to P via

$$P(\tau(p)) = \frac{1}{2\pi i}\int_C dz[Th(p)T^{-1} - z]^{-1} = -T^{-1}\left(\frac{1}{2\pi i}\int_C d\bar{z}[h(p) - z]^{-1}\right)T = T^{-1}P(p)T\,.$$

We have picked up two minus signs in this computation. One due to pulling $\frac{1}{2\pi i}$ into T^{-1}. The other due to pulling dz into T^{-1}, which turns the integral into $\int_C d\bar{z} = \int_{\overline{C}} dz = -\int_C dz$, where $\overline{C}$ is orientation reversed version of the curve C. The pointwise relation $P(\tau(p)) = T^{-1}P(p)T$ can be restated as relation of maps $X \to L(\mathbb{C}^M)$ by

$$P \circ \tau = T^{-1}PT\,. \tag{7.6}$$

Note that by the chain rule $d(P\circ\tau)_p = dP_{\tau p}d\tau_p$, which implies $dP_{\tau p} = d(P\circ\tau)_p d\tau_p^{-1} = d(P\circ\tau)_p d\tau_{\tau p}$, where we have used that τ is its own inverse. For the physical case of $X = S^2$ or $\mathbb{T}^2$, there are standard coordinates (k^i) on X that cover X up to a set of measure zero. Moreover, in these coordinates, τ is given by $\boldsymbol{k} \mapsto -\boldsymbol{k}$, so that $\partial_i \tau^j(p) = -\delta_i^j$ for all $p \in X$. We thus compute

$$\partial_i P(\tau(p)) = \partial_j(P\circ\tau)(p)\partial_i\tau^j(\tau(p)) = -\partial_i(P\circ\tau)(p) = -T^{-1}\partial_i P(p)T\,.$$

[10] The proof is inspired on the treatment in [11].

This result can be used to relate $\mathrm{Tr}[F_{12}(\tau(p))]$ to $\mathrm{Tr}[F_{12}(p)]$ through

$$\begin{aligned}\mathrm{Tr}[F_{12}(\tau(p))] &= \mathrm{Tr}[P(\tau p)[\partial_i P(\tau p), \partial_j P(\tau p)]] \\ &= \mathrm{Tr}[T^{-1}P(p)TT^{-1}[\partial_i P(p), \partial_j P(p)]T] \\ &= \mathrm{Tr}[(P(p)[\partial_i P(p), \partial_j P(p)])^\dagger] \\ &= \mathrm{Tr}[F_{21}(p)] \\ &= -\mathrm{Tr}[F_{12}(p)]\,,\end{aligned}$$

where in the third line we have used a subtlety in manipulating the traces in the presence of antiunitary operators. Namely, that for $A, B \in L(\mathbb{C}^M)$ with B antiunitary, and $\{\psi_a\}$ an orthonormal basis for $\mathbb{C}^M$, we have

$$\mathrm{Tr}[B^{-1}AB] = \sum_a \langle \psi_a, B^{-1}AB\psi_a\rangle = \sum_a \langle AB\psi_a, B\psi_a\rangle = \mathrm{Tr}[A^\dagger]\,,$$

where the last inequality holds since $\{B\psi_a\}$ is an orthonormal basis for $\mathbb{C}^M$ whenever $\{\psi_a\}$ is. The upshot is that $\mathrm{Tr}[F_{12}]$ is odd under time reversal. Therefore, integrating $\mathrm{Tr}[F_{12}]$ over S^2 or $\mathbb{T}^2$ will yield zero. The first Chern number of a two-dimensional, time reversal symmetric insulator thus vanishes.

Seemingly then, we must conclude that time reversal symmetric insulators are topologically trivial. And, with respect to the notion of topological equivalence we have considered so far, this conclusion is correct. However, for the class of time reversal symmetric insulators a more stringent notion of topological equivalence is necessary. Whereas Definition 2.1 only requires the adiabatic deformation between insulators to preserve the insulating gap, we should for time reversal symmetric insulators additionally require that during the entire deformation the time-reversal symmetry is preserved. With respect to that notion of topological equivalence it turns out that there are distinct equivalence classes of time reversal symmetric insulators.

7.3 Fu-Kane-Mele invariant

In their study of the Quantum Spin Hall effect, which is the time reversal symmetric version of the Quantum Hall effect, Kane and Mele noted a form of $\mathbb{Z}_2$ topological order [13, 12]. The idea of a $\mathbb{Z}_2$ topological invariant for time reversal symmetric insulators has since been developed further by Fu, Kane and Mele to what is now known as the Fu-Kane-Mele (FKM) invariant. We review the construction of the Fu-Kane-Mele invariant [14], and prove its well-definedness in the context of continuum models. Furthermore, we show how it simplifies for systems with inversion symmetry [15].

Consider a Hamiltonian $h : X \to \mathrm{Herm}(2M, \mathbb{C})$ for a two- or three-dimensional system that is time reversal symmetric. Due to Kramers' degeneracy there is necessarily an even number $2M$ of total bands and also an even number $2m$ of valence bands. Let us assume that there are no other degeneracies than Kramers' degeneracy. Since the first Chern number vanishes, the valence bundle is trivial. Hence, we can find a global continuous basis of eigenvectors $\{\psi_a\}_{a=1}^{2m}$ for the valence bundle. The action of time reversal on this basis can be captured in a unitary matrix valued function $w : X \to U(2m)$ defined by

$$w_{ab}(\boldsymbol{k}) = \langle \psi_a(-\boldsymbol{k}), T\psi_b(\boldsymbol{k})\rangle\,.$$

At the time-reversal invariant momenta $\lambda_i = -\lambda_i \in X$, the matrix $w_{ab}(\lambda_i)$ is antisymmetric. Therefore, the Pfaffian $\mathrm{Pf}\, w(\lambda_i)$ is defined. The key property of the Pfaffian of

an antisymmetric matrix is that it squares to the determinant. For a unitary matrix, the determinant has unit modulus, and so the Pfaffian and square root of the determinant can differ only by a sign. For two time-reversal invariant momenta λ_i and λ_j, one can then define the so-called time reversal polarization ν_{ij} to be

$$(-1)^{\nu_{ij}} = \frac{\sqrt{\det w(\lambda_i)}}{\operatorname{Pf} w(\lambda_i)} \frac{\sqrt{\det w(\lambda_j)}}{\operatorname{Pf} w(\lambda_j)} \,. \tag{7.7}$$

It is crucial in this definition that the branch cuts of the square root are chosen such that $\sqrt{\det w}$ is continuous on a path connecting λ_i to λ_j .

By construction ν_{ij} is only well-defined modulo 2 . However, we explicitly chose a frame $\{\psi_a\}$ in the construction of ν_{ij}. It has to be argued that $\nu_{ij} \mod 2$ does not depend on this choice of frame. Let us focus our attention on two-dimensional continuum models for which $X = S^2$, viewed as a compactification of $\mathbb{R}^2$ by identifying the points at infinity. There are two time-reversal invariant momenta, namely $\lambda_0 = 0$ and $\lambda_\infty = \infty$, which in the compactification S^2 are the north and south pole, respectively.

Consider a second global frame $\{\psi'_a\}$ for the valence bundle related to $\{\psi_a\}$ via a gauge transformation $g : S^2 \to U(2m)$ through

$$\psi_a(\boldsymbol{k}) = g_{ab}(\boldsymbol{k})\psi'_b(\boldsymbol{k}) \,.$$

According to [14, Eq. 3.23], the time reversal polarization $\nu'_{0\infty}$ of Eq. (7.7) with respect to the frame $\{\psi'_a\}$ can be equivalently expressed as

$$\nu'_{0\infty} = \frac{1}{2\pi i} \int_0^\pi dk \, \frac{d}{dk} \log \det[w'(k)] - \frac{1}{\pi i} \log \left(\frac{\operatorname{Pf}[w'(\pi)]}{\operatorname{Pf}[w'(0)]} \right) , \tag{7.8}$$

where w' is the time reversal matrix with respect to the frame $\{\psi'_a\}$. The integral is performed over any loop $\gamma : [-\pi, \pi] \to S^2$ such that $\gamma(0) = 0$ and $\gamma(-\pi) = \gamma(\pi) = \infty$. From the definition of $w(\boldsymbol{k})$ it is straightforward to verify that the matrices of the time-reversal operation in the two frames are related as

$$w'(\boldsymbol{k}) = g^T(-\boldsymbol{k}) w(\boldsymbol{k}) g(\boldsymbol{k}) \,.$$

Their determinants are then related as

$$\det w'(\boldsymbol{k}) = \det(g(-\boldsymbol{k})) \det(w(\boldsymbol{k})) \det(g(\boldsymbol{k})) \,,$$

where we used that $\det A^T = \det A$. Furthermore, using the relation $\operatorname{Pf}(B^T A B) = \det(B) \operatorname{Pf}(A)$, their Pfaffians at the time-reversal invariant momenta are related as

$$\operatorname{Pf} w'(\lambda_i) = (\det g(\lambda_i)) \operatorname{Pf} w(\lambda_i) \,,$$

for $i = 0, \infty$. Hence, the time reversal polarizations $\nu'_{0\infty}$ and $\nu_{0\infty}$ are related as

$$\begin{aligned}
\nu'_{0\infty} &= \frac{1}{2\pi i} \int_0^\pi dk \, \frac{d}{dk} \log \det[w'(k)] - \frac{1}{\pi i} \log \left(\frac{\operatorname{Pf}[w'(\pi)]}{\operatorname{Pf}[w'(0)]} \right) \\
&= \frac{1}{2\pi i} \int_0^\pi dk \, \frac{d}{dk} \log \det[g(-k) w(k) g(k)] - \frac{1}{\pi i} \log \left(\frac{\det[g(\pi)] \operatorname{Pf}[w(\pi)]}{\det[g(0)] \operatorname{Pf}[w(0)]} \right) \\
&= \nu_{0\infty} + \frac{1}{2\pi i} \int_0^\pi dk \, \frac{d}{dk} (\log \det[g(-k)] + \log \det[g(k)]) - \frac{1}{\pi i} \log \left(\frac{\det[g(\pi)]}{\det[g(0)]} \right) .
\end{aligned}$$

Let us fix a branch cut for the logarithm. Performing the integrals then yields

$$\nu'_{0\infty} - \nu_{0\infty} = n_- + n_+ + \frac{1}{\pi i} \log\left(\frac{\det[g(\pi)]}{\det[g(0)]}\right) - \frac{1}{\pi i} \log\left(\frac{\det[g(\pi)]}{\det[g(0)]}\right) = n_- + n_+ \,,$$

where n_- is the signed number of times $\det g(-k)$ crosses the branch cut and n_+ is the signed number of times $\det g(k)$ crosses the branch cut, when k goes from 0 to π. Restricted to any loop on S^2 the continuous function $\det g$ has zero winding number because any loop on S^2 is contractible. Therefore, we must have that $n_+ - n_- = 0$. Hence,

$$\nu'_{0\infty} - \nu_{0\infty} = 2n_- \in 2\mathbb{Z}$$

which means that the time reversal polarization on S^2 is well-defined mod 2 . This justifies Eq. (7.7) as a $\mathbb{Z}_2$ invariant for time reversal symmetric continuum models $h : S^2 \to \mathrm{Herm}(2M, \mathbb{C})$. The argument to establish the gauge invariance only relied on the fact that $\pi_1(S^2) = 0$. Since $\pi_1(S^d) = 0$ for $d \geq 2$ it immediately generalizes to continuum models in any dimension $d \geq 2$. For continuum models then, the FKM invariant is defined to be $(-1)^{\nu_{0\infty}}$ which as we have argued above is a $\mathbb{Z}_2$ valued topological invariant, i.e. it is independent of any choice of gauge.

For periodic models the definition is more complicated since there are more time-reversal invariant momenta. For example, $\mathbb{T}^2$ has 4 time-reversal invariant momenta: $(0,0)$, $(0,\pi)$, $(\pi,0)$ and (π,π). One can divide the time-reversal invariant momenta into pairs $\{(0,0),(0,\pi)\}$ and $\{(\pi,0),(\pi,\pi)\}$, and compute the time reversal polarizations for each of these pairs. The sum over all time reversal polarization of the pairs modulo 2 is the FKM invariant for the periodic model. A similar method defines the FKM invariant over $\mathbb{T}^3$. We refer to [14] and [15] for the details. For the remainder of this section, however, we focus on continuum models.

Evaluation of the FKM invariant for specific models is constrained by the fact that finding a global continuous choice of frame for the valence bundle is in general hard to do explicitly. However, if the Hamiltonian, in addition to time-reversal symmetry, also has so-called *parity symmetry*, then the evaluation of the $\mathbb{Z}_2$ invariant can be stated in a different way [15], which is much easier to compute. In particular, if $\{\psi_a(\lambda_0)\}_{a=1}^{2m}$ and $\{\psi_a(\lambda_\infty)\}_{a=1}^{2m}$ are bases for the negative energy eigenstates at the time-reversal invariant momenta consisting of eigenvectors of the parity operator, then

$$(-1)^{\nu_{0\infty}} = \prod_{a=1}^{m} \xi_{2a}(\lambda_0)\xi_{2a}(\lambda_\infty)\,, \tag{7.9}$$

where $\xi_{2a}(\lambda_i)$ is the parity eigenvalue of $\psi_{2a}(\lambda_i)$ with $i = 0, \infty$. This is a significant simplification, because we only need to find eigenvectors at finitely many isolated points.

Let establish Eq. (7.9) more firmly. On a wavefunction $|\boldsymbol{x}, s\rangle$ with a spatial part and a spin part describing a particle in three spatial dimensions, the parity operator has the effect of inverting space, but preserving spin so that

$$|\boldsymbol{x}, s\rangle \mapsto |-\boldsymbol{x}, s\rangle \ .$$

For this reason, we sometimes refer to parity symmetry as *inversion symmetry*. On the effective Hilbert space bundle $X \times \mathbb{C}^{2M}$, the parity operator is represented by an operator that maps from fibers over $\boldsymbol{k}$ to fibers over $-\boldsymbol{k}$ by a unitary $\mathcal{P} \in U(\mathbb{C}^{2M})$ that squares to the identity. A Hamiltonian $h : S^d \to \mathrm{Herm}(2M, \mathbb{C})$ is parity symmetric if

$$h(-\boldsymbol{k}) = \mathcal{P} h(\boldsymbol{k}) \mathcal{P}^{-1} \,. \tag{7.10}$$

Note that this means that $h(\lambda_i)$ commutes with $\mathcal{P}$ at the time-reversal invariant momenta. Hence, at each λ_i the operators $h(\lambda_i)$ and $\mathcal{P}$ have a common basis of eigenvectors.

These common bases at the time-reversal invariant momenta can be extended to a global frame for the valence bundle whenever $d \leq 3$. To see this, recall first that we have assumed that Kramers' degeneracies are the only degeneracies in our model. This implies that the valence bundle splits as a direct sum of $U(2)$ bundles, where each factor corresponds to one Kramers' pair. Due to time-reversal symmetry, the first Chern number of each of these $U(2)$ bundles vanishes. This means, on the one hand, that the $U(2)$ bundle corresponding to a Kramers' pair is trivial and, on the other hand, that its determinant line bundle is also trivial. The triviality of the determinant line bundle implies that the $U(2)$ bundle reduces to an $SU(2)$ bundle. It thus follows that the valence bundle for a time reversal symmetric insulator with no degeneracies other than Kramers' splits as a direct sum of trivial $SU(2)$ bundles.

Let us consider a single trivial $SU(2)$ bundle over S^d with a global frame $\{\varphi_a\}_{a=1}^2$. Assume that for each time-reversal invariant momentum, we have found a basis of parity eigenvectors $\{\psi_a(\lambda_i)\}_{a=1}^2$. For each λ_i, we can extend $\{\psi_a(\lambda_i)\}$ to a frame over a little open ball around λ_i. Because $\pi_{d-1}(SU(2)) = \pi_{d-1}(S^3) = 0$ for $d \leq 3$, an arbitrary global frame $\{\varphi_a\}$ outside the fixed points can be glued continuously to the local frames $\{\psi_a\}$ along the boundaries of the balls around the λ_i. In this way, a global trivialization is obtained which at the time-reversal invariant momenta coincides with the bases $\{\psi_a(\lambda_i)\}_{a=1}^2$.

Suppose then that we have constructed a global orthonormal frame $\{\psi_a\}_{a=1}^{2m}$ such that $\psi_a(\lambda_i)$ is a an eigenstate of $\mathcal{P}$ with parity eigenvalue $\xi_a(\lambda_i)$ and that the energy eigenvalues of $\{\psi_{2a-1}, \psi_{2a}\}$ at λ_i are equal. Because of the gauge invariance of the FKM invariant, we may assume without loss of generality that $\psi_{2a}(\lambda_i) = T\psi_{2a-1}(\lambda_i)$. Indeed, at λ_i, the states $\psi_{2a-1}(\lambda_i)$ and $\psi_{2a}(\lambda_i)$ are orthogonal within the two-dimensional eigenspace of $\epsilon_{2a}(\lambda_i)$, but so are $\psi_{2a-1}(\lambda_i)$ and $T\psi_{2a-1}(\lambda_i)$. Therefore, $\psi_{2a}(\lambda_i) = e^{i\theta_a(\lambda_i)}T\psi_{2a-1}(\lambda_i)$ for some phase $\theta_a(\lambda_i)$. Defining

$$\begin{cases} \psi'_{2a-1} = e^{-i\theta_a}\psi_{2a-1} \\ \psi'_{2a} = \psi_{2a} \end{cases}$$

for some continuous extension of θ_a from the fixed point set to the full momentum space, makes $\{\psi'_{2a-1}(\lambda_i), \psi'_{2a}(\lambda_i)\}$ into a Kramers pair. This gauge transformation does not change the fact that the $\psi'_a(\lambda_i)$ are still parity eigenvectors with the same parity eigenvalues $\xi_a(\lambda_i)$ as the original $\psi_a(\lambda_i)$.

Let us then assume that we have a frame $\{\psi_a\}$ of the valence bundle, which at the time-reversal invariant momenta consists of parity eigenvectors that form Kramers' pairs $\{\psi_{2a-1}(\lambda_i), \psi_{2a}(\lambda_i)\}$. Note that $[T, \mathcal{P}] = 0$ because T acts only on spin while $\mathcal{P}$ acts only on space. Therefore, $\xi_{2a-1}(\lambda_i) = \xi_{2a}(\lambda_i)$ for the parity eigenvalues of $\psi_{2a-1}(\lambda_i)$ and $\psi_{2a}(\lambda_i) = T\psi_{2a-1}(\lambda_i)$. Recall that in the construction of the FKM invariant we defined the matrix

$$w_{ab}(\boldsymbol{k}) = \langle \psi_a(-\boldsymbol{k}), T\psi_b(\boldsymbol{k}) \rangle .$$

With $\mathcal{P}$ available in addition to T we can define another matrix

$$v_{ab}(\boldsymbol{k}) = \langle \psi_a(\boldsymbol{k}), \mathcal{P}T\psi_b(\boldsymbol{k}) \rangle .$$

These two matrices are related through

$$v(-\boldsymbol{k}) = w(\boldsymbol{k})v(\boldsymbol{k})^* w(\boldsymbol{k})^T . \tag{7.11}$$

The matrix v is antisymmetric for all $\boldsymbol{k}$ and hence $\operatorname{Pf} v(\boldsymbol{k})$ is defined for all $\boldsymbol{k}$. Moreover, v is unitary, because

$$\begin{aligned}\sum_{b=1}^{2m} v_{ab}^{\dagger} v_{bc} &= \sum_{b=1}^{2m} \overline{\langle \psi_b(\boldsymbol{k}), \mathcal{P}T\psi_a(\boldsymbol{k})\rangle} \langle \psi_b(\boldsymbol{k}), \mathcal{P}T\psi_c(\boldsymbol{k})\rangle \\ &= \sum_{b=1}^{2m} \langle \mathcal{P}T\psi_a(\boldsymbol{k}), \psi_b(\boldsymbol{k})\rangle \langle \psi_b(\boldsymbol{k}), \mathcal{P}T\psi_c(\boldsymbol{k})\rangle \\ &= \sum_{b=1}^{2m} \langle \mathcal{P}T\psi_a(\boldsymbol{k}), \mathcal{P}T\psi_c(\boldsymbol{k})\rangle \\ &= \delta_{ac}\,.\end{aligned}$$

Therefore, $\operatorname{Pf} v(\boldsymbol{k}) = e^{i\phi(\boldsymbol{k})}$ for some continuous phase ϕ. After performing the gauge transformation

$$\begin{cases} \psi_1' = e^{i\phi}\psi_1 & a = 1 \\ \psi_a' = \psi_a & a \neq 1 \end{cases},$$

we have that $\operatorname{Pf} v' \equiv 1$. Note that although this gauge transformation messes up the Kramers' pair structure, it preserves the parity eigenvalues $\xi_a(\lambda_i)$. Using Eq. (7.11), we see that $\operatorname{Pf} v'(-\boldsymbol{k}) = \det w'(\boldsymbol{k}) \operatorname{Pf}[v'(\boldsymbol{k})]^*$, so that in the present gauge we have

$$\det w' \equiv 1.$$

The issue of choosing branches for the square root in Eq. (7.7) can thus easily be dealt with and we obtain $\sqrt{\det w'} \equiv 1$.

To evaluate $\nu_{0\infty}$ from Eq. (7.7) it only remains to obtain an expression for $\operatorname{Pf} w'(\lambda_i)$. Since $\mathcal{P}^2 = 1$, we have that

$$w'_{ab}(\boldsymbol{k}) = \langle \psi'_a(-\boldsymbol{k}), \mathcal{P}(\mathcal{P}T)\psi'_b(\boldsymbol{k})\rangle$$

so that at the time-reversal invariant momenta

$$w'_{ab}(\lambda_i) = \xi_a(\lambda_i) v'_{ab}(\lambda_i)\,.$$

Then, $(\operatorname{Pf} w'(\lambda_i))^2 = \det w'(\lambda_i) = \det v'(\lambda_i) \prod_{a=1}^{2m} \xi_a(\lambda_i)$. Recall that the parity eigenvalues $\xi_{2a-1}(\lambda_i)$ and $\xi_{2a}(\lambda_i)$ are equal. So, taking the square root we obtain

$$\operatorname{Pf} w'(\lambda_i) = \operatorname{Pf} v'(\lambda_i) \prod_{a=1}^{m} \xi_{2a}(\lambda_i) = \prod_{a=1}^{m} \xi_{2a}(\lambda_i)\,,$$

where the sign of the square root is fixed by the fact that in the case that all $\xi_a(\lambda_i) = 1$, we would have $w'(\lambda_i) = v'(\lambda_i)$. Because of the gauge invariance of $\nu_{0\infty} \mod 2$, this proves the equivalence of Eq. (7.7) and Eq. (7.9).

In conclusion, for continuum models of time reversal symmetric insulators, the FKM invariant is the $\mathbb{Z}_2$-valued topological invariant $(-1)^{\nu_{0\infty}}$. This invariant can in general be computed using Eq. (7.7). If the system has an additional parity symmetry, then the simpler formula Eq. (7.9) applies.

7.4 Example: A low-energy model for Bi_2Se_3

As an example of a time-reversal symmetric insulator we study a low-energy model for Bi_2Se_3. Bi_2Se_3 is a three-dimensional insulator that has both time reversal and inversion symmetry. It also has a three-fold rotation symmetry around what is conventionally taken to be the z axis. Based on these symmetry constraints, a low-energy effective tight-binding model taking into account the four bands closest to the Fermi level has been constructed in [37]. The effective tight-binding Hilbert space is $\mathbb{C}^4$ and Hamiltonian is given by

$$h_{\mathrm{Bi_2Se_3}}(\boldsymbol{k}) = \epsilon_0(\boldsymbol{k}) I_4 + \begin{pmatrix} \mathcal{M}(\boldsymbol{k}) & A_1 k_z & 0 & A_2 k_- \\ A_1 k_z & -\mathcal{M}(\boldsymbol{k}) & A_2 k_- & 0 \\ 0 & A_2 k_+ & \mathcal{M}(\boldsymbol{k}) & -A_1 k_z \\ A_2 k_+ & 0 & -A_1 k_z & -\mathcal{M}(\boldsymbol{k}) \end{pmatrix}, \tag{7.12}$$

where $k_\pm = k_x \pm i k_y$, $\mathcal{M}(\boldsymbol{k}) = M - B_1 k_z^2 - B_2 k_\perp^2$ and $\epsilon_0(\boldsymbol{k}) = C + D_1 k_z^2 + D_2 k_\perp^2$, with $k_\perp^2 = k_x^2 + k_y^2$. This Hamiltonian is written with respect to the ordered basis $\{|\uparrow, \mathrm{Bi}^+\rangle, |\uparrow, \mathrm{Se}^-\rangle, |\downarrow, \mathrm{Bi}^+\rangle, |\downarrow, \mathrm{Se}^-\rangle\}$, where Bi^+ indicates a positive parity orbital corresponding to a Bismuth atom and Se^- indicates a negative parity orbital corresponding to a Selenide atom. The arrows $\uparrow$ and $\downarrow$ indicate the spin in the z direction of the electron in the orbital. To study the topology of Eq. (7.12) we can ignore the $\epsilon_0(\boldsymbol{k})I$ term since this does not affect the eigenvectors. For Bi_2Se_3 the parameters A_1, A_2, B_1, B_2 and M are all positive. To simplify the calculations, let us rescale the first four parameters A_1, A_2, B_1 and B_2 to 1. This rescaling is just a homotopy that preserves all the symmetries, so it does not alter the topological invariant of the model. The model we will thus be studying is

$$h(\boldsymbol{k}) = \begin{pmatrix} M - k^2 & k_z & 0 & k_- \\ k_z & -M + k^2 & k_- & 0 \\ 0 & k_+ & M - k^2 & -k_z \\ k_+ & 0 & -k_z & -M + k^2 \end{pmatrix}, \tag{7.13}$$

with $M > 0$. Our goal is to compute the FKM invariant of this model using Eq. (7.9). In order to do this, we need to obtain a basis of eigenvectors of $h(\boldsymbol{k} = 0)$ and $h(\boldsymbol{k} = \infty)$ that are also parity eigenstates.

Let us introduce the gamma matrices

$$\Gamma_1 = \sigma_x \otimes \tau_x, \quad \Gamma_2 = \sigma_y \otimes \tau_x, \quad \Gamma_3 = \sigma_z \otimes \tau_x, \quad \Gamma_4 = I_2 \otimes \tau_z,$$

where σ_i denotes a Pauli matrix in spin space and τ_i denotes Pauli matrix in orbital space. In terms of these gamma matrices, the Hamiltonian (7.13) can be written as

$$h(\boldsymbol{k}) = k_x \Gamma_1 + k_y \Gamma_2 + k_z \Gamma_3 + (M - k^2)\Gamma_4. \tag{7.14}$$

Moreover, with respect to the basis $\{|\uparrow, \mathrm{Bi}^+\rangle, |\uparrow, \mathrm{Se}^-\rangle, |\downarrow, \mathrm{Bi}^+\rangle, |\downarrow, \mathrm{Se}^-\rangle\}$, the time-reversal operator and parity operators are

$$T = (i\sigma_y \otimes I_2)K \quad \text{and} \quad \mathcal{P} = I_2 \otimes \tau_z.$$

As a first step towards a global basis of eigenvectors for Eq. (7.13) we compute the eigenvalues and eigenvectors of the generic Dirac Hamiltonian

$$h_D = \sum_{j=1}^{4} d_j \Gamma_j. \tag{7.15}$$

Straightforward computation of the characteristic polynomial $\det(h_D - \lambda I_4)$ shows that the eigenvalues of h_D are $\pm d$, where $d^2 = \sum_{j=1}^4 d_j^2$ and that both eigenvalues are doubly degenerate. We are in particular interested in the eigenspace of $\lambda = -d$. By definition, this eigenspace is the kernel of the matrix

$$h_D + d\, I_4 = \begin{pmatrix} d_4 + d & d_3 & 0 & d_1 - id_2 \\ d_3 & -d_4 + d & d_1 - id_2 & 0 \\ 0 & d_1 + id_2 & d_4 + d & -d_3 \\ d_1 + id_2 & 0 & -d_3 & -d_4 + d \end{pmatrix} . \tag{7.16}$$

Two vectors in the kernel are

$$\tilde{v}_1 = \begin{pmatrix} -d_3 \\ d_4 + d \\ -d_1 - id_2 \\ 0 \end{pmatrix} \quad \text{and} \quad \tilde{v}_2 = \begin{pmatrix} d_4 - d \\ d_3 \\ 0 \\ d_1 + id_2 \end{pmatrix} .$$

These vectors are independent as long as $d_1 + id_2 \neq 0$. However, we can make two different orthonormal bases of the negative energy eigenspace by performing the Gramm-Schmidt algorithm, depending on whether we start from $\tilde{v}_1$ or from $\tilde{v}_2$. Starting the Gramm-Schmidt algorithm from $\tilde{v}_1$, we obtain

$$u_1 = \frac{1}{\sqrt{2d(d + d_4)}} \begin{pmatrix} -d_3 \\ d + d_4 \\ -d_1 - id_2 \\ 0 \end{pmatrix} \quad \text{and} \quad u_2 = \frac{1}{\sqrt{2d(d + d_4)}} \begin{pmatrix} -d_1 + id_2 \\ 0 \\ d_3 \\ d + d_4 \end{pmatrix} .$$

On the other hand, starting from $\tilde{v}_2$ yields

$$u_1' = \frac{1}{\sqrt{2d(d - d_4)}} \begin{pmatrix} 0 \\ d_1 - id_2 \\ d_4 - d \\ -d_3 \end{pmatrix} \quad \text{and} \quad u_2' = \frac{1}{\sqrt{2d(d - d_4)}} \begin{pmatrix} d_4 - d \\ d_3 \\ 0 \\ d_1 + id_2 \end{pmatrix} .$$

Although they have been constructed rather *ad hoc*, it is a straightforward check that $\{u_1, u_2\}$ and $\{u_1', u_2'\}$ are indeed two orthonormal bases for the $-d$ eigenspace of h_D. Moreover, the $SU(2)$ matrix

$$g = \frac{1}{\sqrt{d^2 - d_4^2}} \begin{pmatrix} d_1 - id_2 & -d_3 \\ d_3 & d_1 + id_2 \end{pmatrix} \tag{7.17}$$

transforms $\{u_1, u_2\}$ into $\{u_1', u_2'\}$ according to $u_i' = \sum_j g_{ij} u_j$.

Let us now specialize to $\boldsymbol{d}(\boldsymbol{k}) = (k_x, k_y, k_z, M - k^2)$ which corresponds to (7.13). However, instead of taking $M > 0$ as is required for Bi_2Se_3, we will first consider $M < 0$. If $M < 0$, then $d_4(\boldsymbol{k}) = M - k^2 < 0$ for all $\boldsymbol{k} \in \mathbb{R}^3$. Therefore, the basis $\{u_1'(\boldsymbol{k}), u_2'(\boldsymbol{k})\}$ is well-defined for all $\boldsymbol{k} \in \mathbb{R}^3$. Indeed, $d(\boldsymbol{k}) - d_4(\boldsymbol{k}) > 0$ for all $\boldsymbol{k} \in \mathbb{R}^3$ so there are no issues with the normalization. However, it remains to study the behaviour of $\{u_1', u_2'\}$ in the limit $k \to \infty$. In this limit, the quadratic term d_4 will dominate the linear terms d_1, d_2 and d_3. In particular, $d \to -d_4$ when $k \to \infty$, and hence,

$$u_1' \to \begin{pmatrix} 0 & 0 & -1 & 0 \end{pmatrix}^T \quad \text{and} \quad u_2' \to \begin{pmatrix} -1 & 0 & 0 & 0 \end{pmatrix}^T .$$

The basis $\{u_1', u_2'\}$ thus extends to $\boldsymbol{k} = \infty$ and we conclude that $\{u_1', u_2'\}$ actually provides a global continuous frame for the valence bundle over S^3. The parity operator is

$$\mathcal{P} = I_2 \otimes \tau_z = \mathrm{diag}(1, -1, 1, -1)\,.$$

At $\boldsymbol{k} = 0$, the vectors $\{u_1', u_2'\}$ become

$$u_1'(0) = \begin{pmatrix} 0 & 0 & -1 & 0 \end{pmatrix}^T \quad \text{and} \quad u_2'(0) = \begin{pmatrix} -1 & 0 & 0 & 0 \end{pmatrix}^T .$$

These are already parity eigenvectors with parity $\xi_1'(0) = \xi_2'(0) = 1$. Meanwhile, as $k \to \infty$, the vectors $\{u_1', u_2'\}$ become

$$u_1'(\infty) = \begin{pmatrix} 0 & 0 & -1 & 0 \end{pmatrix}^T \quad \text{and} \quad u_2'(\infty) = \begin{pmatrix} -1 & 0 & 0 & 0 \end{pmatrix}^T ,$$

which are again parity eigenvectors with parity $\xi_1'(\infty) = \xi_2'(\infty) = 1$. The $\mathbb{Z}_2$ invariant is the product of the parities of $u_2'(0)$ and $u_2'(\infty)$, which is

$$(-1)^{\nu_{0\infty}} = \xi_2'(0)\xi_2'(\infty) = 1 \cdot 1 = 1\,.$$

For $M < 0$, the model Eq. (7.13) thus describes a trivial insulator.

Now let us consider the situation in which $M > 0$. In this case, the basis $\{u_1', u_2'\}$ is problematic at $\boldsymbol{k} = 0$. Indeed, $d \to d_4$ as $\boldsymbol{k} \to 0$, so we might expect some trouble from the singularity of the normalization factor. The problem that occurs can be seen as follows. Suppose we set $k_x = k_y = 0$ and let k_z approach 0. The first three components of u_1' will tend to zero, but for k_z small, the fourth component behaves as

$$(u_1')_4(0, 0, k_z) = \frac{-d_3}{\sqrt{2d(d - d_4)}}\bigg|_{\boldsymbol{k}=(0,0,k_z)} \simeq \frac{-k_z}{\sqrt{k_z^2}}$$

and thus has a discontinuity at $k_z = 0$, as can be seen on the left in Fig. 7. Therefore, $\{u_1', u_2'\}$ cannot be used as a frame around $\boldsymbol{k} = 0$. On the other hand, the unprimed basis $\{u_1, u_2\}$ is well-defined at $\boldsymbol{k} = 0$, because $d + d_4 > 0$ at $\boldsymbol{k} = 0$, but it is ill-defined when $\boldsymbol{k} \to \infty$. Indeed, if $k_x = k_y = 0$ and we let $k_z \to \infty$, then the first component of u_1 behaves as

$$(u_1)_1(0, 0, k_z) \simeq \frac{-k_z}{\sqrt{k_z^2}}\,,$$

for large k_z, as illustrated on the right in Fig. 7. Hence, $(u_1)_1$ converges to -1 if $k_z \to +\infty$ and to $+1$ if $k_z \to -\infty$, and so does not extend continuously to $\boldsymbol{k} = \infty$. To compute the

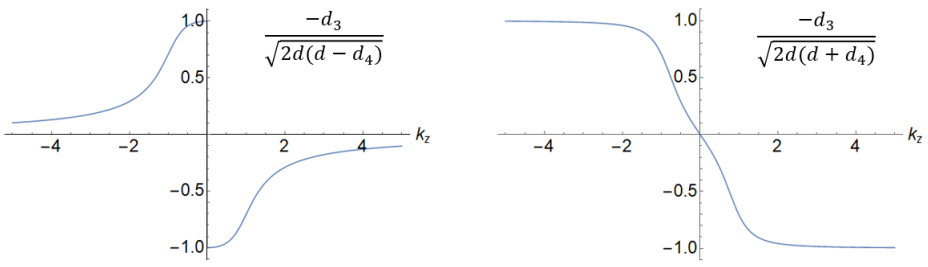

Figure 7: Left: The fourth component of $u_1'(0, 0, k_z)$ is discontinuous at $k_z = 0$. Right: The first component of $u_1(0, 0, k_z)$ is discontinuous at $k_z = \infty$.

FKM invariant we should thus use the basis $\{u_1, u_2\}$ around $\boldsymbol{k} = 0$ and use the basis

$\{u_1', u_2'\}$ around $\boldsymbol{k} = \infty$. The parity of $u_2(0) = (0, 0, 0, 1)^T$ is $\xi_2(0) = -1$ and the parity of $u_2'(\infty) = (-1, 0, 0, 0)^T$ is $\xi_2'(\infty) = 1$. The product gives the $\mathbb{Z}_2$ invariant

$$(-1)^{\nu_{0\infty}} = \xi_2(0)\xi_2'(\infty) = -1 \cdot 1 = -1\,.$$

The case $M > 0$ thus describes the non-trivial phase, which means that Bi_2Se_3 is a topologically non-trivial insulator.

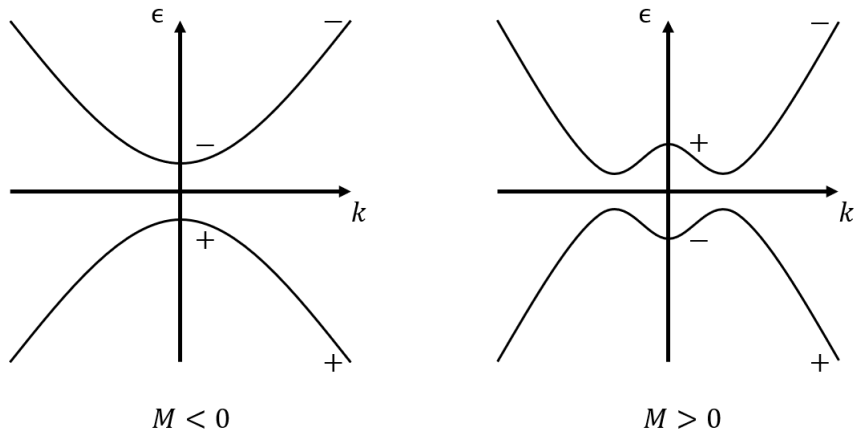

Figure 8: Left: Band structure for $M < 0$. Right: Band structure for $M > 0$. The $+$ and $-$ labels indicate the parity of the bands. For $M > 0$, the $+$ and $-$ parity bands are inverted near $\boldsymbol{k} = 0$.

The above computations of the $\mathbb{Z}_2$ invariant of the low-energy Bi_2Se_3 model can be summarized in a simple physical picture. There are two Kramers degenerate valence bands and two Kramers degenerate conduction bands. For $M < 0$, the valence bands have positive parity and the conduction bands, therefore, must have negative parity. The product of the parities of one copy of the valence bands at $\boldsymbol{k} = 0$ and $\boldsymbol{k} = \infty$ is 1. When M is increased from negative to positive, the bands cross at $M = 0$ in the point $\boldsymbol{k} = 0$ and, subsequently, for $M > 0$ they are "inverted"; The positive parity band now has positive energy at $\boldsymbol{k} = 0$ while the negative parity band has negative energy. Physically, this band inversion is due to spin-orbit coupling [37]. Now, for $M > 0$ the bands would have to cross at nonzero values of $\boldsymbol{k}$, but in reality this degeneracy will be lifted by the "avoided crossing" mechanism. This yields a final band structure as in Fig. 8. It is seen that now for $M > 0$ the parity at $\boldsymbol{k} = 0$ of the valence bands is -1 while at $\boldsymbol{k} = \infty$ it is still 1. The product of the two parities is -1, i.e. we have non-trivial insulator.

8 Classification of "Quaternionic" vector bundles

The first Chern number of a time reversal symmetric insulator vanishes. In mathematical terms, the complex vector bundle associated to a time reversal symmetric insulator via the valence bundle construction is trivial. However, the time-reversal operation induces extra structure on the valence bundle. With this extra structure, the valence bundle becomes a "Quaternionic" vector bundle. The goal of this section is to introduce "Quaternionic" vector bundles and define the so-called FKMM invariant for "Quaternionic" vector bundles over sufficiently simple base spaces. We prove that for continuum models in dimensions two and three the FKMM invariant fully classifies "Quaternionic" vector bundles. Moreover,

we show that the FKMM invariant is equivalent to the Fu-Kane-Mele invariant in the continuum models.

8.1 "Quaternionic" vector bundles

The operation of time reversal introduces extra structure on the valence bundle of an insulator. On the one hand, the momentum space becomes endowed with an involution $\tau : X \to X$ that corresponds to the action $\boldsymbol{k} \mapsto -\boldsymbol{k}$ of time reversal on momentum. On the other hand, there is an antilinear transformation $T = U_T K$ on the Hilbert space $\mathbb{C}^M$. This extra structure can be formalized by the notion of "Quaternionic" structure on a complex vector bundle over an involutive base space.

An involutive space is a pair (X, τ) of a topological space X together with a homeomorphism $\tau : X \to X$ such that $\tau^2 = \mathrm{id}$. Let $\pi : E \to X$ be a complex vector bundle. A *"Quaternionic" structure* on E is a homeomorphism $\Theta : E \to E$ such that

(Q1) Θ lifts the involution τ, in the sense that $\pi \circ \Theta = \tau \circ \pi$,

(Q2) $\Theta|_{E_x} : E_x \to E_{\tau(x)}$ is antilinear, meaning $\Theta(\lambda e + f) = \bar{\lambda}\Theta(e) + \Theta(f)$ for $\lambda \in \mathbb{C}$ and $e, f \in E_x$,

(Q3) $\Theta^2 = -\mathrm{id}$.

Closely related to the notion of "Quaternionic" structure is the so-called *"Real" structure*. A "Real" structure on a complex vector bundle over an involutive base space is defined in exactly the same way as a "Quaternionic" structure, except that the third property (Q3) in the above is changed to

(R3) $\Theta^2 = \mathrm{id}$.

A pair (E, Θ) of a complex vector bundle E together with a "Quaternionic" structure will be referred to as a "Quaternionic" vector bundle. If Θ is a "Real" structure, then we call (E, Θ) a "Real" vector bundle. If E is the valence bundle associated to some insulator, i.e. if E is a trivial subbundle of $X \times \mathbb{C}^M$, then $\Theta : (x, \psi) \mapsto (\tau(x), T\psi)$ clearly defines a "Quaternionic" structure on E because T is antilinear and squares to -1. Implicit in our notion of insulator is that the single-particle problem describes fermions so that indeed $T^2 = -1$. If the single-particle problem would be bosonic, then T would square to 1 and we would obtain a "Real" vector bundle. However, our actual interest in "Real" vector bundles is that they play a key role in the development of the FKMM invariant for "Quaternionic" vector bundles, as will be discussed later.

The reason for putting the putting the word "Quaternionic" in quotations is that one must not confuse the above defined notion of "Quaternionic" vector bundle with the more standard notion of quaternionic vector bundle, namely, a vector bundle whose fibers are quaternionic vector spaces. The two notions are related though. In particular, the newly defined "Quaternionic" vector bundle is a generalization of the usual quaternionic vector bundle. A quaternionic vector bundle $E \to X$ is naturally endowed with two complex structures $I, J : E \to E$, which are bundle automorphism of the underlying real vector bundle $E_{\mathbb{R}}$ encoding the multiplication by i and j respectively. Endowing the underlying real vector bundle with only one of the two complex structure, say I, the pair $(E_{\mathbb{R}}, I)$ can be thought of as a complex vector bundle $E_{\mathbb{C}}$. The second complex structure J becomes a fiberwise antilinear map on $E_{\mathbb{C}}$ that squares to minus the identity. Therefore, J is a "Quaternionic" structure on $E_{\mathbb{C}} \to X$, where X is endowed with the trivial involution $\tau = \mathrm{id}$. The pair $(E_{\mathbb{C}}, J)$ is then a "Quaternionic" vector bundle.

In the converse spirit, if (E, Θ) is a "Quaternionic" vector bundle over (X, τ) and $x \in X$ is a fixed point of τ, then the fiber E_x is isomorphic to $\mathbb{H}^m$, where $\mathbb{H}$ denotes the quaternions. Indeed, $\Theta|_{E_x} : E_x \to E_x$ is an antilinear map that squares to -1, so it plays the role of the quaternionic unit j in E_x. A particular consequence of this fact is that E has *complex* rank $2m$. We will be interested in "Quaternionic" bundles over base spaces with fixed points, namely the time-reversal invariant momenta, so we will study "Quaternionic" bundles whose underlying complex vector bundle has even complex rank.

We now develop the notions of morphism, triviality and frames for "Quaternionic" bundles. To distinguish these from the corresponding notions of the underlying complex vector bundle they should receive the adjective "Quaternionic". To tighten up the notation, we will often abbreviate these adjectives to just Q.

A *Q-morphism* between two Q-bundles (E, Θ) and (E', Θ') is a morphism $\varphi : E \to E'$ of the underlying complex vector bundles that satisfies

$$\varphi \circ \Theta = \Theta' \circ \varphi \,.$$

If φ satisfies the equivariance condition $\varphi \circ \Theta = \Theta' \circ \varphi$ and is an isomorphism of complex vector bundles, then it is a Q-isomorphism. The Q-isomorphism classes of rank $2m$ Q-bundles over an involutive space (X, τ) will be denoted by $\mathrm{Vect}_Q^{2m}(X, \tau)$, where we stress that $2m$ is the complex rank of the underlying complex vector bundle. For $X \times \mathbb{C}^{2m} \to X$ the trivial complex vector bundle of rank $2m$ over an involutive space (X, τ), we define the *trivial Q-structure* to be

$$\Theta_0(x, v) = (\tau(x), Q\bar{v}) \,, \qquad \text{where} \quad Q = \begin{pmatrix} 0 & -I_m \\ I_m & 0 \end{pmatrix} \,,$$

where the matrix Q is written with respect to the standard basis of $\mathbb{C}^{2m}$. A Q-bundle of rank $2m$ is *Q-trivial* if it is Q-isomorphic to the trivial Q-bundle $(X \times \mathbb{C}^{2m}, \Theta_0)$. A Q-bundle is *locally* Q-trivial if for each $x \in X$ there exists a Q-isomorphism between $\varphi : (E|_U, \Theta|_U) \to (U \times \mathbb{C}^{2m}, \Theta_0)$ over a neighbourhood U of x that is τ-invariant, meaning that $\tau(U) = U$.

Just like for ordinary complex vector bundles triviality can also be understood in terms of frames, Q-triviality can be understood in terms of Q-frames. A Q-structure Θ induces a map $\tau_\Theta : \Gamma(E) \to \Gamma(E)$ on sections given by

$$\tau_\Theta(s) = \Theta \circ s \circ \tau \,.$$

A *Q-frame* over a τ-invariant open set $U \subseteq X$ is defined to be a frame of E of the form

$$\{s_1, \ldots, s_m, \tau_\Theta(s_1), \ldots, \tau_\Theta(s_m)\} \,.$$

In physics lingo, a Q-frame is a frame consisting of Kramers' pairs $(s_a, \tau_\Theta(s_a))$. The standard basis $\{e_1, \ldots, e_{2m}\}$ of $\mathbb{C}^{2m}$ is a global Q-frame for the trivial Q-bundle $(X \times \mathbb{C}^{2m}, \Theta_0)$.[11] Indeed, for $a = 1, \ldots, m$, we have $Q\bar{e}_a = Qe_a = e_{m+a}$. We have used that $e_a = (0, \ldots, 1, \ldots, 0)$ with the 1 appearing in the a-th slot to see that $\bar{e}_a = e_a$. More generally, a Q-bundle is Q-trivial if and only if it admits a global Q-frame. To see this, suppose we are given a Q-trivialization $\varphi : E \to X \times \mathbb{C}^{2m}$ and define $s_a = \varphi^{-1}(e_a)$. For $a = 1, \ldots, 2m$ it follows that

$$s_{m+a} = \varphi^{-1}(e_{m+a}) = \varphi^{-1}(\tau_{\Theta_0} e_a) = (\varphi^{-1} \circ \Theta_0)(e_a \circ \tau) = (\Theta \circ \varphi^{-1})(e_a \circ \tau) = \tau_\Theta(s_a) \,.$$

[11] We do not distinguish in notation between the vector $e_a \in \mathbb{C}^{2m}$ and the section $e_a : X \to X \times \mathbb{C}^{2m}$ defined as $x \mapsto (x, e_a)$.

Conversely, it can be checked that given a Q-frame $\{s_1, \ldots, s_m, \tau_\Theta(s_1), \ldots, \tau_\Theta(s_m)\}$ we can construct a Q-trivialization $\varphi : E \to X \times \mathbb{C}^{2m}$ by defining $\varphi(s_a) = e_a$ and $\varphi(\tau_\Theta(s_a)) = e_{m+a}$ for $a = 1, \ldots, m$ and extending linearly.

Whereas complex vector bundles are locally trivial by definition, it actually requires some work to show that Q-bundles are locally Q-trivial.

Proposition 8.1. A Q-bundle is locally Q-trivial.

Proof. Let (E, Θ) be a Q-bundle over an involutive space (X, τ) and let $x \in X$. The goal is to obtain a Q-frame over a τ-invariant neighbourhood of x. We treat the two cases $\tau(x) = x$ and $\tau(x) \neq x$ separately.

Suppose $\tau(x) = x$. The main task is to construct a Q-basis for E_x. This can be done inductively. We start by picking a vector $s_1(x) \in E_x$ nonzero. We claim that acting with the Q-structure Θ yields a linearly independent vector $\Theta(s_1(x)) \in E_x$. If it were not linearly independent, then $\Theta(s_1(x)) = \lambda s_1(x)$ for some $\lambda \in \mathbb{C} \setminus \{0\}$. Applying Θ on both sides gives $-s_1(x) = \bar{\lambda}\Theta(s_1(x))$ which implies that $\lambda = -\bar{\lambda}^{-1}$. This is a contradiction, since writing $\lambda = re^{i\phi}$ the requirement $\lambda = -\bar{\lambda}^{-1}$ is equivalent to $r = -1/r$, which cannot be satisfied for $r > 0$.

By way of induction, suppose that $s_1(x), \ldots, s_k(x)$ and $\Theta(s_1(x)), \ldots, \Theta(s_k(x))$ are all independent and pick one more independent vector $s_{k+1}(x)$. The claim is that $\Theta(s_{k+1}(x))$ is independent of all these vectors. Again, we argue by contradiction. Suppose we can write

$$\Theta(s_{k+1}(x)) = \sum_{i=1}^{k} \lambda_a \Theta(s_a(x)) + \sum_{b=1}^{k} \mu_b s_b(x) + \mu_{k+1} s_{k+1}(x)\,,$$

for some $\lambda_a, \mu_b \in \mathbb{C}$ with at least one of them nonzero. Applying Θ to both sides yields

$$-s_{k+1}(x) = -\sum_{a=1}^{k} \bar{\lambda}_a s_a(x) + \sum_{b=1}^{k} \bar{\mu}_b \Theta(s_b(x)) + \bar{\mu}_{k+1} \Theta(s_{k+1}(x))\,.$$

Plugging in the expression of $\Theta(s_{k+1}(x))$ from the first equation into the second equation we obtain

$$\sum_{a=1}^{k} \alpha_a \Theta(s_a(x)) + \sum_{b=1}^{k} \beta_b s_b(x) + (1 + |\mu_{k+1}|^2) s_{k+1}(x) = 0\,, \tag{8.1}$$

for some $\alpha_a, \beta_b \in \mathbb{C}$, whose precise form is not important. By the induction hypothesis, $s_1(x), \ldots, s_{k+1}(x)$ and $\Theta(s_1(x)), \ldots, \Theta(s_k(x))$ are independent. Therefore, all coefficients in Eq. (8.1) must be zero. This is a contradiction since $1 + |\mu_{k+1}|^2 > 0$ for any $\mu_{k+1} \in \mathbb{C}$. Hence, $\Theta(s_{k+1}(x))$ is also independent. This inductive process thus leads to a Q-basis $\{s_1(x), \ldots, s_m(x), \Theta(s_1(x)), \ldots, \Theta(s_m(x))\}$.

We now extend the vectors $s_a(x) \in E_x$ to global sections $s_a \in \Gamma(E)$. Since x is a fixed point, $\Theta(s_a(x)) = \Theta(s_a(\tau(x)) = \tau_\Theta(s_a)(x)$ and so the set of sections $\{s_1, , \ldots, \tau_\Theta(s_m)\}$ is linearly independent at the fixed point x, because there it reduces to the basis we constructed above. Since being linearly independent is an open condition, there is an open neighbourhood U of x on which $\{s_1, \ldots, s_m, \tau_\Theta(s_1), \ldots, \tau_\Theta(s_m)\}$ are still linearly independent. We can shrink U to a τ-invariant neighbourhood by considering $U' = U \cap \tau(U)$. We have thus constructed a Q-frame over a τ-invariant neighbourhood U' of the fixed point x.

If $\tau(x) \neq x$, we construct a Q-frame over the closed set $Y = \{x, \tau(x)\}$ in a similar fashion to the above. Start by picking $s_1(x) \in E_x$ and $s_1(\tau(x)) \in E_{\tau(x)}$. This defines a section s_1 over Y. The map τ_Θ squares to minus one and is antilinear, so exactly the

same algebra as in the case $\tau(x) = x$ leads to an inductive construction of a Q-frame $\{s_1, \ldots, s_m, \tau_\Theta(s_1), \ldots, \tau_\Theta(s_m)\}$ over Y. We extend to global sections and subsequently restrict to an open U on which the extended sections are linearly independent. We thus obtain a Q-frame over $U \cap \tau(U)$. □

Every Q-bundle is locally Q-trivial, but not every Q-bundle is Q-trivial, globally, that is. It turns out that Q-triviality is exactly the property we need to distinguish the different phases of the time-reversal invariant insulator in dimensions two and three. In the next section, we develop an invariant which tells us whether or not a given Q-bundle is Q-trivial.

8.2 Homotopy formulation of the FKMM invariant

Classifying Q-bundles over an arbitrary base space (X, τ) is a difficult task. Luckily, the spaces that come up when dealing with topological insulators are quite simple. We will capture this simplicity in the following technical assumptions. Firstly, we assume that all the fixed points of τ are isolated and that there are a finite, but nonzero, amount of them. Secondly, we assume that all "Real" line bundles over (X, τ) are R-trivial. In [18], DeNittis and Gomi have developed an invariant for Q-bundles over these types of base spaces called the FKMM invariant.[12] Let us review their construction.

Let (X, τ) be an involutive space satisfying the two assumptions above and (E, Θ) a Q-bundle over (X, τ). By the first assumption τ has fixed points, so the underlying complex vector bundle E has even rank $2m$ for some $m \in \mathbb{N}$. The associated determinant line bundle $\det E := \bigwedge^{2m} E$ is then endowed with a "Real" structure $\det \Theta$. To see this, let $\{s_1, \ldots, s_{2m}\}$ be a local frame of E such that $s_1 \wedge \cdots \wedge s_{2m}$ is a local frame of $\det E$. Note that any local frame of $\det E$ is of this form. On this frame, we have by definition that

$$(\det \Theta)(s_1 \wedge \cdots \wedge s_{2m}) := \Theta s_1 \wedge \cdots \wedge \Theta s_{2m} .$$

Since Θ is a homeomorphism that is antilinear on fibers, the same holds for $\det \Theta$. Moreover, since $\Theta^2 = -1$, it follows that $(\det \Theta)^2$ acts as $(-1)^{2m} = 1$. Therefore, $\det \Theta$ is indeed a "Real" structure.

By the second assumption, the "Real" line bundle $(\det E, \det \Theta)$ must be R-trivial. This means that it has a global nonvanishing section s that satisfies the equivariance property

$$(\det \Theta) \circ s = s \circ \tau .$$

Such an equivariant section is also referred to as an R-section. At a fixed point $x \in X^\tau$, an R-section s satisfies $s(x) = (\det \Theta)(s(x))$. Since $\det \Theta|_{E_x}$ is antilinear and squares to the identity, it acts as complex conjugation in the fiber and, in this sense, s can be thought of as being real at the fixed points. The idea behind the FKMM invariant is to compare the global sections of $(\det E, \det \Theta)$ to a canonical orientation of the real lines over the fixed points X^τ coming from the Q-structure on E. The existence of this canonical orientation is the content of the following Proposition.

Proposition 8.2. If (X, τ) is a space with trivial involution $\tau = \mathrm{id}_X$ and (E, Θ) is Q-bundle over X, then $(\det E, \det \Theta)$ has a canonical orientation.

Proof. Let (E, Θ) be a Q-bundle over an involutive space (X, τ) with trivial involution. Let $(U_\alpha, \{s_1^\alpha, \ldots, s_{2m}^\alpha\})$ be a collection of trivializing Q-frames of (E, Θ). By equivariance, we know that the section $s_{\det}^\alpha := s_1^\alpha \wedge \cdots \wedge s_{2m}^\alpha$ of $\det E|_{U_\alpha}$ is real. We claim that the

[12] DeNittis and Gomi have named the FKMM invariant after Furuta, Kametani, Matsue and Minami, whose unpublished work [38] was a main source of inspiration for [18].

collection $(U_\alpha, s^\alpha_{\det})$ defines an orientation. What has to be shown is that on overlaps $U_\alpha \cap U_\beta$ the signs of $s^\alpha_{\det}$ and $s^\beta_{\det}$ agree.

Let $(g_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{GL}(\mathbb{C}^{2m}))$ denote the transition functions of the Q-trivialization $(U_\alpha, \{s^\alpha_1, \dots, s^\alpha_{2m}\})$. The transition functions are equivariant with respect to the Q-structure Θ_0 on the product bundle $U \times \mathbb{C}^{2m}$. Therefore,

$$g(x) = -Q\overline{g(\tau x)}Q = -Q\overline{g(x)}Q\,.$$

If $g(x)v = \lambda v$, then, using the above equation,

$$g(x)Q\bar{v} = -Q\overline{g(x)}Q^2\bar{v} = Q\overline{g(x)v} = \bar{\lambda}Q\bar{v}\,.$$

The eigenvalues of $g(x)$ thus appear in conjugate pairs which makes the determinant of $g(x)$ strictly positive. The transitions functions of $(\det E, \det \Theta)$ are precisely the determinants of the transition functions of (E, Θ). It follows that on $U_\alpha \cap U_\beta$, the sections $s^\alpha_{\det}$ and $s^\beta_{\det}$ are positive multiples of one another, which proves the claim.

There is thus a well-defined orientation induced by the Q-trivialization $(U_\alpha, \{s^\alpha_1, \dots\})$. The orientation is independent of choice of Q-trivialization since two Q-trivializations are again related by equivariant transition functions which have positive determinant. The orientation is thus independent of the chosen Q-trivialization and thus it is canonical. □

For any involutive space (X, τ), the fixed point set X^τ is a space with trivial involution. Therefore, if (E, Θ) is a Q-bundle over (X, τ), the determinant bundle $(\det E|_{X^\tau}, \det \Theta|_{X^\tau})$ restricted to X^τ has a canonical orientation. If $t : X \to \det E$ is a global R-section, then at the fixed points it is real and we can compare it to the canonical orientation.

Let s_{X^τ} be an equivariant section of $\det E|_{X^\tau}$ that agrees with the canonical orientation. Any Q-trivialization of (E, Θ) gives rise to such a section. Since t is an equivariant section, there is an $f : X^\tau \to \mathbb{C}^*$ that is equivariant with respect to τ on X^τ and complex conjugation on $\mathbb{C}^*$ such that

$$t|_{X^\tau} = f \cdot s_{X^\tau}\,.$$

Indeed, for $x \in X^\tau$, the equivariance of t and s_{X^τ} with respect to $\det \Theta$ implies

$$t(x) = (\det \Theta)(t(x)) = (\det \Theta)(f(x)s_{X^\tau}(x)) = \overline{f(x)}(\det \Theta)(s_{X^\tau}(x)) = \overline{f(x)}s_{X^\tau}(x)$$

so that $f(\tau(x)) = f(x) = \overline{f(x)}$.

The function f depends on the choice of t. If t' is another nonvanishing equivariant section of $\det E$, then $t' = g \cdot t$ for some equivariant function $g : X \to \mathbb{C}^*$ and we have

$$t'|_{X^\tau} = g|_{X^\tau} \cdot t|_{X^\tau} = (g|_{X^\tau} \cdot f) \cdot s_{X^\tau}\,.$$

This construction of comparing a global section to the canonical orientation as represented by s_{X^τ} thus yields an equivariant map $f : X^\tau \to \mathbb{C}^*$, which is defined up to multiplication by the restriction of global equivariant maps $g : X \to \mathbb{C}^*$. Of course, f also still depends on the choice of s_{X^τ}. Another choice of s_{X^τ} corresponds to rescaling f by a positive real function. To take this redundancy into account, we consider the equivariant homotopy class $[f] \in [X^\tau, \mathbb{C}^*]_{\mathbb{Z}_2}$, where we recall that two equivariant maps are equivariantly homotopic if and only if they are homotopic via a homotopy that consists of equivariant maps for all times. The definition of the FKMM invariant can now be stated as follows.

Definition 8.3 (FKMM invariant)**.** Let (E, Θ) be a Q-bundle, let t be a global equivariant section of $(\det E, \det \Theta)$, let s_{X^τ} represent the canonical orientation of $\det E|_{X^\tau}$, and let

$f : X^\tau \to \mathbb{C}^*$ be such that $t|_{X^\tau} = f \cdot s_{X^\tau}$. The *FKMM invariant* of (E, Θ) is defined to be the class

$$[f] \in [X^\tau, \mathbb{C}^*]_{\mathbb{Z}_2} / [X, \mathbb{C}^*]_{\mathbb{Z}_2} \,,$$

where the action of $[g] \in [X, \mathbb{C}^*]_{\mathbb{Z}_2}$ on $[f] \in [X^\tau, \mathbb{C}^*]_{\mathbb{Z}_2}$ is given by $[g|_{X^\tau} \cdot f]$.

To see that this construction really yields an invariant for "Quaternionic" vector bundles, it must be shown that the class $[f]$ only depends on the isomorphism type of (E, Θ). Suppose that $\varphi : (E, \Theta) \to (E', \Theta')$ is a Q-isomorphism. Then, $\det \varphi : (\det E, \det \Theta) \to (\det E', \det \Theta')$ is an R-isomorphism. Let t be global equivariant section of $(\det E, \det \Theta)$ and let $f : X^\tau \to \mathbb{C}^*$ be such that

$$t|_{X^\tau} = f \cdot s_{X^\tau} \,.$$

The section $t' = (\det \varphi)(t)$ is a global nonvanishing equivariant section of $(\det E', \det \Theta')$ since by the equivariance of φ,

$$t'(\tau(x)) = (\det \varphi)(t(\tau(x))) = (\det \varphi)(\det \Theta)(t(x)) = (\det \Theta')(\det \varphi)(t(x)) = (\det \Theta')(t'(x)).$$

Moreover, recall from the proof of Proposition 8.2 that s_{X^τ} is constructed locally as the wedge product of a local Q-frame $\{s_1, \ldots, s_{2m}\}$ for (E, Θ). Since $\{\varphi(s_1), \ldots, \varphi(s_{2m})\}$ is a local Q-frame for (E', Θ'), it follows that

$$s'_{X^\tau} = (\det \varphi)(s_{X^\tau}) \,.$$

Therefore,

$$t'|_{X^\tau} = (\det \varphi)(t)|_{X^\tau} = (\det \varphi)(f \cdot s_{X^\tau}) = f \cdot (\det \varphi)(s_{X^\tau}) = f \cdot s'_{X^\tau} \,,$$

which means that $[f]$ also represents the FKMM invariant for (E', Θ'). The FKMM invariant is thus a proper invariant for "Quaternionic" vector bundles over an involutive space X satisfying the two technical assumptions stated earlier, in the sense that it defines a map

$$\kappa : \mathrm{Vect}_Q^{2m}(X, \tau) \to [X^\tau, \mathbb{C}^*]_{\mathbb{Z}_2} / [X, \mathbb{C}^*]_{\mathbb{Z}_2} \,.$$

It is actually the case that if $\dim X \leq 3$, then the map κ is even injective. For the proof of this fact we refer to [18, Theorem 1.1]. We note here that a useful ingredient in the proof is that for $\dim X \leq 3$, it is enough to consider rank 2 Q-bundles (i.e. $m = 1$) [18, Theorem 2.5]. The reason is that rank $2m$ Q-bundles (E, Θ) for $m > 1$ always split as a direct sum

$$(E, \Theta) \cong (E', \Theta') \oplus (X \times \mathbb{C}^{2m-2}, \Theta_0) \,,$$

were E' is rank 2. The proofs of both the injectivity and the reduction to the case of rank 2 when $\dim X \leq 3$ heavily rely on X having a so-called $\mathbb{Z}_2$-CW structure. This is not a very stringent condition and, in particular, the time reversal sphere and torus satisfy it.

8.3 The FKMM invariant for time reversal spheres

Recall that the time reversal sphere $\widetilde{S}^d = (S^d, \tau)$ is the sphere endowed with the involution

$$\tau : (k_1, \ldots, k_d, k_{d+1}) \mapsto (-k_1, \ldots, -k_d, k_{d+1}) \,.$$

The FKMM invariant for Q-bundles over $\widetilde{S}^d$ takes values in the quotient of equivariant homotopy classes $[(S^d)^\tau, \mathbb{C}^*]_{\mathbb{Z}_2} / [S^d, \mathbb{C}^*]_{\mathbb{Z}_2}$, where it is understood that the equivariance is

with respect to the involution τ on S^d and complex conjugation on $\mathbb{C}^*$. The quotient $[(S^d)^\tau, \mathbb{C}^*]_{\mathbb{Z}_2}/[S^d, \mathbb{C}^*]_{\mathbb{Z}_2}$ turns out to be very simple. For $d = 1$ it is trivial while for $d \geq 2$ it is $\mathbb{Z}_2$. It is easy to evaluate $[(S^d)^\tau, \mathbb{C}^*]_{\mathbb{Z}_2}$. First, by normalizing a representative in $[f] \in [(S^d)^\tau, \mathbb{C}^*]_{\mathbb{Z}_2}$ we see that $[S^d, \mathbb{C}^*]_{\mathbb{Z}_2} \cong [S^d, U(1)]_{\mathbb{Z}_2}$. Secondly, the time reversal sphere $\widetilde{S}^d$ has only two fixed points, the north pole $N = (0, \ldots, 0, 1)$ and the south pole $S = (0, \ldots, 0, -1)$. Moreover, at these fixed points, equivariance with respect to complex conjugation requires that $f(N)$ and $f(S)$ be real valued. Therefore, the set of equivariant homotopy classes $[(S^d)^\tau, \mathbb{C}^*]_{\mathbb{Z}_2}$ is isomorphic to $\mathrm{Map}(\{N, S\}, \{1, -1\})$. Since the constant map -1 is in $[S^d, \mathbb{C}^*]_{\mathbb{Z}_2}$ there are at most two classes in the quotient so that

$$[(S^d)^\tau, \mathbb{C}^*]_{\mathbb{Z}_2}/[S^d, \mathbb{C}^*]_{\mathbb{Z}_2} \subseteq \mathbb{Z}_2 .$$

The two possible classes are represented by maps $f_0, f_1 : (S^d)^\tau \to U(1)$ defined through

$$f_0(N) = f_0(S) = 1 \quad \text{and} \quad f_1(N) = -f_1(S) = 1 .$$

If $d = 1$, then there is a map in $g \in [S^1, \mathbb{C}^*]_{\mathbb{Z}_2}$ such that $g(N) = -g(S) = 1$. To see this, picture rotating the time reversal circle $\widetilde{S}^1$ with its involution $(k_1, k_2) \mapsto (-k_1, k_2)$ clockwise by $\pi/2$-radians so that it becomes the unit circle in $\mathbb{C}$ with complex conjugation as its involution. In other words, the rotated version of the time reversal circle is $U(1)$. The identity map $g = \mathrm{id} : U(1) \to U(1)$ is equivariant with respect to complex conjugation and has opposite signs at the two fixed points. Therefore,

$$[(S^1)^\tau, \mathbb{C}^*]_{\mathbb{Z}_2}/[S^1, \mathbb{C}^*]_{\mathbb{Z}_2} \cong \{0\} ,$$

which means that all Q-bundles over $\widetilde{S}^1$ are trivial.

Let us now consider the case $d = 2$. The claim is that the two classes defined by f_0 and f_1 are really distinct in this case. To show this, we have to argue that $[S^2, \mathbb{C}^*]_{\mathbb{Z}_2} \cong [S^2, U(1)]_{\mathbb{Z}_2}$ does not contain a map g with the property $g(N) = 1$ and $g(S) = -1$. We do this by contradiction. Suppose $g : S^2 \to U(1)$ is equivariant with respect to $(k_1, k_2, k_3) \mapsto (-k_1, -k_2, k_3)$ on S^2 and complex conjugation on $U(1)$, and satisfies $g(0, 0, 1) = 1$ and $g(0, 0, -1) = -1$. Consider a great circle C through the north and south poles. Because of the equivariance, the value of g on half of the great circle from N to S determines g on the other half. As we move $\boldsymbol{k}$ from N to S along the first half great circle, $g(\boldsymbol{k})$ follows a path γ in $U(1)$ which winds around a half integer amount of times since it has to start at 1 but end at -1. As we go from N to S along the opposite half great circle, $g(\boldsymbol{k})$ winds around $U(1)$ again a half integer amount of times, because we just have follow the complex conjugated path $\bar{\gamma}$. Now let's reverse the orientation of the conjugated path, so that it corresponds to going from S to N. Then, as we move $\boldsymbol{k}$ around the great circle, we first follow the path γ on $U(1)$ and then the path $(\bar{\gamma})^{-1}$. Crucially, the half integer portions of the winding do not cancel each other but rather add up to one full rotation around $U(1)$, as is illustrated in Fig. 9. The result is that $g|_C : C \to U(1)$ winds an odd number of times around $U(1)$. We conclude that that $g|_C : C \to U(1)$ has a nonzero winding number.

However, $g : S^2 \to U(1)$ is continuous. If we let D denote one of the hemispheres bounded by the great circle C, then the restriction $g|_D : D \to U(1)$ induces a homotopy from $g|_C$, a map with nonzero winding number from $S^1 \to S^1$, to a constant map. But we know that such a homotopy cannot exist, so we have reached a contradiction. We have thus established that there is *no* equivariant $g : S^2 \to U(1)$ such that $g(N) = 1$ and $g(S) = -1$. The same argument applies for S^d with $d \geq 3$ since we can reach the same contradiction as above by just considering the restriction g to $S^d \cap (\{0\}^{d-2} \times \mathbb{R}^3) \cong S^2$.

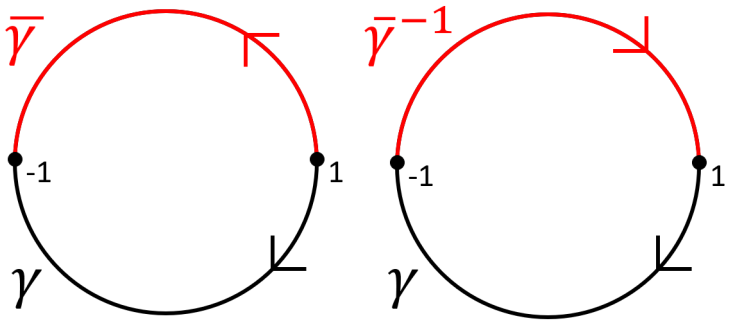

Figure 9: Left: Half integer winding portions of the paths γ and $\bar{\gamma}$ corresponding to moving $\boldsymbol{k}$ from N to S along the two different halves of the great circle C. Right: Concatenating half integer winding portions of γ and $\bar{\gamma}^{-1}$ leads to one full rotation around $U(1)$.

We have shown that when $d \geq 2$, maps in $[S^d, U(1)]_{\mathbb{Z}_2}$ necessarily have the same sign at the north and south poles, either both 1 or both -1. This means that we cannot change the relative sign of a map in $[(S^d)^\tau, U(1)]_{\mathbb{Z}_2}$ by multiplying with the restriction of a map from $[S^d, U(1)]_{\mathbb{Z}_2}$. In other words, the classes of f_0 and f_1 in the quotient $[(S^d)^\tau, \mathbb{C}^*]_{\mathbb{Z}_2}/[S^d, \mathbb{C}^*]_{\mathbb{Z}_2}$ are distinct which implies that

$$[(S^d)^\tau, \mathbb{C}^*]_{\mathbb{Z}_2}/[S^d, \mathbb{C}^*]_{\mathbb{Z}_2} \cong \mathbb{Z}_2\,.$$

The FKMM invariant for time reversal spheres in $d \geq 2$ is thus $\mathbb{Z}_2$ valued.

Recall that the Fu-Kane-Mele (FKM) invariant for time reversal symmetric topological insulators is also $\mathbb{Z}_2$ valued. However, it is not obvious that the FKM invariant and the FKMM invariant agree. Indeed, the FKM invariant has been called an "invariant" because it is independent of the choice of frame used in its construction. This only makes reference to the complex vector bundle structure though. To show that the FKM invariant agrees with the FKMM invariant, it must be shown the FKM invariant is actually invariant under Q-bundle isomorphisms.

Let us first note that the FKM invariant has ostensibly been defined in a more restrictive setting than the FKMM invariant. Indeed, the FKM invariant has been defined for valence bundles corresponding to time reversal symmetric topological insulators with no degeneracies other than Kramers'. When a Q-bundle (E, Θ) comes from such a valence bundle construction, then it is naturally equipped with a bundle metric with respect to which the time-reversal operator Θ is antiunitary. Moreover, since the first Chern number of E vanishes, the determinant line bundle is trivial so that, restricting to a single Kramers' pair, we may assume without essential loss of generality that E is a trivial $SU(2)$ bundles over S^d. However, in $d = 2, 3$ it is always the case that Q-bundles correspond to $SU(2)$ bundles such that the "Quaternionic" structure is antiunitary with respect to a bundle metric [18]. We may thus safely restrict ourselves to this case.

Suppose (E, Θ) and (E', Θ') are two Q-bundles over $\widetilde{S}^d$ for $d = 2$ or 3. Choose global orthonormal frames $\{\psi_1, \psi_2\}$ and $\{\psi'_1, \psi'_2\}$ for E and E' respectively. With respect to these frames, a Q-isomorphism $\varphi : (E, \Theta) \to (E', \Theta')$ can be interpreted as a matrix valued function $\Phi : S^d \to SU(2)$ whose components are defined through

$$\psi'_{a'}(\boldsymbol{k}) = \sum_{a=1}^{2} \Phi_{a'a}(\boldsymbol{k})\psi_a(\boldsymbol{k})\,.$$

The FKM invariant Eq. (7.7) of E' is constructed from the matrix

$$w'_{a'b'}(\boldsymbol{k}) = \langle \psi'_{a'}(-\boldsymbol{k}), \Theta' \psi'_{b'}(\boldsymbol{k}) \rangle \,.$$

Using the equivariance $\varphi \circ \Theta = \Theta' \circ \varphi$ it follows that

$$w'_{a'b'}(\boldsymbol{k}) = (\Phi^*)_{a'a}(-\boldsymbol{k}) w_{ab}(\boldsymbol{k}) (\Phi^T)_{bb'}(-\boldsymbol{k}) \,. \tag{8.2}$$

Since Φ is a unitary matrix, the matrices $\Phi^*(-\boldsymbol{k})$ and $\Phi^T(-\boldsymbol{k})$ are inverses of each other and so Eq. (8.2) implies

$$\det w'(\boldsymbol{k}) = \det w(\boldsymbol{k}) \,.$$

Therefore, $\sqrt{\det w'(0)}\sqrt{\det w'(\infty)} = \sqrt{\det w(0)}\sqrt{\det w(\infty)}$.

To show that the FKM invariants of (E, Θ) and (E', Θ') agree we need to gain control over the Pfaffians as well. Now, Pfaffians are only well-behaved with respect to orthogonal transformations, not similarity transformations as the one above. However, the Pfaffian $\operatorname{Pf} w(\boldsymbol{k})$ is a purely local quantity so we will show that we can always alter the frame $\{\psi_1, \psi_2\}$ by a gauge transformation $U : S^d \to SU(2)$ on E so that $\Phi_{a'a}(0) = \Phi_{a'a}(\infty) = \delta_{a'a}$. This then shows that $w(\boldsymbol{k})$ and $w'(\boldsymbol{k})$ are identical at the fixed points and thus their Pfaffians agree.

The way to choose the gauge transformation U is as follows. At $\boldsymbol{k} = 0$, we have that

$$\psi'_{a'}(0) = \sum_{a=1}^{2} \Phi_{a'a}(0) \psi_a(0) \,.$$

We thus need to choose $U(0) = \Phi^{-1}(0)$. Similarly, at $\boldsymbol{k} = \infty$, we need $U(\infty) = \Phi^{-1}(\infty)$. It only remains to find a continuous extension to S^d. Since $SU(2)$ is path connected, there is a continuous path from $\Phi^{-1}(0)$ to the identity matrix I. If we consider a little disk around $\boldsymbol{k} = 0$, then parameterizing the path in $SU(2)$ by the radius inside the disk gives us a continuous deformation of $\Phi^{-1}(0)$ at $\boldsymbol{k} = 0$ to I on the boundary of the disk. The same construction applies around $\boldsymbol{k} = \infty$. Simply extending U by the identity outside the two little disks around 0 and ∞ we obtain global continuous gauge transformation with the property that in this gauge, $\operatorname{Pf} w(0) = \operatorname{Pf} w'(0)$ and $\operatorname{Pf} w(\infty) = \operatorname{Pf} w'(\infty)$. By the gauge invariance of the FKM invariant over time reversal spheres, we have thus proven that the FKM invariants of (E, Θ) and (E', Θ') agree.

We conclude that the FKM invariant and the FKMM invariant both measure the non-triviality of a Q-bundle over the time reversal sphere. In physical terms, they are both an obstruction to finding a global basis of Kramers' pairs. Since they are $\mathbb{Z}_2$ invariants, they necessarily agree. Moreover, in $d = 2, 3$, the FKMM invariant even provides an isomorphism

$$\kappa : \operatorname{Vect}^2_Q(S^d, \tau) \to \mathbb{Z}_2,$$

because, on the one hand, it is injective, and, on the other hand, non-trivial Q-bundles exist so that $\operatorname{Vect}^2_Q(S^d, \tau) \neq 0$. For example, in $d = 2$ there is the quantum spin Hall effect and in $d = 3$ we have the Bi_2Se_3 model.[13]

8.4 The FKMM invariant for time reversal tori

We have established the FKMM in detail for "Quaternionic" bundles over the time reversal spheres. Let us here briefly discuss the FKMM invariant for Q-bundles over the time reversal torus $\widetilde{\mathbb{T}}^d = (\mathbb{T}^d, \tau)$. In $d = 2$, the torus $\mathbb{T}^d$ is represented as the square $[-\pi, \pi] \times [-\pi, \pi]$

[13] DeNittis and Gomi [18] establish the equivalence of FKM and FKMM in a slightly more general setting, which among other things includes the time reversal torus.

with opposite sides identified in the usual way, and the involution is given by $\tau(\boldsymbol{k}) = -\boldsymbol{k}$. There are four fixed points of τ, namely, $(0,0)$, $(\pi,0)$, (π,π) and $(0,\pi)$. The FKMM invariant takes values in the quotient $[(\mathbb{T}^2)^\tau, U(1)]_{\mathbb{Z}_2}/[\mathbb{T}^2, U(1)]_{\mathbb{Z}_2}$. Using elementary equivariant homotopy arguments, we can show that $[(\mathbb{T}^2)^\tau, U(1)]_{\mathbb{Z}_2}/[\mathbb{T}^2, U(1)]_{\mathbb{Z}_2} \cong \mathbb{Z}_2$.

First of all, $[(\mathbb{T}^2)^\tau, U(1)]_{\mathbb{Z}_2}$ consists of all different configurations ± 1's at the fixed points. In order to compute the quotient by $[\mathbb{T}^2, U(1)]_{\mathbb{Z}_2}$, we have to check how many of these configurations of signs can be attained by a global equivariant map $g : \mathbb{T}^2 \to U(1)$. It turns out that only sign configurations for which the product of all the signs is 1 can be attained by such a g. Since the constant map $-1 : \mathbb{T}^2 \to U(1)$ is in $[\mathbb{T}^2, U(1)]_{\mathbb{Z}_2}$, we may always fix the sign at $\boldsymbol{k} = (0,0)$ to be 1. There are then four distinct sign configurations for which the product of all the signs is 1. The three nontrivial configurations corresponding to nonconstant maps are listed in Fig. 10.

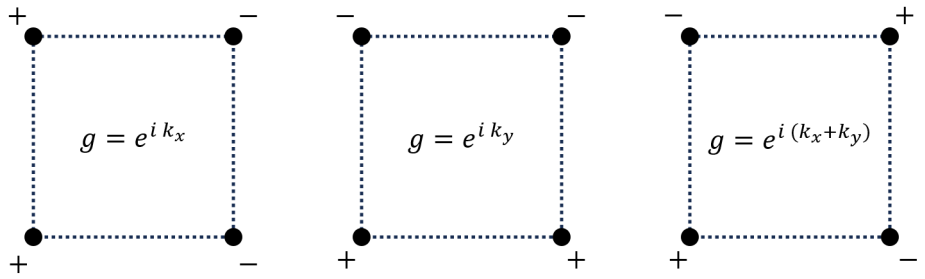

Figure 10: Sign configurations at the fixed points of $\widetilde{\mathbb{T}}^2$ that can be realized by globally equivariant maps. The fixed points are represented here as the vertices of the square $[0,\pi] \times [0,\pi]$.

However, there are no global equivariant maps $g : \mathbb{T}^2 \to U(1)$ that have an odd number of minus signs at the fixed points. To see this, consider the sign configuration displayed in Fig. 11, where now we display the full time reversal torus. Suppose now that $g : \mathbb{T}^2 \to U(1)$ is a globally continuous map that restricts to the sign configuration displayed in Fig. 11. Restricted to $k_x = 0$, the equivariance of g implies $g(0,-k_y) = \overline{g(0,k_y)}$. As k_y goes from 0 to π, the map $g(0,k_y)$ winds an integer number of times around $U(1)$. As k_y goes from 0 to $-\pi$, the map $g(0,k_y) = \overline{g(0,-k_y)}$ winds the same integer number times around $U(1)$, but now in the opposite direction. Reversing the orientation on the $[0,-\pi]$ segment, we find that $g(0,k_y)$ winds an *even* number of times around $U(1)$ as k_y winds around the circle at $k_x = 0$. Now, at $k_x = \pi$, the map $g(\pi,k_y)$ winds a half integer amount of times around $U(1)$. The map $g(\pi,k_y)$ winds a half integer amount of times in the opposite direction, because

$$g(\pi,k_y) = \overline{g(-\pi,-k_y)} = \overline{g(\pi,-k_y)}.$$

Similar to what we saw for the sphere in Fig. 9, the half integer winding portions add up when considering $g(\pi,k_y)$ as k_y goes from $-\pi$ to π, and so g winds around $U(1)$ an *odd* number of times when k_y winds around the circle at $k_x = \pi$. This is a contradiction. The winding number of g must be the same around $k_x = 0$ and $k_x = \pi$ because g is continuous. Therefore, no global equivariant g exists that realizes a sign configuration at the fixed points with an odd number of minus signs. We conclude that

$$[(\mathbb{T}^2)^\tau, U(1)]_{\mathbb{Z}_2}/[\mathbb{T}^2, U(1)]_{\mathbb{Z}_2} \cong \mathbb{Z}_2\,,$$

where the two distinct classes correspond to sign configuration with an even or odd number of minus signs.

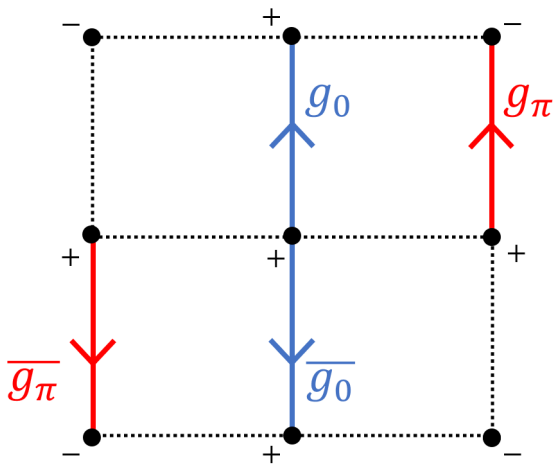

Figure 11: Sign configuration at the fixed points of $\widetilde{\mathbb{T}}^2$ displayed on the square $[-\pi, \pi] \times [-\pi, \pi]$. Opposite sides of the square are identified. The blue and red lines indicate restrictions of a global equivariant map $g : \mathbb{T}^2 \to U(1)$.

The above argument generalizes to higher dimensional tori roughly as follows. The set $[(\mathbb{T}^d)^\tau, U(1)]_{\mathbb{Z}_2}$ of all the sign configurations at the fixed points is isomorphic to $\mathbb{Z}_2^{2^d}$. Indeed, a d-dimensional cube has 2^d vertices and for each vertex we can pick either 1 or -1 in $\mathbb{Z}_2$. Just like in the two-dimensional case, every map $g_j(\boldsymbol{k}) = e^{ik_j}$ for $j = 1, \ldots, d$ and the constant map -1 identifies pairs of sign configurations. In fact, each of these $d+1$ maps reduces the total number of sign configurations by a factor of 2. This leads to following result for the quotient

$$[(\mathbb{T}^d)^\tau, U(1)]_{\mathbb{Z}_2} / [\mathbb{T}^d, U(1)]_{\mathbb{Z}_2} \cong \mathbb{Z}_2^{2^d} / \mathbb{Z}_2^{d+1} \cong \mathbb{Z}_2^{2^d - d - 1}.$$

This answer is in correspondence with the result of [18], which is obtained using advanced cohomological tools. In $d = 2$ our proof is complete, however, for $d > 2$ it should be established more convincingly that the maps g_j and -1 are the only maps that may be used to identify configurations.

The key ingredient in the construction of the FKMM invariant as a quotient of equivariant homotopy groups is that the determinant bundle associated to (E, Θ) is a trivial "Real" line bundle. So far, we have dealt with this simply by assuming that all "Real" line bundles over (X, τ) are trivial. It is, however, not obvious that this is the case for the time reversal sphere or torus. In the next section we will develop a cohomological classification of "Real" line bundles which allows us to justify this assumption, at least in the case of spheres.

9 Classification of "Real" line bundles

A "Real" line bundle (L, Θ) over an involutive space (X, τ) is a complex line bundle L together with a "Real" structure. Recall that the "Real" structure Θ is a homeomorphism $\Theta : L \to L$ that lifts the involution τ, acts antilinearly on fibers and squares to the identity. A key assumption in the construction of the FKMM invariant for "Quaternionic" vector bundles is the triviality of the associated determinant line bundle as a "Real" line bundle. In this section, we develop a classification of "Real" line bundles based on a generalization of Čech and sheaf cohomology to an equivariant setting. The main result is that "Real"

line bundles over an involutive space are classified by the equivariant cohomology group $H^2(X;\mathbb{Z}_2,\widetilde{\mathbb{Z}})$. For time reversal spheres, we can explicitly show that this cohomology group vanishes which means that all "Real" line bundles over the time reversal sphere are trivial. This justifies the construction of the FKMM invariant in the previous section for continuum models.

9.1 Čech cohomology

Čech cohomology is a tool to recover topological information about a space from combinatorial data associated to an open cover of the space. The combinatorial data consists of so-called "sections of sheaves" over the overlaps of open sets in the cover. Hence, before we can define Čech cohomology we have to introduce the notion of a sheaf. Sheaves capture a particular property of functions. Namely, if X is a topological space and $f_1 : U_1 \to \mathbb{R}$ and $f_2 : U_2 \to \mathbb{R}$ are two functions such that $f_1|_{U_1\cap U_2} = f_1|_{U_1\cap U_2}$, then there is a unique function $f : U_1 \cup U_2 \to \mathbb{R}$ that restricts to f_1 on U_1 and to f_2 on U_2. We say that functions defined over different open set can be "glued" together when they agree on the overlaps. Sheaves axiomatize this notion of gluing, and generalize it to more general objects than just real valued functions.

It is customary to introduce sheaves in two stages. The first stage is to define what is called a presheaf. A *presheaf* $\mathcal{F}$ of abelian groups over a topological space X consists of the following data:

1. an abelian group $\mathcal{F}(U)$ for every open set $U \subseteq X$,
2. a map $\rho^U_V : \mathcal{F}(U) \to \mathcal{F}(V)$ for every inclusion of opens $V \subseteq U$.

The maps ρ^U_V are called "restriction" maps and they are required to satisfy $\rho^U_U = \mathrm{id}$ and $\rho^V_W \circ \rho^U_V = \rho^U_W$ for any three opens $W \subseteq V \subseteq U$. The elements of $\mathcal{F}(U)$ are called "sections over U" and we will sometimes denote the abelian group $\mathcal{F}(U)$ by $\Gamma(U,\mathcal{F})$. For us the most important presheaves are the presheaves of functions with values in some abelian group A. For example, if X is any topological space, the constant functions with values in $\mathbb{Z}$ are a presheaf. We denote this presheaf by $\mathbb{Z}^0_X$. The continuous functions with values in $\mathbb{Z}$ are also a presheaf that we denote by $\mathbb{Z}_X$. Another example of a presheaf is the continuous functions with values in the abelian group $(\mathbb{C},+)$. We denote this presheaf by $\mathcal{O}_X$. Explicitly,

1. for $U \subseteq X$ open, $\mathcal{O}_X(U) = C^0(U,\mathbb{C})$,
2. for $V \subseteq U$ an inclusion of opens, $\rho^U_V : \mathcal{O}_X(U) \to \mathcal{O}_X(V)$ is just given by the restriction of functions.

The second stage is to restrict the notion of presheaf to that of sheaf. Let $\mathcal{F}$ be a presheaf, then $\mathcal{F}$ is called a *sheaf* if it satisfies the following two properties. Let $U \subseteq X$ be open and let $\mathcal{U} = \{U_\alpha\}_{\alpha\in\mathcal{A}}$ be an open cover of U,

1. if $s,t \in \mathcal{F}(U)$ are sections such that $\rho^U_{U_\alpha}(s) = \rho^U_{U_\alpha}(t)$ for all $\alpha \in \mathcal{A}$, then $s = t$,
2. if $\{s_\alpha \in \mathcal{F}(U_\alpha)\}_{\alpha\in\mathcal{A}}$ is a collection of sections such that $\rho^{U_\alpha}_{U_\alpha\cap U_\beta}(s_\alpha) = \rho^{U_\beta}_{U_\alpha\cap U_\beta}(s_\beta)$ for all $\alpha,\beta \in \mathcal{A}$, then there is $s \in \mathcal{F}(U)$ such that $\rho^U_{U_\alpha}(s) = s_\alpha$ for all $\alpha \in \mathcal{A}$.

The first property states that if two sections agree everywhere locally, then they agree globally. The second property states that if we have a collection of local sections that agree

on overlaps, then they glue to a global section. The combination of the two properties states that this gluing is unique.

The presheaf $\mathcal{O}_X$ introduced above satisfies both of these properties and so $\mathcal{O}_X$ is a sheaf. The presheaf $\mathbb{Z}^0_X$ of constant integer valued functions is in general *not* a sheaf because if X contains two disjoint opens subsets,[14] then the second sheaf property fails. For example, if X is the set $\{0,1\}$ with the discrete topology, then the sets $\{0\}$ and $\{1\}$ are two opens whose union is X. Consider the local sections $s_0 = 7$ and $s_1 = 3$ on $\{0\}$ and $\{1\}$, respectively. There is no section $s : \{0,1\} \to \mathbb{Z}$ that restricts to 7 on $\{0\}$ and to 3 on $\{1\}$ because any global sections s has to be constant on X. On the other hand, the presheaf $\mathbb{Z}_X$, of continuous, integer-valued functions, does not have this problem. Indeed, the sections of $\mathbb{Z}_X$ are the *locally* constant functions, which may take different values on disjoint opens, and so $\mathbb{Z}_X$ does form a sheaf.

The sheaves we will deal with are usually of the type of locally constant functions or continuous functions with values in some abelian group such as $\mathbb{Z}$ or $\mathbb{C}$. The advantage of still using the language of abstract sheaves is two-fold. On the one hand, we can develop the theory of Čech cohomology with coefficients in an arbitrary sheaf $\mathcal{F}$ and then at the end plug in $\mathcal{O}_X$ or $\mathbb{Z}_X$ or a similar sheaf. On the other hand, Čech cohomology can be viewed as an approximation to the more abstract notion of sheaf cohomology, which will be a useful tool later for proving theorems about Čech cohomology and doing computations.

With this short introduction to sheaves out of the way, let us now define Čech cohomology with coefficients in a sheaf. The story of Čech cohomology starts with the notion of Čech cohomology "with respect to an open cover". Fix a topological space X and a sheaf $\mathcal{F}$ on X. Choose an open cover $\mathcal{U} = \{U_\alpha\}_{\alpha\in\mathcal{A}}$ of X. To this data we associate a cochain complex $(\check{C}^*(\mathcal{U};\mathcal{F}),\delta)$ called the *Čech cochain complex with respect to the open cover* $\mathcal{U}$. The Čech cochain group in degree n is

$$\check{C}^n(\mathcal{U};\mathcal{F}) = \Big\{(f_{\alpha_0\ldots\alpha_n}) \in \prod_{\alpha_0,\ldots,\alpha_n} \Gamma(U_{\alpha_0\ldots\alpha_n},\mathcal{F}) :\ f_{\alpha_0\ldots\alpha_i\alpha_{i+1}\ldots\alpha_n} = -f_{\alpha_0\ldots\alpha_{i+1}\alpha_i\ldots\alpha_n}\Big\}$$

where the notation $U_{\alpha_0\ldots\alpha_n}$ means to take the intersection $U_{\alpha_0}\cap\cdots\cap U_{\alpha_n}$. In other words, an n-Čech cochain is defined by taking a section of the sheaf $\mathcal{F}$ over every $n+1$-fold overlap of the cover in an antisymmetrized way.

The Čech differential $\delta : \check{C}^n(\mathcal{U};\mathcal{F}) \to \check{C}^{n+1}(\mathcal{U};\mathcal{F})$ is given by the formula

$$(\delta f)_{\alpha_0\ldots\alpha_{n+1}} = \sum_{i=0}^{n+1}(-1)^i\rho_i(f_{\alpha_0\ldots\hat{\alpha}_i\ldots\alpha_{n+1}})\,,$$

where ρ_i is the restriction from $U_{\alpha_0\ldots\hat{\alpha}_i\ldots\alpha_n}$ to $U_{\alpha_0\ldots\alpha_n}$, and the hat on top of an index indicates that we skip that index. The *Čech cohomology with respect to the open cover* $\mathcal{U}$ with values in the sheaf $\mathcal{F}$ is the cohomology of the Čech cochain complex $(\check{C}^*(\mathcal{U};\mathcal{F}),\delta)$ for the cover $\mathcal{U}$. We denote this cohomology by $\check{H}^*(\mathcal{U};\mathcal{F})$.

The Čech cohomology in degree 0 is equal to the global sections of the sheaf $\mathcal{F}$. Let's see how this comes about. By definition,

$$\check{H}^0(\mathcal{U};\mathcal{F}) = \ker(\delta : \check{C}^0(\mathcal{U};\mathcal{F}) \to \check{C}^1(\mathcal{U};\mathcal{F}))\,.$$

Let $(f_\alpha) \in \check{C}^0(\mathcal{U};\mathcal{F})$, i.e. for every open $U_\alpha \in \mathcal{U}$, we specify a section f_α of $\mathcal{F}$ over U_α. For every double overlap $U_{\alpha_0\alpha_1}$, we have

$$(\delta f)_{\alpha_0\alpha_1} = \rho^{\alpha_1}_{\alpha_0\alpha_1}(f_{\alpha_1}) - \rho^{\alpha_0}_{\alpha_0\alpha_1}(f_{\alpha_0})\,.$$

[14] Pretty much any sensible topological space which is not the point contains two disjoint opens.

The cochain (f_α) is in the kernel of δ if and only if $(\delta f)_{\alpha_0\alpha_1} = 0$ for all double overlaps $U_{\alpha_0\alpha_1}$. Now notice that $(\delta f)_{\alpha_0\alpha_1} = 0$ if and only if $\rho^{\alpha_1}_{\alpha_0\alpha_1}(f_{\alpha_1}) = \rho^{\alpha_0}_{\alpha_0\alpha_1}(f_{\alpha_0})$. So, if (f_α) is in $\ker\delta$, then the collection (f_α) of local sections agree on overlaps. By the sheaf property there is thus a unique global section $f \in \Gamma(X,\mathcal{F})$ such that $f|_{U_\alpha} = f_\alpha$ for all $\alpha \in \mathcal{A}$. In other words, $\check{H}^0(\mathcal{U};\mathcal{F}) = \Gamma(X,\mathcal{F})$.

Interestingly, the degree zero Čech cohomology with respect to the cover $\mathcal{U}$ is independent of the cover. This is not true in higher degrees. In general, the cohomology $H^*(\mathcal{U};\mathcal{F})$ depends on the choice of open cover. To obtain a topological invariant that is intrinsic to the space X we need to consider what happens to $\check{H}^*(\mathcal{U};\mathcal{F})$ when we make the cover $\mathcal{U}$ arbitrarily fine so that it captures all the topological features of X.

If $\mathcal{U} = \{U_\alpha\}_{\alpha\in\mathcal{A}}$ and $\mathcal{V} = \{V_\beta\}_{\beta\in\mathcal{B}}$ are two open covers of X, then we say that $\mathcal{U}$ *refines* $\mathcal{V}$ if for every $\alpha \in \mathcal{A}$ we can find a $\beta \in \mathcal{B}$ such that $U_\alpha \subseteq V_\beta$. Any such choice of association of β's to α's defines a map $r : \mathcal{A} \to \mathcal{B}$, which we call a refinement map. We also say $\mathcal{U} \geq \mathcal{V}$ if $\mathcal{U}$ refines $\mathcal{V}$. A refinement map r defines a map r^* on Čech cochains,

$$r^* : \check{C}^n(\mathcal{V};\mathcal{F}) \to \check{C}^n(\mathcal{U};\mathcal{F}), \qquad (r^*f)_{\alpha_0\ldots\alpha_n} = \rho^V_U(f_{r\alpha_0\ldots r\alpha_n}),$$

where ρ^V_U is the restriction from $V_{r\alpha_0\ldots r\alpha_n}$ to $U_{\alpha_0\ldots\alpha_n}$. It is easily seen that r^* commutes with the differential δ, so that r^* is in fact a map of cochain complexes. Therefore, there is an induced map in cohomology

$$r^* : \check{H}^*(\mathcal{V};\mathcal{F}) \to \check{H}^*(\mathcal{U};\mathcal{F}), \qquad [f] \mapsto [r^*f].$$

It can be shown that the induced map on cohomology is independent of the choice of refinement map $r : \mathcal{A} \to \mathcal{B}$. If r and $\tilde{r}$ are two refinement maps for $\mathcal{U} \geq \mathcal{V}$, then

$$k : \check{C}^{n+1}(\mathcal{V};\mathcal{F}) \to \check{C}^n(\mathcal{U};\mathcal{F}), \qquad (kf)_{\alpha_0\ldots\alpha_n} = \sum_{i=0}^{n}(-1)^i\rho^V_U(f_{r\alpha_0\ldots r\alpha_i\tilde{r}\alpha_i\ldots\tilde{r}\alpha_n}),$$

where ρ^V_U is the appropriate restriction map, can be verified to be a cochain homotopy from r^* to $\tilde{r}^*$ by a direct, but somewhat lengthy computation [39]. This implies that $r^* = \tilde{r}^*$.

The notion of "arbitrarily fine cover" is defined as the limit of refinements of covers. Consider the directed set $(\mathcal{K}_X, \leq)$ of all open covers

$$\mathcal{K}_X = \{\mathcal{U} : \mathcal{U} \text{ is open cover of } X\}$$

together with the refinement relation $\leq$.[15] We define the Čech cohomology of X with coefficients in $\mathcal{F}$ as

$$\check{H}^*(X;\mathcal{F}) = \varinjlim \check{H}^*(\mathcal{U};\mathcal{F}),$$

where the limit is a direct limit taken over refinements of open covers of X. The direct limit is defined as follows. We take the disjoint union of the cohomologies $\check{H}^*(\mathcal{U};\mathcal{F})$ over all open covers and divide out by an equivalence relation induced by the refinement maps. More precisely,

$$\varinjlim \check{H}^*(\mathcal{U};\mathcal{F}) = \bigsqcup_{\mathcal{U}\in\mathcal{K}_X} \check{H}^*(\mathcal{U};\mathcal{F})/\sim,$$

where $[f] \in \check{H}^*(\mathcal{U};\mathcal{F})$ is equivalent to $[g] \in \check{H}^*(\mathcal{V};\mathcal{F})$ if there is some common refinement $\mathcal{W}$ of $\mathcal{U}$ and $\mathcal{V}$ with refinement maps $r^* : \check{H}^*(\mathcal{U};\mathcal{F}) \to \check{H}^*(\mathcal{W};\mathcal{F})$ and $s^* : \check{H}^*(\mathcal{V};\mathcal{F}) \to$

[15] Technically, the open covers of a space X do not form a set, because there are too many open covers. This issue can be dealt with. For the sake of explaining the ideas, we will ignore this technicality here and refer for the details to [40].

$\check{H}^*(\mathcal{W};\mathcal{F})$ such that $r^*[f] = s^*[g]$. Intuitively, two elements are equivalent if they eventually become equal when the open cover is fine enough.

To check that $\sim$ is actually an equivalence relation we use the following key observation. We have seen that if $\mathcal{V} \leq \mathcal{U}$ then there is a map $r^* : \check{H}^*(\mathcal{V};\mathcal{F}) \to \check{H}^*(\mathcal{U};\mathcal{F})$ which is induced by choosing *any* refinement map $r : \mathcal{U} \to \mathcal{V}$. So , if $\mathcal{U}$ and $\mathcal{V}$ are equal, we can take $\mathrm{id} : \mathcal{A} \to \mathcal{A}$ as refinement map which induces the identity on cohomology as well. This will give the reflexivity of $\sim$. Moreover, if $\mathcal{U} \geq \mathcal{V} \geq \mathcal{W}$ for some third cover $\mathcal{W} = \{W_\gamma\}_{\gamma\in\mathcal{C}}$ and we have refinement maps $r : \mathcal{A} \to \mathcal{B}$ and $s : \mathcal{B} \to \mathcal{C}$, then $s \circ r : \mathcal{A} \to \mathcal{C}$ is a refinement map for $\mathcal{U} \geq \mathcal{W}$. We have that $(s \circ r)^* = r^* \circ s^*$ for the induced maps in cohomology since this holds already at the cochain level. This will give the transitivity of $\sim$. The fact that $\sim$ is symmetric follows immediately from the construction.

Let us now demonstrate what the Čech cohomology defined above can do for us. Our primary interest in Čech cohomology is that it is a useful tool for classifying line bundles. Recall that we classified complex line bundles over a space X by the first Chern class in $H^2(X;\mathbb{Z})$. As we will see now, Čech cohomology also gives a classification of complex line bundles.

Proposition 9.1. Let X be a space and let $\mathcal{O}_X^*$ be the sheaf of non-vanishing, continuous complex valued functions on X. There is a canonical bijection

$$\mathrm{Vect}^1_{\mathbb{C}}(X) \cong \check{H}^1(X;\mathcal{O}_X^*)\,.$$

Proof. We first construct a map $L : \check{H}^1(X;\mathcal{O}_X^*) \to \mathrm{Vect}^1_{\mathbb{C}}(X)$ that associates a complex line bundle to a class $c \in \varinjlim \check{H}^1(\mathcal{U};\mathcal{O}_X^*)$. Let $c \in \check{H}^1(X;\mathcal{O}_X^*)$. A basic property of the direct limit is that the class c in the direct limit can be represented by a single cocycle $(g_{\alpha\beta}) \in \check{H}^1(\mathcal{U};\mathcal{O}_X^*)$ for some cover $\mathcal{U} = \{U_\alpha\}_{\alpha\in\mathcal{A}}$. We construct a line bundle from this cocycle by gluing copies of $U_\alpha \times \mathbb{C}$ together using the $g_{\alpha\beta}$. More precisely, we define the space

$$\widetilde{L}(g) = \bigsqcup_{\alpha\in\mathcal{A}} U_\alpha \times \mathbb{C}$$

and the equivalence relation

$$U_\alpha \times \mathbb{C} \ni (x,z) \sim (y,w) \in U_\beta \times \mathbb{C} \iff (y,w) = (x, g_{\alpha\beta}(x)z)\,.$$

The relation $\sim$ is an equivalence relation precisely because $(g_{\alpha\beta})$ is a Čech cocycle. Indeed, $\sim$ is reflexive since $g_{\alpha\alpha} = 1$, the identity element in the multiplicative group $\mathbb{C}^*$, by skew-symmetry of the Čech indices. It is symmetric since $g_{\alpha\beta}g_{\beta\alpha} = g_{\alpha\alpha} = 1$ by $(\delta g)_{\alpha\beta\alpha} = 1$. It is transitive since $g_{\beta\gamma}g_{\alpha\beta} = g_{\alpha\gamma}$ because $(\delta g)_{\alpha\beta\gamma} = 1$. We can thus take the quotient of $\widetilde{L}(g)$ by $\sim$. We claim that the map

$$c \mapsto L(c) := \widetilde{L}(g)/\sim$$

is well-defined, in the sense that a different choice of $(g_{\alpha\beta})$ to represent c yields an isomorphic line bundle. Suppose $(g'_{\alpha\beta})$ is another cocycle over the same cover $\mathcal{U}$ that represents the same class in $\check{H}^1(\mathcal{U};\mathcal{O}_X^*)$ then there is an $(f_\alpha) \in \check{C}^0(\mathcal{U};\mathcal{O}_X^*)$ such that $g'_{\alpha\beta} = f_\beta g_{\alpha\beta} f_\alpha^{-1}$. Indeed, this is precisely saying that $g' = g + \delta f$, but then in multiplicative notation. The map $\widetilde{F} : \widetilde{L}(g') \to \widetilde{L}(g)$ defined by

$$U_\alpha \times \mathbb{C} \ni (x,z) \mapsto (x, f_\alpha(x)z) \in U_\alpha \times \mathbb{C}$$

descends to the quotients by $\sim$ and gives an isomorphism between the line bundles $\widetilde{L}(g)/\sim$ and $\widetilde{L}(g')/\sim$. Now suppose that $(g'_{\alpha'\beta'})$ is a cocycle over some possibly different cover

$\mathcal{U}' = \{U_{\alpha'}\}_{\alpha' \in \mathcal{A}'}$ such that $(g'_{\alpha'\beta'}) \in \check{H}^1(\mathcal{U}'; \mathcal{O}_X^*)$ also represents $c \in \check{H}^1(X; \mathcal{O}_X^*)$. In that case, there must be a some cover $\mathcal{V} = \{V_i\}_{i \in \mathcal{I}}$ and refinement maps $r : \mathcal{A} \to \mathcal{I}$ and $r' : \mathcal{A}' \to \mathcal{I}$ such that rg and $r'g'$ differ by a coboundary δf. So by the same argument as above, we will at least have $(\widetilde{L}(rg)/\sim) \cong (\widetilde{L}(r'g')/\sim)$. It remains to argue that the line bundle $\widetilde{L}(g)/\sim$ is stable under refinement. In other words, that $\widetilde{L}(g)/\sim$ and $\widetilde{L}(rg)/\sim$ are isomorphic. This is indeed the case because the refined bundle is obtained from the original bundle by just restricting the transition functions we already had. This does not change the line bundle. The map $L : c \mapsto \widetilde{L}(g)/\sim$ is thus well-defined.

The map L is injective because if c and c' are two cochain represented by $g \in \check{H}(\mathcal{U}; \mathcal{O}_X)$ and $g' \in \check{H}(\mathcal{U}'; \mathcal{O}_X)$, respectively, such that $\widetilde{L}(g)/\sim$ and $\widetilde{L}(g')/\sim$ are isomorphic, then over a common refinement of $\mathcal{U}$ and $\mathcal{U}'$ this isomorphism gives rise to a cochain (f) such that $r'g' = rg\delta f$. Finally, the map L is surjective because given any line bundle E we can pick a trivializing cover $(\mathcal{U}, (\varphi_\alpha))$ with transition functions $(g_{\alpha\beta})$ and show that $\widetilde{L}(g)/\sim$ is isomorphic to E via

$$U_\alpha \times \mathbb{C} \ni (x, z) \mapsto \varphi_\alpha^{-1}(x, z)\,.$$

We conclude that the map L constructed above realizes the required bijection.

□

9.2 Sheaf cohomology

We have defined the Čech cohomology $\check{H}^*(X; \mathcal{F})$ of the space X with coefficients in the sheaf $\mathcal{F}$ by taking the direct limit over Čech cohomologies of open covers $\mathcal{U}$ of X. However, even for a topological space as simple as S^d we can never compute this direct limit because there are just way to many open covers to consider. Also, it might be very hard to even compute $\check{H}^*(\mathcal{U}; \mathcal{F})$ for a single open cover because there could be intractably many overlaps to consider. If we ever want to compute something, we need more tools. To get to these tools, we will review sheaf cohomology, which provides another point of view on the cohomology $\check{H}^*(X; \mathcal{F})$.

We start by defining the proper notion of a map between sheaves. Let $\mathcal{F}$ and $\mathcal{G}$ be two sheaves, then a *morphism of sheaves* $\varphi : \mathcal{F} \to \mathcal{G}$ is given by specifying a morphism of abelian groups $\varphi(U) : \mathcal{F}(U) \to \mathcal{G}(U)$ for every open $U \subseteq X$ such that the $\varphi(U)$ are compatible with the restriction maps of $\mathcal{F}$ and $\mathcal{G}$. More precisely, if $V \subseteq U$ is an inclusion of opens, we require

$$(\rho_{\mathcal{G}})^U_V \circ \varphi(U) = \varphi(V) \circ (\rho_{\mathcal{F}})^U_V\,,$$

where $\rho_{\mathcal{F}}$ and $\rho_{\mathcal{G}}$ are the restriction maps of $\mathcal{F}$ and $\mathcal{G}$, respectively.

To formulate the notions of injectivity and surjectivity for maps of sheaves, we need to introduce the notion of stalks. Let $\mathcal{F}$ be a sheaf over X. We define the stalk $\mathcal{F}_x$ of $\mathcal{F}$ at the point $x \in X$ by describing its elements which are called germs. A *germ* of $\mathcal{F}$ at $x \in X$ is an equivalence class $[(U, s)]$ of pairs consisting of an open neighbourhood U of x and a section s over U. Two pairs (U, s) and (U', s') are equivalent if there is a neighbourhood $V \subseteq U \cap U'$ of x such that

$$\rho^U_V(s) = \rho^{U'}_V(s')\,.$$

The *stalk* of $\mathcal{F}$ at $x \in X$ is the set of all germs at x. If $\mathcal{F}$ is a sheaf of abelian groups, then the stalk $\mathcal{F}_x$ is an abelian group. Indeed, the group operation on sections carries over to the germs.[16]

[16]The definition of stalk can be phrased more concisely in terms of the direct limit that we saw before when defining Čech cohomology. The stalk $\mathcal{F}_x$ of $\mathcal{F}$ at a point $x \in X$ can be defined as

$$\mathcal{F}_x = \varinjlim \mathcal{F}(U)\,,$$

A map of sheaves $\varphi : \mathcal{F} \to \mathcal{G}$ induces maps on stalks $\varphi_x : \mathcal{F}_x \to \mathcal{G}_x$ for all $x \in X$. If $[(U, s)] \in \mathcal{F}_x$, then

$$\varphi_x([(U, s)]) = [(U, \varphi(U)(s))] .$$

We say that $\varphi : \mathcal{F} \to \mathcal{G}$ is injective whenever $\varphi_x : \mathcal{F}_x \to \mathcal{G}_x$ is injective for all $x \in X$. Similarly, φ is surjective if φ_x is surjective for all $x \in X$. Moreover, if $\mathcal{F} \to \mathcal{G} \to \mathcal{H}$ is a sequence of sheaves, then it is exact if and only if the sequence of stalks $\mathcal{F}_x \to \mathcal{G}_x \to \mathcal{H}_x$ is exact for all $x \in X$.[17]

The key ingredient in the definition of sheaf cohomology is that of an injective resolution of a sheaf. If $\mathcal{F}$ is a sheaf, then a *resolution* of $\mathcal{F}$ is an exact sequence of sheaves of the form

$$0 \longrightarrow \mathcal{F} \longrightarrow \mathcal{F}^0 \xrightarrow{d} \mathcal{F}^1 \xrightarrow{d} \mathcal{F}^2 \longrightarrow \dots .$$

A resolution of $\mathcal{F}$ is an *injective resolution* if each $\mathcal{F}^n$ is an injective sheaf. An injective sheaf is a sheaf satisfying a certain universal property. At the moment, it is not so important what this universal property precisely entails. The important points are the following. First of all, injective sheaves have some nice properties with regards to extending local sections to global sections. We will get into this later. Secondly, any sheaf $\mathcal{F}$ injects into an injective sheaf $\mathcal{I}$ [41]. This fact is usually referred to by saying that the category of sheaves has "enough injectives".

A sheaf $\mathcal{F}$ always admits a resolution of injective sheaves. Indeed, since there are enough injectives we can find an injective sheaf $\mathcal{I}^0$ such that $\mathcal{F}$ injects into $\mathcal{I}^0$. There is thus a short exact sequence

$$0 \longrightarrow \mathcal{F} \longrightarrow \mathcal{I}^0 \longrightarrow \mathcal{C}^0 \longrightarrow 0 ,$$

where $\mathcal{C}^0 = \mathcal{I}^0 / \operatorname{im} \mathcal{F}$ is the cokernel of $\mathcal{F} \to \mathcal{I}^0$. The definition of the cokernel sheaf can be made precise by the method of so-called sheafification. However, for our purposes it suffices to know that it exists, because now we can find $\mathcal{I}^1$ such that $\mathcal{C}^0$ injects into $\mathcal{I}^1$. We can thus extend the diagram above to the following diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathcal{F} & \longrightarrow & \mathcal{I}^0 & \longrightarrow & \mathcal{C}^0 & \longrightarrow & 0 \\ & & & & & \searrow^{d} & & \searrow & \\ & & & & & & \mathcal{I}^1 & \longrightarrow & \mathcal{C}^1 \longrightarrow 0 \end{array} ,$$

where $\mathcal{C}^1$ is the cokernel of $\mathcal{C}^0 \to \mathcal{I}^1$. Since $\mathcal{C}^0 \to \mathcal{I}^1$ is injective, the kernel of the dashed arrow d is precisely the kernel of $\mathcal{I}^0 \to \mathcal{C}^0$. It then follows from the exactness of the top row that

$$\ker(\mathcal{I}^0 \to \mathcal{I}^1) = \operatorname{im}(\mathcal{F} \to \mathcal{I}^0) .$$

The sequence $0 \to \mathcal{F} \to \mathcal{I}^0 \to \mathcal{I}^1 \to \mathcal{C}^1 \to 0$ is thus exact. Repeating the construction we end up with an exact sequence of sheaves

$$0 \longrightarrow \mathcal{F} \longrightarrow \mathcal{I}^0 \xrightarrow{d} \mathcal{I}^1 \xrightarrow{d} \mathcal{I}^2 \xrightarrow{d} \dots \tag{9.1}$$

which is the injective resolution of $\mathcal{F}$ we were looking for.

---

where the direct limit is taken over all open neighbourhoods U of X.

[17] One might wonder why we choose to define these notions of injectivity and surjectivity via stalks. And indeed, injectivity can also be defined via the so-called kernel sheaf and surjectivity via the image sheaf. The definition of the image sheaf, however, uses the notion of sheafification. Sheafification, in turn, uses stalks. For ease of presentation, we gave the definitions immediately in terms of stalks.

Sheaf cohomology is now defined as follows. Let $\mathcal{F}$ be a sheaf over a space X and take an injective resolution of $\mathcal{F}$ as constructed above. Since the injective resolution is exact, $d^2 = 0$ as a map between sheaves so that

$$0 \longrightarrow \mathcal{I}^0(X) \xrightarrow{d(X)} \mathcal{I}^1(X) \xrightarrow{d(X)} \mathcal{I}^2(X) \xrightarrow{d(X)} \dots \tag{9.2}$$

defines a cochain complex. Although the injective resolution Eq. (9.1) is an exact sequence of sheaves, the cochain complex (9.2) is in general not an exact sequence. The exactness of the resolution only implies that the sequence of stalks

$$\mathcal{I}^0_x \xrightarrow{d_x} \mathcal{I}^1_x \xrightarrow{d_x} \mathcal{I}^2_x \xrightarrow{d_x} \dots$$

is exact for all $x \in X$. A section $s \in \mathcal{I}^n(X)$ in the kernel of $d(X)$ thus locally always has a primitive, in the sense that for each $x \in X$ there is $t_x \in \mathcal{I}^{n-1}_x$ such that $d_x t_x = s_x$. However, these local primitives $(t_x)_{x\in X}$ might not glue to a global section $t \in \mathcal{I}^{n-1}(X)$. The cochain Eq. (9.2) therefore possible has non-trivial cohomology and this cohomology is what we define to be the *sheaf cohomology* $H^*(X;\mathcal{F})$. The construction of sheaf cohomology is a special case of the construction of a so-called "right derived functor" and is therefore independent of the choice of injective resolution [40].

It is a nontrivial fact, of which we sketch the proof in Appendix B, that sheaf cohomology and Čech cohomology are isomorphic over paracompact Hausdorff spaces [40, Theorem 6.88]. We thus restrict ourselves to paracompact Hausdorff spaces from now on.

Using the definition we can now compute the sheaf cohomology of the sheaf $\mathbb{Z}_{\{*\}}$ over the point. The first task is to construct an injective resolution. The sheaf $\mathbb{Z}_{\{*\}}$ can be identified with the abelian group $\mathbb{Z}$, and for groups being injective is the same as being divisible [42, Tag 01D6]. A group G is divisible if for every $g \in G$ and $n \in \mathbb{N}$ there exists $y \in G$ such that $ny = g$. For example, the additive group of rational numbers $\mathbb{Q}$ is divisible and, hence, injective. The quotient $\mathbb{Q}/\mathbb{Z}$ is also injective. So, the short exact sequence

$$0 \longrightarrow \mathbb{Z} \longrightarrow \mathbb{Q} \longrightarrow \mathbb{Q}/\mathbb{Z} \longrightarrow 0$$

realizes an injective resolution of $\mathbb{Z}$. The sheaf cohomology $H^*(\{*\};\mathbb{Z}_{\{*\}})$ is thus equal to the cohomology of $\mathbb{Q} \to \mathbb{Q}/\mathbb{Z} \to 0$. The zeroth degree cohomology is $H^0(\{*\};\mathbb{Z}_{\{*\}}) = \mathbb{Z}$ and all cohomology groups in positive degree vanish. In fact, the same proof works for any abelian group, because we can always embed an abelian group into an injective group and a quotient of an injective group is injective.

A resolution of the sheaf $\mathbb{Z}_X$ over a space X which is not just a point could yield an unwieldy object. However, if X is contractible, then the following theorem still allows us to compute the sheaf cohomology of $\mathbb{Z}_X$.

Theorem 9.2. Sheaf cohomology with values in a constant sheaf is homotopy invariant.

See [41] for a proof. The implication of Theorem 5.1 is that if X is a contractible space, then $H^n(X;\mathbb{Z}_X) = H^n(\{*\};\mathbb{Z}_{\{*\}})$, which we have just computed above.

Getting your hands on an injective resolution of a sheaf $\mathcal{F}$ is in general a difficult task. So, just like Čech cohomology, computing sheaf cohomology is thus also difficult and it seems as though we have not made any progress. However, as is usual in cohomology theory one does not use the definitions to compute cohomology, we want to use theorems to compute cohomology. Remember that one of the issues holding us back from computing Čech cohomology is that we cannot take the direct limit over all refinements of covers of a space, unless maybe the space consists of a countable number of points. By relating Čech and sheaf cohomology, it turns out that it is possible to choose a "good" cover of X which already gives us the full Čech cohomology. This is the content of Leray's theorem.

Theorem 9.3 (Leray)**.** If $\mathcal{U}$ is an open cover of X such that $H^n(U_{\alpha_0\ldots\alpha_k};\mathcal{F}) = 0$ for all $k \geq 0$ and $n \geq 1$, then $\check{H}^n(\mathcal{U};\mathcal{F}) \cong H^n(X;\mathcal{F})$ for $n \geq 0$.

This statement can for example be found in [43]. It can be proven by considering the double complex $C^{p,q} = \check{C}^p(\mathcal{U};\mathcal{I}^q)$, where $0 \to \mathcal{F} \to \mathcal{I}^*$ is an injective resolution. The cohomology of this double complex can be computed by means of spectral sequences. Below, we will adapt Leray's Theorem to an equivariant setting and provide the details of the proof in that case. Presently, we note that, as we have seen above, if $\mathcal{F}$ is a constant sheaf, then the Čech cohomology of a contractible open set U vanishes in positive degree by homotopy invariance for constant sheaves. So, if we are dealing with constant sheaves and we can find an open cover $\mathcal{U}$ such that all finite intersections $U_{\alpha_0\ldots\alpha_k}$ are contractible, then there is no need to take the limit because the Čech cohomology of this cover already gives us the Čech cohomology of the space X. What is more, for some spaces one can already realize a "good" cover with only a few open sets. This can make the computation of Čech cohomology a much more tractable problem.

9.3 Equivariant Čech and sheaf cohomology

Our objective is now to generalize Čech cohomology and sheaf cohomology to an equivariant setting, or, in other words, to a setting where our topological spaces and the sheaves on them come equipped with a $\mathbb{Z}_2$ action.

First, let us generalize the notion of sheaf over a space X to that of equivariant sheaf over an involutive space (X,τ). For an ordinary sheaf $\mathcal{F}$ over (X,τ), we let $\tau^*\mathcal{F}$ denote the pullback sheaf, which is defined through $(\tau^*\mathcal{F})(U) = \mathcal{F}(\tau(U))$ for opens $U \subseteq X$. The stalk at $x \in X$ of the pullback sheaf is isomorphic to the stalk at $\tau(x)$, i.e. $(\tau^*\mathcal{F})_x \cong \mathcal{F}_{\tau(x)}$. An *equivariant sheaf* $(\mathcal{F},\theta)$ over an involutive space (X,τ) is a sheaf $\mathcal{F}$ over X together with a sheaf isomorphism $\theta : \mathcal{F} \to \tau^*\mathcal{F}$ such that the induced map on stalks $\theta_x : \mathcal{F}_x \to \mathcal{F}_{\tau(x)}$ satisfies $\theta_{\tau(x)} \circ \theta_x = \mathrm{id}_x$. For example, the sheaves $\mathbb{Z}_X$ and $\mathcal{O}_X^*$ can be easily upgraded to equivariant sheaves when they are defined over an involutive space (X,τ) by defining θ_x on stalks to be multiplication by -1 or complex conjugation, respectively. This is possible, because the stalks are canonically isomorphic to $\mathbb{Z}$ and $\mathbb{C}$. We will denote these two equivariant sheaves by $\widetilde{\mathbb{Z}}_X$ and $\widetilde{\mathcal{O}}_X^*$.

We would now like to define a notion of equivariant Čech cohomology over an open cover $\mathcal{U}$ with coefficients in an equivariant sheaf $\widetilde{\mathcal{F}} := (\mathcal{F},\theta)$. However, this is not going to work for just any open cover $\mathcal{U}$ of X. The cover itself also has to be equivariant in some way [44]. Let (X,τ) be an involutive space such that all fixed points of X are isolated. A cover $\mathcal{U} = \{U_\alpha\}_{\alpha\in\mathcal{A}}$ of (X,τ) is called a $\mathbb{Z}_2$*-cover* if there is an involution on the index set $\mathcal{A}$, that we shall denote by $\tau : \mathcal{A} \to \mathcal{A}$, such that

$$U_{\tau(\alpha)} = \tau(U_\alpha)\,.$$

We say the $\mathbb{Z}_2$-cover $\mathcal{U}$ is *without fixed points* if the involution $\tau : \mathcal{A} \to \mathcal{A}$ on the index set has no fixed points. From now on, we will assume our $\mathbb{Z}_2$-covers to be without fixed points.

Čech cohomology then generalizes as follows. Let $(\mathcal{F},\theta)$ be an equivariant sheaf and $\mathcal{U} = \{U_\alpha\}_{\alpha\in\mathcal{A}}$ a $\mathbb{Z}_2$-cover without fixed points. There is an action of $\mathbb{Z}_2$ on the Čech cochains $(f) \in \check{C}^n(\mathcal{U};\widetilde{\mathcal{F}})$ given by

$$\tau : \check{C}^n(\mathcal{U};\widetilde{\mathcal{F}}) \to \check{C}^n(\mathcal{U};\widetilde{\mathcal{F}})\,, \qquad (\tau(f))_{\alpha_0\ldots\alpha_n} = \theta(f_{\tau(\alpha_0)\ldots\tau(\alpha_n)})\,.$$

The cochains that are invariant under this $\mathbb{Z}_2$ action are called the $\mathbb{Z}_2$*-invariant cochains*, which we denote by $\check{C}^n(\mathcal{U};\widetilde{\mathcal{F}})^{\mathbb{Z}_2}$. The equivariant Čech cohomology $\check{H}^*(\mathcal{U};\mathbb{Z}_2,\widetilde{\mathcal{F}})$ of the

$\mathbb{Z}_2$-cover $\mathcal{U}$ is defined to be the cohomology of the cochain complex of $\mathbb{Z}_2$-invariant cochains $\check{C}^*(\mathcal{U};\widetilde{\mathcal{F}})^{\mathbb{Z}_2}$. By taking the direct limit over refinements of $\mathbb{Z}_2$-covers we obtain the *equivariant Čech cohomology* $\check{H}^*(X;\mathbb{Z}_2,\widetilde{\mathcal{F}})$.

Sheaf cohomology also has an equivariant version. Recall that we defined sheaf cohomology as the cohomology of the global sections of an injective resolution. The key ingredient for this definition was the fact that every sheaf of abelian groups can be embedded injectively in an injective sheaf. This result remains true in the equivariant setting as proven by Grothendieck in [45, Proposition 5.1.2]. The definition of equivariant sheaf cohomology now proceeds exactly like the definition of sheaf cohomology, but now all sheaves and maps involved are equivariant. So, if $\widetilde{\mathcal{F}} := (\mathcal{F},\theta)$ is an equivariant sheaf over an involutive space (X,τ), then the equivariant sheaf cohomology $H^*(X;\mathbb{Z}_2,\widetilde{\mathcal{F}})$ is defined to be the cohomology of

$$\widetilde{\mathcal{I}}^0(X) \longrightarrow \widetilde{\mathcal{I}}^1(X) \longrightarrow \widetilde{\mathcal{I}}^2(X) \longrightarrow \dots \ ,$$

where $0 \to \widetilde{\mathcal{F}} \to \widetilde{\mathcal{I}}^*$ is an injective resolution of equivariant sheaves. Again, by viewing the equivariant sheaf cohomology as a right derived functor, we obtain its independence on the choice of resolution. The proof that establishes the equivalence of sheaf and Čech cohomology can be extended to the equivariant world. The technicalities are deferred to Appendix B.

Computing equivariant sheaf cohomology and Čech cohomology suffers at least from the same difficulties as ordinary sheaf and Čech cohomology. Namely, how do we get our hands on a resolution? And, can we compute the direct limit over refinements of $\mathbb{Z}_2$-covers? In the ordinary setting the Leray theorem got rid of the problem of the direct limit over refinements. The Leray theorem states that the Čech cohomology with respect to a "good" cover is isomorphic to the Čech cohomology. An equivariant generalization of the Leray theorem is possible. Suppose we are given an equivariant sheaf $\widetilde{\mathcal{F}}$. By a "good" $\mathbb{Z}_2$-cover we mean a $\mathbb{Z}_2$-cover $\mathcal{U} = \{U_\alpha\}_{\alpha\in\mathcal{A}}$ without fixed fixed points such that $H^n(U_{\alpha_0\dots\alpha_k};\mathcal{F}) = 0$ for all $k \geq 0$ and $n \geq 1$. Note that this condition is a condition in ordinary sheaf cohomology, just like in the original Leray theorem. The following generalizes the Leray theorem to the equivariant setting.

Theorem 9.4. Let (X,τ) be an involutive space and $\widetilde{\mathcal{F}} = (\mathcal{F},\theta)$ a $\mathbb{Z}_2$-sheaf over (X,τ) and $\mathcal{U}$ a "good" $\mathbb{Z}_2$-cover. Then, $\check{H}^n(\mathcal{U};\mathbb{Z}_2,\widetilde{\mathcal{F}}) \cong H^n(X;\mathbb{Z}_2,\widetilde{\mathcal{F}})$ for all $n \geq 0$.

The proof uses techniques involving double complexes and spectral sequences of double complexes covered for example in [46].

Proof. Let $0 \to \widetilde{\mathcal{F}} \to \widetilde{\mathcal{I}}^*$ be an injective resolution of $\mathbb{Z}_2$-sheaves. Define the double complex

$$C^{p,q} := C^p(\mathcal{U};\widetilde{\mathcal{I}}^q)^{\mathbb{Z}_2} = \Big(\prod_{\alpha_0,\dots,\alpha_p} \Gamma(U_{\alpha_0\dots\alpha_p},\widetilde{\mathcal{I}}^q)\Big)^{\mathbb{Z}_2}$$

with differentials

$$\delta : C^{p,q} \to C^{p+1,q} \quad \text{and} \quad d : C^{p,q} \to C^{p,q+1}\,,$$

where δ is the Čech differential and d is the differential induced by the maps of the resolution. The total differential of the double complex is $D = \delta + (-1)^p d$, which is a map

$$D : \bigoplus_{p+q=n} C^{p,q} \to \bigoplus_{p+q=n+1} C^{p,q}\,.$$

It can be checked that $D^2 = 0$, so that $\left(\bigoplus_{p+q=n} C^{p,q}, D\right)$ is a cochain complex. The cohomology of this cochain complex is, by definition, the cohomology of the double complex. We can compute the cohomology of the double complex by means of a spectral sequence in two ways, namely taking "horizontal filtrations" or "vertical filtrations".

Let us first go horizontally by computing the δ-cohomology for fixed q. This yields the Čech cohomology of the cover $\mathcal{U}$ with coefficients in the $\mathbb{Z}_2$-sheaf $\widetilde{\mathcal{I}}^q$. The first page of the spectral sequence when using the horizontal filtration is thus

$$E_{h,1}^{p,q} = \check{H}^p(\mathcal{U}; \mathbb{Z}_2, \widetilde{\mathcal{I}}^q)\,.$$

Since $\widetilde{\mathcal{I}}^q$ is an injective $\mathbb{Z}_2$-sheaf, its $\mathbb{Z}_2$-Čech cohomology vanishes for any $\mathbb{Z}_2$-cover in positive degrees, see Proposition B.5. Hence, $E_{h,1}^{p,q} = 0$ for $p > 0$. On the other hand, for $p = 0$, we have the zeroth degree $\mathbb{Z}_2$-Čech cohomology, which are just the global $\mathbb{Z}_2$-invariant sections,

$$E_{h,1}^{0,q} = \check{H}^0(\mathcal{U}; \mathbb{Z}_2, \widetilde{\mathcal{I}}^q) = \Gamma^{\mathbb{Z}_2}(X, \widetilde{\mathcal{I}}^q)\,.$$

By the theory of spectral sequences, it follows that the cohomology of the double complex is equal to the cohomology of $(\Gamma^{\mathbb{Z}_2}(X, \widetilde{\mathcal{I}}^*), d)$, which is $H^*(X; \mathbb{Z}_2, \widetilde{\mathcal{F}})$ by definition of the $\mathbb{Z}_2$-sheaf cohomology.

We now compute the double cohomology again by first going vertically. We claim that computing the d-cohomology for fixed $p \geq 0$ yields

$$E_{v,1}^{p,q} = \Big(\prod_{\alpha_0 \ldots \alpha_p} H^q(U_{\alpha_0 \ldots \alpha_p}; \widetilde{\mathcal{F}}) \Big)^{\mathbb{Z}_2},$$

for all $q \geq 0$. The $\mathbb{Z}_2$-invariance requirement here means that a sequence of cocycles $(f_{\vec{\alpha}})$ representing a product of cohomology classes has to be $\mathbb{Z}_2$ invariant as a sequence of cochains. For $q = 0$,

$$E_{v,1}^{p,0} = \ker d : \Big(\prod_{\alpha_0 \ldots \alpha_p} \Gamma(U_{\alpha_0 \ldots \alpha_p}, \widetilde{\mathcal{I}}^0) \Big)^{\mathbb{Z}_2} \to \Big(\prod_{\alpha_0 \ldots \alpha_p} \Gamma(U_{\alpha_0 \ldots \alpha_p}, \widetilde{\mathcal{I}}^1) \Big)^{\mathbb{Z}_2}.$$

An invariant cochain $(f_{\vec{\alpha}})$ is in this kernel if and only if $df_{\vec{\alpha}} = 0$ for all $\vec{\alpha}$, so that by definition of the degree 0 sheaf cohomology

$$E_{v,1}^{p,0} \cong \Big(\prod_{\alpha_0 \ldots \alpha_p} H^0(U_{\alpha_0 \ldots \alpha_p}; \widetilde{\mathcal{F}}) \Big)^{\mathbb{Z}_2}.$$

For $q > 0$ the argument is more involved. We have by definition that

$$E_{v,1}^{p,q} = \frac{\ker d : \left(\prod_{\alpha_0 \ldots \alpha_p} \Gamma(U_{\alpha_0 \ldots \alpha_p}, \widetilde{\mathcal{I}}^q) \right)^{\mathbb{Z}_2} \to \left(\prod_{\alpha_0 \ldots \alpha_p} \Gamma(U_{\alpha_0 \ldots \alpha_p}, \widetilde{\mathcal{I}}^{q+1}) \right)^{\mathbb{Z}_2}}{\operatorname{im} d : \left(\prod_{\alpha_0 \ldots \alpha_p} \Gamma(U_{\alpha_0 \ldots \alpha_p}, \widetilde{\mathcal{I}}^{q-1}) \right)^{\mathbb{Z}_2} \to \left(\prod_{\alpha_0 \ldots \alpha_p} \Gamma(U_{\alpha_0 \ldots \alpha_p}, \widetilde{\mathcal{I}}^{q}) \right)^{\mathbb{Z}_2}}.$$

It is again the case that a cochain $(f_{\vec{\alpha}}) \in \left(\prod_{\alpha_0 \ldots \alpha_p} \Gamma(U_{\alpha_0 \ldots \alpha_p}, \widetilde{\mathcal{I}}^q) \right)^{\mathbb{Z}_2}$ is in the kernel of d if and only if $df_{\vec{\alpha}} = 0$ for all $\vec{\alpha}$. However, before we can conclude that each $f_{\vec{\alpha}}$ defines a cohomology class in $H^q(U_{\vec{\alpha}}; \widetilde{\mathcal{F}})$, it has to be checked that demanding $\mathbb{Z}_2$-invariance commutes with taking the quotient. In other words, we have to prove that two sequences of $\mathbb{Z}_2$-invariant cocycles $(f_{\vec{\alpha}})$ and $(f'_{\vec{\alpha}})$ of degree q represent the same elements in cohomology for every $\vec{\alpha}$ if and only if there is a $\mathbb{Z}_2$-*invariant* cochain $(g_{\vec{\alpha}})$ in degree $q-1$ such that

$f_{\vec{\alpha}} - f'_{\vec{\alpha}} = dg_{\vec{\alpha}}$ for all $\vec{\alpha}$. From right to left the statement is trivial, but from left to right it is not.

Assume $(f_{\vec{\alpha}})$ and $(f'_{\vec{\alpha}})$ are two $\mathbb{Z}_2$-invariant cochains in degree $q > 0$ that represent the same cohomology classes, meaning that for every $\vec{\alpha}$ we can find $g_{\vec{\alpha}}$ in degree $q-1$ such that $f_{\vec{\alpha}} - f'_{\vec{\alpha}} = dg_{\vec{\alpha}}$ for all $\vec{\alpha}$. The cochain formed by $(g_{\vec{\alpha}})$ may not be $\mathbb{Z}_2$ invariant. We will deform it such that it becomes so.

Since d is a $\mathbb{Z}_2$-map, it commutes with θ so that $d(\theta g_{\vec{\alpha}}) = \theta(dg_{\vec{\alpha}})$ for all $\vec{\alpha}$. The $\mathbb{Z}_2$-invariance of f and f' implies that

$$\theta(dg_{\vec{\alpha}}) = \theta(f_{\vec{\alpha}} - f'_{\vec{\alpha}}) = (\tau(f - f'))_{\tau\vec{\alpha}} = (f - f')_{\tau\vec{\alpha}} = dg_{\tau\vec{\alpha}}\,.$$

Therefore, $d(\theta g_{\vec{\alpha}} - g_{\tau\vec{\alpha}}) = 0$. The resolution $0 \to \mathcal{F} \to \mathcal{I}^*$ is an exact sequence of sheaves. By the assumption that $H^*(U_{\vec{\alpha}}; \mathcal{F}) = 0$ in positive degrees, it follows that the sequence

$$0 \to \mathcal{F}(U_{\vec{\alpha}}) \to \mathcal{I}^*(U_{\vec{\alpha}}) \tag{9.3}$$

is exact in all degrees. Note that exactness at $\mathcal{F}(U_{\vec{\alpha}})$ and $\mathcal{I}^0(U_{\vec{\alpha}})$ always holds due to left exactness of the section functor $\Gamma(U_{\vec{\alpha}}, -)$. The exactness in higher degrees is secured by the assumption. We can thus find a primitives for $\theta g_{\vec{\alpha}} - g_{\tau\vec{\alpha}}$ in this sequence. Let $h_{\vec{\alpha}}$ in degree $q-2$ of the sequences of the type (9.3) be such that

$$\theta g_{\vec{\alpha}} = g_{\tau\vec{\alpha}} + dh_{\tau\vec{\alpha}}\,. \tag{9.4}$$

If $q = 1$, then $q - 2 = -1$ and the $h_{\vec{\alpha}}$ is drawn from $\mathcal{F}(U_{\vec{\alpha}})$. Applying θ to (9.4) yields $g_{\vec{\alpha}} = \theta g_{\tau\vec{\alpha}} + \theta dh_{\tau\vec{\alpha}}$, since $\theta^2 = 1$. We obtain the relation

$$g_{\vec{\alpha}} - \theta dh_{\tau\alpha} = \theta g_{\tau\vec{\alpha}} = g_{\vec{\alpha}} + dh_{\vec{\alpha}}$$

which implies

$$dh_{\vec{\alpha}} = -\theta dh_{\tau\vec{\alpha}}\,. \tag{9.5}$$

Since we assume a cover without fixed points, we can partition the indices $\vec{\alpha}$ in two disjoint sets $\mathcal{A}_+$ and $\mathcal{A}_-$ such that $\tau\mathcal{A}_+ = \mathcal{A}_-$, for example by partitioning on the first index of $\vec{\alpha} = (\alpha_0 \dots \alpha_n)$. Define now the cochain (g') as follows,

$$g'_{\vec{\alpha}} = \begin{cases} g_{\vec{\alpha}} & \text{if } \vec{\alpha} \in \mathcal{A}_+ \\ g_{\vec{\alpha}} + dh_{\vec{\alpha}} & \text{if } \vec{\alpha} \in \mathcal{A}_- \end{cases}.$$

It is clear that $dg'_{\vec{\alpha}} = dg_{\vec{\alpha}} = f_{\vec{\alpha}} - f'_{\vec{\alpha}}$, since $d^2 = 0$. Moreover, (g') is $\mathbb{Z}_2$ invariant. Indeed, if $\vec{\alpha} \in \mathcal{A}_+$, then

$$\theta g'_{\vec{\alpha}} = \theta g_{\vec{\alpha}} = g_{\tau\vec{\alpha}} + dh_{\tau\vec{\alpha}} = g'_{\tau\vec{\alpha}}\,,$$

and if $\vec{\alpha} \in \mathcal{A}_-$, then

$$\theta g'_{\vec{\alpha}} = \theta(g_{\vec{\alpha}} + dh_{\vec{\alpha}}) = g_{\tau\vec{\alpha}} + dh_{\tau\vec{\alpha}} + \theta dh_{\vec{\alpha}} = g_{\tau\vec{\alpha}} = g'_{\tau\vec{\alpha}}\,,$$

where we have used (9.5) (after relabelling the indices). This proves that indeed, if two invariant cocycles are cohomologous, then they differ by the image under d of a $\mathbb{Z}_2$-invariant cochain, which proves the claim.

The assumption that $H^q(U_{\alpha_0 \dots \alpha_n}; \mathcal{F}) = 0$ for all $q > 0$ yields that $E^{p,q}_{v,1} = 0$ for $q > 0$. Computing the δ-cohomology of $E^{p,0}_{v,1}$ yields again the cohomology of the double complex, but this time in the guise of the $\mathbb{Z}_2$-Čech cohomology $\check{H}^*(\mathcal{U}; \mathbb{Z}_2, \widetilde{\mathcal{F}})$. We thus conclude that $\check{H}^*(\mathcal{U}; \mathbb{Z}_2, \widetilde{\mathcal{F}}) \cong H^*(X; \mathbb{Z}_2, \widetilde{\mathcal{F}})$ since both of these compute the cohomology of the same double complex. □

The assumption $H^n(U_{\alpha_0\dots\alpha_k};\mathcal{F})=0$ for all $k\geq 0$ and $n\geq 1$ in Theorem 9.4 is in general not straightforward to verify. However, if we consider coefficients in the equivariant sheaf $\widetilde{\mathbb{Z}}_X$, then the assumption is true for $\mathbb{Z}_2$-covers all of whose intersections are contractible to a point. Recall that $\widetilde{\mathbb{Z}}_X$ is the sheaf $\mathbb{Z}_X$ endowed with the $\mathbb{Z}_2$ action θ that acts on stalks via multiplication by -1. Forgetting the $\mathbb{Z}_2$ structure, the sheaf $\widetilde{\mathbb{Z}}_X$ is just the constant sheaf $\mathbb{Z}_X$, and, for constant sheaves, the sheaf cohomology is homotopy invariant. So, Theorem 9.4 says that when we want the compute the $\mathbb{Z}_2$-Čech cohomology of some involutive space (X,τ) with coefficients in $\widetilde{\mathbb{Z}}_X$, then it suffices to compute the $\mathbb{Z}_2$-Čech cohomology on a cover all of whose finite intersections are contractible.

We can thus compute the cohomology with coefficients in $\widetilde{\mathbb{Z}}_X$ for some simple spaces X. In the following, let us drop the space X from the notation of the sheaf $\widetilde{\mathbb{Z}}_X$ and just denote this sheaf by $\widetilde{\mathbb{Z}}$ with the space under consideration being understood. The simplest space is of course just a point with the trivial involution $\tau:\{*\}\to\{*\}$. A $\mathbb{Z}_2$-cover without fixed points and all of whose finite intersections are contractible is given by $\mathcal{U}=\{*_+,*_-\}$, i.e. two copies of the point indexed by $+$ and $-$. The involution on the index set $\{+,-\}$ acts as $\tau(+)=-$. The $\mathbb{Z}_2$ action on a cochain (a) is

$$(\tau(a))_+=\theta(a_-)=-a_- \quad\text{and}\quad (\tau(a))_-=\theta(a_+)=-a_+\,.$$

The degree zero $\mathbb{Z}_2$-invariant cochains are

$$C^0(\mathcal{U};\widetilde{\mathbb{Z}})^{\mathbb{Z}_2}=\{(a_+,a_-)\in\mathbb{Z}\oplus\mathbb{Z}:\,a_-=-a_+\}\cong\mathbb{Z}\,.$$

The degree one $\mathbb{Z}_2$-invariant cochains are

$$C^1(\mathcal{U};\widetilde{\mathbb{Z}})^{\mathbb{Z}_2}=\{b_{+-}\in\mathbb{Z}\}\cong\mathbb{Z}\,.$$

Note that by skew symmetry in the indices of the Čech cochain, it suffices to specify only b_{+-} in the above. The skew symmetry constraint and the $\mathbb{Z}_2$-invariance are compatible, since both require $b_{-+}=-b_{+-}$. There are no triple overlaps, so all higher cochain groups are zero. The equivariant Čech cochain complex for the cover $\{*_+,*_-\}$ of the point is thus

$$0\longrightarrow\mathbb{Z}\xrightarrow{\delta}\mathbb{Z}\longrightarrow 0\ ,$$

where if $(a)=(a_+,a_-)\in C^0(\mathcal{U};\widetilde{\mathbb{Z}})^{\mathbb{Z}_2}$, then $(\delta a)_{+-}=a_--a_+=2a_-$. So, $\ker\delta=0$ and $\operatorname{im}\delta=2\mathbb{Z}$. The cover $\mathcal{U}$ and the sheaf $\widetilde{\mathbb{Z}}$ satisfy the assumptions of Theorem 9.4. The cohomology of the point is thus

$$H^n(*;\mathbb{Z}_2,\widetilde{\mathbb{Z}})\cong H^n(\mathcal{U};\mathbb{Z}_2,\widetilde{\mathbb{Z}})=\begin{cases}\mathbb{Z}_2 & \text{if } n=1\\ 0 & \text{if } n\neq 1\end{cases}\,.$$

A peculiarity of the equivariant cohomology of the point with $\widetilde{\mathbb{Z}}$ coefficients is that the first cohomology of the point is not zero, but it is $\mathbb{Z}_2$.

We can also compute the $\mathbb{Z}_2$-cohomology with coefficients in $\widetilde{\mathbb{Z}}$ for the time reversal circle. Recall that the time reversal circle $\widetilde{S}^1$ is the circle S^1 with involution $\tau(x,y)=(-x,y)$. A $\mathbb{Z}_2$-cover $\mathcal{U}$ of $\widetilde{S}^1$ without fixed points and with all finite intersections contractible is shown in Fig. 12. The index set of $\mathcal{U}$ is $\mathcal{A}=\{N_+,N_-,S_+,S_-,x_+,x_-\}$. The open set corresponding to $N_\pm$ is the northern hemisphere, to $S_\pm$ the southern hemisphere, to x_+ the right hemisphere and to x_- the left hemisphere. The fixed points are the north pole and the south pole, which is why we need two copies of the northern hemisphere and two copies of the southern hemisphere. We have single, double and triple overlaps, all of which

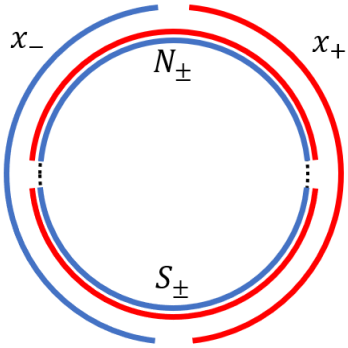

Figure 12: Schematic representation of the "good" cover of the time reversal circle with the six opens labelled by $\mathcal{A} = \{N_+, N_-, S_+, S_-, x_+, x_-\}$.

are contractible. The involution on the index set sends a subindex $+$ to a subindex $-$, e.g. $\tau(N_+) = N_-$. The degree zero cochains are

$$C^0(\mathcal{U};\widetilde{\mathbb{Z}})^{\mathbb{Z}_2} = \{(a_{N_+}, a_{N_-}, a_{S_+}, a_{S_-}, a_{x_+}, a_{x_-}) \in \mathbb{Z}^6 : a_{i_+} = -a_{i_-} \text{ for } i = N, S, x\} \cong \mathbb{Z}^3 .$$

For the higher degree cochains, we note that to specify a cochain in $C^p(\mathcal{U};\widetilde{\mathbb{Z}})^{\mathbb{Z}_2}$, we only need to specify an integer for each element of a "generating set of overlaps", that is, for a set of overlaps such that taking permutations of the indices or τ on all of the indices yields all possible combinations. For the double overlap, for example, we know that $b_{N_-x_-} = -b_{N_+x_+}$, so we only need to specify an integer for one of these two. Doing so, the degree one and two cochains are

$$\begin{aligned} C^1(\mathcal{U};\widetilde{\mathbb{Z}})^{\mathbb{Z}_2} &\cong \{(b_{N_+N_-}, b_{S_+S_-}, b_{N_+x_+}, b_{N_+x_-}, b_{S_+x_+}, b_{S_+x_-}) \in \mathbb{Z}^6\} , \\ C^2(\mathcal{U};\widetilde{\mathbb{Z}})^{\mathbb{Z}_2} &\cong \{(c_{N_+N_-x_+}, c_{S_+S_-x_+}) \in \mathbb{Z}^2\} . \end{aligned}$$

All higher order cochains are zero because we do not have higher order overlaps. The cochain complex for $\widetilde{S}^1$ thus looks like

$$0 \longrightarrow \mathbb{Z}^3 \xrightarrow{\delta^0} \mathbb{Z}^6 \xrightarrow{\delta^1} \mathbb{Z}^2 \longrightarrow 0 \ .$$

For $a = (a_{N_+}, a_{S_+}, a_{x_+}) \in C^0(\mathcal{U};\widetilde{\mathbb{Z}})^{\mathbb{Z}_2}$, we have that $\delta^0 a$ is given by

$$\begin{aligned} (\delta^0 a)_{N_+N_-} &= -2a_{N_+} & (\delta^0 a)_{N_+x_+} &= a_{x_+} - a_{N_+} & (\delta^0 a)_{N_+x_-} &= -a_{x_+} - a_{N_+} \\ (\delta^0 a)_{S_+S_-} &= -2a_{S_+} & (\delta^0 a)_{S_+x_+} &= a_{x_+} - a_{S_+} & (\delta^0 a)_{S_+x_-} &= -a_{x_+} - a_{S_+} , \end{aligned}$$

and for $b = (b_{N_+N_-}, b_{S_+S_-}, b_{N_+x_+}, b_{N_+x_-}, b_{S_+x_+}, b_{S_+x_-}) \in C^1(\mathcal{U};\widetilde{\mathbb{Z}})$, we have that $\delta^1 b$ is given by

$$\begin{aligned} (\delta^1 b)_{N_+N_-x_+} &= -b_{N_+x_-} - b_{N_+x_+} + b_{N_+N_-} \\ (\delta^1 b)_{S_+S_-x_+} &= -b_{S_+x_-} - b_{S_+x_+} + b_{S_+S_-} . \end{aligned}$$

It is clear that $\delta^1 : \mathbb{Z}^6 \to \mathbb{Z}^2$ is surjective, so the second cohomology vanishes. It is also straightforward to see that $\ker \delta^0 = 0$. Indeed, $\delta^0 a = 0$ if and only if all its components vanish. The vanishing of the N_+N_- component requires $a_{N_+} = 0$ and the S_+S_- component

requires $a_{S^+} = 0$. The fact that $a_{x_+} = 0$ then follows for example from the vanishing of the N_+x_+ component. The zeroth cohomology thus also vanishes.

To deduce the first cohomology we really have to work out $\ker \delta^1 / \operatorname{im} \delta^0$. A cochain $(b) \in C^1(\mathcal{U}; \widetilde{\mathbb{Z}})^{\mathbb{Z}_2}$ is a cocycle if and only if $\delta^1 b = 0$, which is true if and only if $b_{N_+x_-} = b_{N_+N_-} - b_{N_+x_+}$ and $b_{S_+x_-} = b_{S_+S_-} - b_{S_+x_+}$. We can use these relations to eliminate $b_{N_+x_-}$ and $b_{S_+x_-}$. Therefore,

$$\ker \delta^1 \cong \{(b_{N_+N_-}, b_{S_+S_-}, b_{N_+x_+}, b_{S_+x_+}) \in \mathbb{Z}^4\}\,.$$

The image of δ^0 contains two independent copies of $2\mathbb{Z}$ in the N_+N_- and S_+S_- components which come from the a_{N_+} and the a_{S_+} degrees of freedom in degree zero. Quotienting by a_{N_+} and a_{S_+} reduces $\mathbb{Z}^4$ to $(\mathbb{Z}_2)^2 \oplus \mathbb{Z}^2$. The a_{x_+} degree of freedom appears diagonally in the $\mathbb{Z}^2$ corresponding to the N_+x_+ and S_+x_+ components and so quotienting by it will reduce $\mathbb{Z}^2$ to $\mathbb{Z}$. We thus conclude that

$$H^n(\widetilde{S}^1; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \cong \begin{cases} (\mathbb{Z}_2)^2 \oplus \mathbb{Z} & \text{if } n = 1 \\ 0 & \text{if } n \neq 1 \end{cases}\,. \tag{9.6}$$

9.4 Classification of "Real" line bundles

Similar to how complex line bundles are classified by a Čech cohomology group, "Real" line bundles are classified by an equivariant Čech cohomology group. A priori, the classification is in terms of Čech cohomology with coefficients in $\widetilde{\mathcal{O}}_X$, the $\mathbb{Z}_2$-sheaf of continuous complex valued functions endowed with the $\mathbb{Z}_2$-action of complex conjugation. However, using the so-called exponential sequence, the classification can be reformulated in terms of Čech cohomology with values in $\widetilde{\mathbb{Z}}_X$. In Appendix C, the relevant Čech cohomology groups for time reversal spheres are computed. It is found that "Real" line bundles over time reversal spheres are always trivial, therefore justifying the construction of the FKMM invariant for "Quaternionic" vector bundles over time reversal spheres.

The classification of "Real" line bundles over an involutive space (X, τ) is given by the following proposition, which is adapted from [44, Proposition 1.1.1].

Proposition 9.5. Let (X, τ) be an involutive space. There is a canonical isomorphism

$$\mathrm{Vect}^1_R(X, \tau) \cong \check{H}^1(X; \mathbb{Z}_2, \widetilde{\mathcal{O}}^*_X)\,,$$

where $\check{H}^1(X; \mathbb{Z}_2, \widetilde{\mathcal{O}}^*_X)$ is the $\mathbb{Z}_2$-Čech cohomology.

Proof. Given a cohomology class $c \in \check{H}^1(X; \mathbb{Z}_2, \widetilde{\mathcal{O}}^*_X)$ we construct from it a complex line bundle $L(c)$ in exactly the same fashion as in the non-equivariant case treated in Proposition 9.1. The complex line bundle $L(c)$ constructed there now has to be endowed with a "Real" structure. Suppose $(g_{\alpha\beta}) \in \check{H}^1(\mathcal{U}; \mathbb{Z}_2, \widetilde{\mathcal{O}}^*_X)$ represents the class c for some $\mathbb{Z}_2$-cover $\mathcal{U} = \{U_\alpha\}_{\alpha \in \mathcal{A}}$. Recall that $L(c) \cong \widetilde{L}(g)/\sim$, where $\widetilde{L}(g) = \bigsqcup_{\alpha \in \mathcal{A}} U_\alpha \times \mathbb{C}$ and

$$U_\alpha \times \mathbb{C} \ni (x, z) \sim (y, w) \in U_\beta \times \mathbb{C} \iff (y, w) = (x, g_{\alpha\beta}(x) z)\,.$$

Define the "Real" structure $\widetilde{\Theta} : \widetilde{L}(g) \to \widetilde{L}(g)$ to be $\widetilde{\Theta}(x, z) = (\tau(x), \bar{z})$. This "Real" structure induces a "Real" structure on the quotient by $\sim$, precisely because $(g_{\alpha\beta})$ is a $\mathbb{Z}_2$-invariant cocycle. To see this, suppose $U_\alpha \times \mathbb{C} \ni (x, z) \sim (y, w) \in U_\beta \times \mathbb{C}$, then we have to show that $U_{\tau\alpha} \times \mathbb{C} \ni (\tau(x), \bar{z}) \sim (\tau(y), \bar{w}) \in U_{\tau\beta} \times \mathbb{C}$. In other words, we have to prove that $\bar{w} = g_{\tau\alpha\tau\beta}(\tau(x))\bar{z}$. We know that $w = g_{\alpha\beta}(x) z$. Taking the complex conjugate yields $\bar{w} = \overline{g_{\alpha\beta}(x)}\bar{z}$. The $\mathbb{Z}_2$-invariance of (g) means that $\theta_x(g_{\alpha\beta}(x)) = g_{\tau\alpha\tau\beta}(\tau(x))$. The

action of θ_x on the stalks of $\mathcal{O}_X^*$ is complex conjugation. Hence, the $\mathbb{Z}_2$-invariance gives us

$$\overline{g_{\alpha\beta}(x)} = \theta_x(g_{\alpha\beta}(x)) = g_{\tau\alpha\tau\beta}(\tau(x))\,,$$

so that, indeed, $\bar{w} = \overline{g_{\alpha\beta}(x)}\bar{z} = g_{\tau\alpha\tau\beta}(\tau(x))\bar{z}$ proving that $(\tau(x),\bar{z}) \sim (\tau(y),\bar{w})$. The map $\widetilde{\Theta}$ thus induces a "Real" structure on $L(c)$.

By the same arguments as before, the "Real" line bundles that we obtain for different choices of representatives $(g_{\alpha\beta})$ of c are isomorphic as "Real" line bundles. Indeed, different representatives are related by a $\mathbb{Z}_2$-invariant boundary, which gives rise to an equivariant isomorphism. The proofs of injectivity and surjectivity also proceed in the same way, where surjectivity now rests on the fact that a "Real" line bundle always has a "Real" trivializing $\mathbb{Z}_2$-cover. □

To show that "Real" line bundles over some involutive space (X,τ), and thus the determinant bundles associated to "Quaternionic" vector bundles of even rank, are trivial then amounts to proving that $\check{H}^1(X;\mathbb{Z}_2,\widetilde{\mathcal{O}}_X^*) = 0$. Computing Čech cohomology for the sheaf $\mathcal{O}_X^*$ is rather impractical, because the spaces of sections are infinite-dimensional function spaces. We would rather deal with finite-dimensional spaces. The trick is to consider the short exact sequence of $\mathbb{Z}_2$-sheaves

$$0 \longrightarrow \widetilde{\mathbb{Z}} \xrightarrow{2\pi i(\cdot)} \widetilde{\mathcal{O}}_X \xrightarrow{\exp(\cdot)} \widetilde{\mathcal{O}}_X^* \longrightarrow 0\,. \tag{9.7}$$

The short exact sequence Eq. (9.7) is called the *exponential sequence*. It is important to note that the sheaf $\mathcal{O}_X$, with group operation of addition of functions, is made into an equivariant sheaf $\widetilde{\mathcal{O}}_X$ by endowing it with the $\mathbb{Z}_2$ action of multiplication by -1, whilst the sheaf $\mathcal{O}_X^*$, with group operation of multiplication of functions, is made into an equivariant sheaf $\widetilde{\mathcal{O}}_X^*$ by endowing it with the $\mathbb{Z}_2$ action of complex conjugation. The exponential sequence gives rise to a long exact sequence in Čech cohomology that contains the following portion

$$\check{H}^1(X;\mathbb{Z}_2,\widetilde{\mathcal{O}}_X) \longrightarrow \check{H}^1(X;\mathbb{Z}_2,\widetilde{\mathcal{O}}_X^*) \longrightarrow \check{H}^2(X;\mathbb{Z}_2,\widetilde{\mathbb{Z}}) \longrightarrow \check{H}^2(X;\mathbb{Z}_2,\widetilde{\mathcal{O}}_X)\,. \tag{9.8}$$

The ordinary sheaf $\mathcal{O}_X$ is *fine*, meaning that it has partitions of unit. Therefore, its ordinary sheaf cohomology $H^*(X;\mathcal{O}_X)$ in positive degree vanishes for all spaces X. A "good" cover of (X,τ) is then realized by taking a cover $\mathcal{U} = \{X_+, X_-\}$ consisting of two copies X. Now we can use Theorem 9.4 to compute $\check{H}^*(X;\mathbb{Z}_2,\widetilde{\mathcal{O}}_X)$ by computing $\check{H}^*(X;\mathbb{Z}_2,\widetilde{\mathcal{O}}_X)$ for the cover $\mathcal{U} = \{X_+, X_-\}$. The cochain groups in degree zero and one are

$$\begin{aligned}
\check{C}^0(\mathcal{U},\widetilde{\mathcal{O}}_X)^{\mathbb{Z}_2} &= \{(f_+,f_-) :\, f_+(x) = -f_-(\tau(x))\}\\
\check{C}^1(\mathcal{U},\widetilde{\mathcal{O}}_X)^{\mathbb{Z}_2} &= \{(g_{+-},g_{-+}) :\, g_{+-}(x) = -g_{-+}(\tau(x)) = g_{+-}(\tau(x))\}\,,
\end{aligned}$$

where for the degree one cochains we have used both the equivariance as well as the usual skew-symmetry of Čech cohomology. Since there are no higher order overlaps, the cochain complex is

$$0 \longrightarrow \check{C}^0(\mathcal{U},\widetilde{\mathcal{O}}_X)^{\mathbb{Z}_2} \xrightarrow{\delta} \check{C}^1(\mathcal{U},\widetilde{\mathcal{O}}_X)^{\mathbb{Z}_2} \longrightarrow 0\ ,$$

where $\delta(f_+,f_-)_{+-}(x) = f_-(x) - f_+(x) = f_-(x) + f_-(\tau(x))$. The degree zero cohomology consists of cochains (f_+,f_-) in the kernel of δ satisfying $f_-(x) = -f_-(\tau(x))$, and can thus

be identified with equivariant sections of $\widetilde{\mathcal{O}}_X$, as expected. Since there are no cochains in degrees $n \geq 2$, it follows immediately that $\check{H}^n(X;\mathbb{Z}_2,\widetilde{\mathcal{O}}_X) = 0$ for $n \geq 2$. Finally, the degree one cohomology $\check{H}^1(X;\mathbb{Z}_2,\widetilde{\mathcal{O}}_X)$ also vanishes. To see this, note that $\operatorname{im}\delta$ and $\check{C}^1(\mathcal{U},\widetilde{\mathcal{O}}_X)^{\mathbb{Z}_2}$ are both given by the functions $h : X \to \mathbb{C}$ such that $h(x) = h(\tau(x))$.

By the vanishing of $\check{H}^n(X;\mathbb{Z}_2,\widetilde{\mathcal{O}}_X)$ for $n \geq 1$ and the exactness of the sequence in Eq. (9.8), it follows then that $\check{H}^1(X;\mathbb{Z}_2,\mathcal{O}_X^*) \cong \check{H}^2(X;\mathbb{Z}_2,\mathbb{Z})$. This allows us to restate the classification of "Real" line bundles in the following Corollary.

Corollary 9.6. Let (X,τ) be an involutive space. There is an isomorphism

$$\mathrm{Vect}^1_R(X,\tau) \cong \check{H}^2(X;\mathbb{Z}_2,\widetilde{\mathbb{Z}})\,.$$

We have already seen that $\check{H}^2(\widetilde{S}^1;\mathbb{Z}_2,\widetilde{\mathbb{Z}}) = 0$ in Eq. (9.6). In Appendix C, we prove that also $\check{H}^2(\widetilde{S}^d;\mathbb{Z}_2,\widetilde{\mathbb{Z}}) = 0$ for time reversal spheres with $d \geq 2$. These results show that all "Real" line bundles over the time reversal spheres in $d \geq 1$ are trivial and this justifies the construction of the FKMM invariant over time reversal spheres.

For the time reversal tori, however, the combinatorics of the Čech cohomology, even after restricting to the smallest possible "good" cover, remain rather unwieldy. Therefore, to justify the construction of the FKMM invariant for Q-bundles over the time reversal torus we must still rely on the more advanced cohomological tools of [18].

9.5 Cohomology formulation of the FKMM invariant

The FKMM invariant has been constructed as taking values in a quotient of equivariant homotopy classes of maps. These equivariant homotopy classes are actually isomorphic to equivariant cohomology groups. This allows for a reformulation of the FKMM invariant in terms of equivariant cohomology. For time reversal spheres, the relevant cohomology group is computed in Appendix C and gives another proof of the $\mathbb{Z}_2$ valuedness of the FKMM invariant.

The isomorphism between equivariant homotopy and equivariant cohomology is due to the long exact sequence associated to the exponential sequence.

Proposition 9.7. Let (X,τ) be an involutive space. There is an isomorphism

$$[X,\mathbb{C}^*]_{\mathbb{Z}_2} \cong \check{H}^1(X;\mathbb{Z}_2,\widetilde{\mathbb{Z}})\,.$$

Proof. The Čech cohomology of $\widetilde{\mathcal{O}}_X$ vanishes in positive degree so the long exact sequence coming from the exponential sequence contains the portion

$$\check{H}^0(X;\mathbb{Z}_2,\widetilde{\mathcal{O}}_X) \xrightarrow{\exp} \check{H}^0(X;\mathbb{Z}_2,\widetilde{\mathcal{O}}_X^*) \longrightarrow \check{H}^1(X;\mathbb{Z}_2,\widetilde{\mathbb{Z}}) \longrightarrow 0\,.$$

The degree zero Čech cohomology is just the equivariant sections. Then, by exactness,

$$\check{H}^1(X;\mathbb{Z}_2,\widetilde{\mathbb{Z}}) \cong \Gamma(X,\widetilde{\mathcal{O}}_X^*)^{\mathbb{Z}_2}/\exp\Gamma(X,\widetilde{\mathcal{O}}_X)^{\mathbb{Z}_2}\,,$$

where the action of $\exp\Gamma(X,\widetilde{\mathcal{O}}_X)^{\mathbb{Z}_2}$ on $\Gamma(X,\widetilde{\mathcal{O}}_X^*)^{\mathbb{Z}_2}$ is by multiplication. We thus have to show that

$$\Gamma(X,\widetilde{\mathcal{O}}_X^*)^{\mathbb{Z}_2}/\exp\Gamma(X,\widetilde{\mathcal{O}}_X)^{\mathbb{Z}_2} \cong [X,\mathbb{C}^*]_{\mathbb{Z}_2}\,.$$

In other words, we have to show that two equivariant maps $f,g : X \to \mathbb{C}^*$ are $\mathbb{Z}_2$-homotopic if and only there is an equivariant $h : X \to \mathbb{C}$ such that $g = f\exp(h)$.

Let us first consider the implication from right to left. Assume we are given an $h : X \to \mathbb{C}$ such that $g = f\exp(h)$. We decompose the map h into its real and imaginary

parts $h = h_R + ih_I$. Note that h is equivariant with respect to multiplication by -1, because we view $\mathbb{C}$ as an additive group. Therefore,

$$h(\tau(x)) = -h(x)\,.$$

On the other hand, we view $\mathbb{C}^*$ as a multiplicative group with $\mathbb{Z}_2$ action given by complex conjugation, so that $||f(x)|| = ||f(\tau(x))||$ and $||g(x)|| = ||g(\tau(x))||$. The relation $g = f\exp(h)$ then implies that

$$\exp h_R(x) = \frac{||g(x)||}{||f(x)||} = \frac{||g(\tau(x))||}{||f(\tau(x))||} = \exp h_R(\tau(x))\,.$$

The real part of h thus has to be both symmetric and antisymmetric under τ, which can only be true if $h_R \equiv 0$. An equivariant homotopy from f to g is then readily constructed by defining

$$H : X \times [0,1] \to \mathbb{C}^*\,, \quad H(x,t) = f(x)\exp(ith_I(x))\,.$$

Conversely, if $H : X \times [0,1] \to \mathbb{C}^*$ is an equivariant homotopy from f to g, then H/f is a homotopy from 1 to g/f. So, g/f is null-homotopic. Therefore, $h := \log(g/f)$ is well-defined and satisfies $g = f\exp(h)$. We have thus established the chain of bijections

$$\check{H}^1(X;\mathbb{Z}_2,\widetilde{\mathbb{Z}}) \cong \Gamma(X,\widetilde{\mathcal{O}}_X^*)^{\mathbb{Z}_2}/\exp\Gamma(X,\widetilde{\mathcal{O}}_X)^{\mathbb{Z}_2} \cong [X,\mathbb{C}^*]_{\mathbb{Z}_2}\,,$$

which proves the result. □

The FKMM invariant takes values in the quotient $[X^\tau,\mathbb{C}^*]_{\mathbb{Z}_2}/[X,\mathbb{C}^*]_{\mathbb{Z}_2}$. The previous Proposition relates the equivariant homotopy groups to the equivariant cohomology groups. In principle, we can thus interpret the FKMM invariant in terms of a quotient of cohomology groups. However, the final form in which we would like to present the cohomological formulation gets rid of the quotient in favor of so-called relative cohomology.

In the non-equivariant context, if $\mathcal{F}$ is a sheaf over X and $Y \subset X$ is subspace, then we can define the relative Čech cohomology $\check{H}^*(X,Y;\mathcal{F})$ of X relative to Y as follows. For a cover $\mathcal{U} = \{U_\alpha\}_{\alpha\in\mathcal{A}}$ of X, let $\mathcal{U}' = \{U_\alpha\}_{\alpha\in\mathcal{A}'}$ be all those sets that have nonempty intersection with Y. Let $\iota : \mathcal{A}' \to \mathcal{A}$ be the inclusion. The map ι induces a map in cochains. The relative Čech cochains $\check{C}^*(\mathcal{U},\mathcal{U}';\mathcal{F})$ are the cokernels of $\iota : \check{C}^*(\mathcal{U}';\mathcal{F}) \to \check{C}^*(\mathcal{U};\mathcal{F})$. The relative Čech cohomology with respect to the cover $\mathcal{U}$, denoted $\check{H}^*(\mathcal{U},\mathcal{U}';\mathcal{F})$, is the cohomology of the relative Čech cochains. The *relative Čech cohomology* $\check{H}^*(X,Y;\mathcal{F})$ of the pair (X,Y) is the direct limit over covers $\mathcal{U}$ of X of the $\check{H}^*(\mathcal{U},\mathcal{U}';\mathcal{F})$.

By definition of the relative Čech cochains, there is a level wise short exact sequence of cochain complexes

$$0 \longrightarrow C^*(\mathcal{U}',\mathcal{F}) \longrightarrow C^*(\mathcal{U};\mathcal{F}) \longrightarrow C^*(\mathcal{U},\mathcal{U}';\mathcal{F}) \longrightarrow 0\ .$$

This short exact sequence induces a long exact sequence by the usual construction,

$$\ldots \longrightarrow \check{H}^n(\mathcal{U};\mathcal{F}) \longrightarrow \check{H}^n(\mathcal{U}';\mathcal{F}) \longrightarrow \check{H}^{n+1}(\mathcal{U},\mathcal{U}';\mathcal{F}) \longrightarrow \check{H}^{n+1}(\mathcal{U};\mathcal{F}) \longrightarrow \ldots\ .$$

Taking the direct limit preserves exactness. The long exact sequence thus carries over to the relative Čech cohomology of the pair (X,Y). The construction for equivariant Čech cohomology is exactly the same, the only extra assumption we need is that Y is to be a τ-invariant subspace.

Now we can present the final cohomological formulation of the FKMM invariant. Recall that we constructed the FKMM invariant for spaces with finitely many isolated fixed points and only trivial "Real" line bundles, which, as we now know, means that $\check{H}^2(X;\mathbb{Z}_2,\widetilde{\mathbb{Z}}) = 0$. Under these assumptions, we have to following identification.

Lemma 9.8. Let (X, τ) be a space such that $\check{H}^2(X; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) = 0$, then

$$[X^\tau, \mathbb{C}^*]_{\mathbb{Z}_2} / [X, \mathbb{C}^*]_{\mathbb{Z}_2} \cong \check{H}^2(X, X^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \,.$$

Proof. The long exact sequence for the pair (X, X^τ) gives

$$\check{H}^1(X; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \xrightarrow{r} \check{H}^1(X^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \longrightarrow \check{H}^2(X, X^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \longrightarrow 0 \quad .$$

It follows that

$$\check{H}^2(X, X^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \cong \check{H}^1(X; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) / r(\check{H}^1(X^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}})) \cong [X^\tau, \mathbb{C}^*]_{\mathbb{Z}_2} / [X, \mathbb{C}^*]_{\mathbb{Z}_2} \,,$$

where the second isomorphisms is due to Proposition 9.7 and the action of $[g] \in [X, \mathbb{C}^*]_{\mathbb{Z}_2}$ on $[f] \in [X^\tau, \mathbb{C}^*]_{\mathbb{Z}_2}$ is given by multiplying f by the restriction of g, i.e. $([g], [f]) \mapsto [g|_{X^\tau} \cdot f]$. □

We can thus view the FKMM invariant as taking values in the relative cohomology $\check{H}^2(X, X^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}})$. In Appendix C, we prove that for time reversal spheres in $d \geq 2$,

$$\check{H}^2(S^d, (S^d)^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \cong \mathbb{Z}_2 \,,$$

thus establishing again the $\mathbb{Z}_2$ nature of the FKMM invariant in this case.

In [18], DeNittis and Gomi present a similar cohomological formulation of the FKMM invariant. The difference with the formulation here is that, where we use equivariant Čech and sheaf cohomology, DeNittis and Gomi use the so-called equivariant Borel cohomology. In a later work [47] they use this cohomological formulation of the FKMM invariant to further generalize it to a more general class of involutive spaces, whose fixed point set is not restricted to be a finite collection of points and may even be empty. The homotopy formulation is clearly not suited for such a generalization since if the fixed point set is empty, there can be no FKMM invariant in the homotopy formulation.

10 Outlook

In this thesis we have studied the topological classification of insulators. In the first part, we have restricted ourselves to insulators without any discrete symmetry requirements. A precise mathematical "model of choice" for insulators based on band theory and the tight binding approximation has been formulated. A classification of insulators with a certain fixed number of valence bands can be achieved within this formalism, as long as the total number of bands that we take into account is large enough. However, it has turned out that all topological information for insulators in less than or equal to three spatial dimensions is already contained in the two-dimensional, two-band model. This model has been studied extensively, and it has been shown that the most general two-dimensional, two-band model, of the form $h = \boldsymbol{B} \cdot \boldsymbol{\sigma}$, is completely classified by the winding number of $\boldsymbol{B} : X \to S^2$, where $X = \mathbb{T}^2$ or S^2. This result was already known, for example in [4]. In our presentation of the proof of this interpretation of the classification of topological insulators we have emphasized that this result relies heavily on some nice coincidences, most importantly the fact that $\boldsymbol{B} : X \to S^2$ can be identified with the classifying map of the valence bundle of the two-band model. Concrete models for topological insulators, however, might not be presented in a two-band form, but could naturally arise as models with many bands. Although it is abstractly speaking always possible to reduce to a two-band model, it would be desirable to have a concrete, computational method of reducing a higher-band model to a two-band model.

In the second part, we have studied time reversal symmetric insulators, with a focus on continuum models. Using methods readily available from the original construction of the Fu-Kane-Mele (FKM) invariant in the context of periodic models, we have proven its well-definedness also in the context of continuum models. The FKM invariant has been shown to be equivalent to the FKMM invariant for "Quaternionic" vector bundles by DeNittis and Gomi. Inspired by this work, the classification of "Quaternionic" vector bundles over time reversal spheres in terms of the FKMM invariant has been derived using elementary arguments in equivariant homotopy, and also less elementary tools from equivariant Čech and sheaf cohomology. These results imply that the physically motivated FKM invariant is the only topological invariant for continuum models of time reversal symmetric insulators. However, in the classification of "Quaternionic" bundles over the time reversal torus, it still remains to use our methods to prove that all "Real" line bundles over the time reversal torus are trivial. The proof of this fact is more difficult for the torus than for the sphere because of a significant increase in combinatorial overhead. In the case of the sphere, the combinatorics also seemed to get out of hand in $d \geq 3$, but these issues have been neatly avoided by a Mayer-Vieteris sequence argument. Even if the equivariant cohomology of the two-dimensional time reversal torus could be computed from the combinatorics, it remains to be seen how to extend the argument inductively to higher-dimensional tori. The development of a Künneth-type formula for equivariant Čech cohomology would be of great use here.

Finally, the ordinary insulators and the time reversal symmetric insulators are just two of ten topological phases of matter in the so-called *tenfold-way* [48]. Within these ten phases of matter, there are two more phases that correspond to topological insulators. Their classifications can be readily understood in terms of what has been presented in this work. However, six of the ten classes correspond to superconducting phases. One could ask whether the strategy of associating a vector bundle to an insulator and then classifying the vector bundle also works to classify the superconducting phases, and what types of vector bundles would arise in that case.

A Gauge invariance of the polarization tensor

In Section 4.3 we compute the Hall conductivity of a two-dimensional, two-band insulator in a gauge with $A_0 = 0$. We also work in the low-energy, low-temperature and long-wavelength limit. In this Appendix, we argue that the result of this computation is gauge invariant. We do this by arguing that the two diagrams that we have taken into account provide a gauge invariant theory in the limits under consideration. To do this, we use the following observation. Under a gauge transformation $A_\mu \to A_\mu + \partial_\mu \chi$ the physically measurable current J^μ should be invariant.[18] In other words, it must be true that

$$J^\mu = \pi^{\mu\nu} A_\nu = \pi^{\mu\nu}(A_\nu + p_\nu \chi).$$

Hence, at the level of currents, gauge invariance is equivalent to $\pi^{\mu\nu}(\omega, \boldsymbol{p}) p_\nu = 0$. What we will show is that in the long-wavelength and low-temperature limit, the polarization tensor coming from the "fish" and the "Kubo" diagrams satisfies

$$\lim_{\boldsymbol{p}\to 0} \pi^{\mu\nu}(\boldsymbol{p}, \omega) p_\nu = 0\,. \tag{A.1}$$

The low-energy limit $\omega \to 0$ imposes the additional requirement that $\sigma_{ij}(\omega) = \frac{1}{-i\omega}\pi_{ij}(\omega)$ be well-defined. In first instance, let us ignore this additional requirement.

In a general gauge, the action of the two-band model up to quadratics terms in A_μ is given by Eq. (4.15) with one additional term due to the minimal coupling $-i\hbar\omega_n \to -i\hbar\omega_n + eA_0(\boldsymbol{k}, i\omega_n)$,

$$\begin{aligned}
S[\phi^*, \phi; \boldsymbol{A}] = &\sum_n \int \frac{d\boldsymbol{k}}{(2\pi)^2}\, \phi_a^*(\boldsymbol{k}, i\omega_n)\left[-i\hbar\omega_n + h_{aa'}(\boldsymbol{k})\right]\phi_{a'}(\boldsymbol{k}, i\omega_n) \\
&+ \frac{e}{\hbar}\frac{1}{\sqrt{\hbar\beta}} \sum_{n,m} \int \frac{d\boldsymbol{k} d\boldsymbol{q}}{(2\pi)^4}\, \phi_a^*(\boldsymbol{k}+\boldsymbol{q}, i\omega_n + i\omega_m) A_0(\boldsymbol{q}, i\omega_m)\phi_{a'}(\boldsymbol{k}, i\omega_n) \\
&+ \frac{e}{\hbar}\frac{1}{\sqrt{\hbar\beta}} \sum_{n,m} \int \frac{d\boldsymbol{k} d\boldsymbol{q}}{(2\pi)^4}\, \phi_a^*(\boldsymbol{k}+\boldsymbol{q}, i\omega_n + i\omega_m) \left[A_i(\boldsymbol{q}, i\omega_m)\frac{\partial h_{aa'}(\boldsymbol{k})}{\partial k^i}\right]\phi_{a'}(\boldsymbol{k}, i\omega_n) \\
&+ \left(\frac{e}{\hbar}\right)^2 \frac{1}{\hbar\beta} \sum_{n,m,r} \int \frac{d\boldsymbol{k} d\boldsymbol{q} d\boldsymbol{p}}{(2\pi)^6}\, \phi_a^*(\boldsymbol{k}+\boldsymbol{q}+\boldsymbol{p}, i\omega_n + i\omega_m + i\omega_r) \\
&\times \left[A_i(\boldsymbol{q}, i\omega_m) A_j(\boldsymbol{p}, i\omega_r)\frac{\partial^2 h_{aa'}(\boldsymbol{k})}{\partial k^i \partial k^j}\right]\phi_{a'}(\boldsymbol{k}, i\omega_n)\,.
\end{aligned} \tag{A.2}$$

The extra A_0 terms have the effect of introducing an extra Green function

$$-\hbar G^{-1}_{1';ab}(\boldsymbol{k}, i\omega_n; \boldsymbol{k}', i\omega_{n'}) = \frac{e}{\hbar}\frac{1}{\sqrt{\hbar\beta}} \sum_m \int \frac{d\boldsymbol{q}}{(2\pi)^2} A_0(\boldsymbol{q}, i\omega_m)\delta_{ab}\delta(\boldsymbol{k}-\boldsymbol{k}'-\boldsymbol{q})\delta(i\omega_n - i\omega_{n'} - i\omega_m).$$

into the expansion $G^{-1} = G_0^{-1} + G_1^{-1} + G_{1'}^{-1} + G_2^{-1}$. There are now two extra terms coming from the Kubo diagram that contribute to the polarization tensor, namely,

$$-\frac{1}{2}\,\mathrm{Tr}[G_0 G_{1'}^{-1} G_0 G_{1'}^{-1}] \quad \text{and} \quad -\frac{1}{2}\,\mathrm{Tr}[G_0 G_1^{-1} G_0 G_{1'}^{-1}].$$

[18] The index $\mu = 0, 1, 2$ is that of a vector in three-dimensional Minkowski space with metric $\eta = \mathrm{diag}(-1, 1, 1)$. In principle, we should be careful about index placement. For the present computation, however, this turns out to be a non-issue and so we will not worry about index placement. Moreover, the convention below is that repeated indices are summed over, irregardless of their positioning.

From these terms, we may compute completely analogously to Eq. (4.18) that in the long-wavelength limit $\boldsymbol{p} \to 0$,

$$\pi_{00}(i\omega_r) = \left(\frac{e}{\hbar}\right)^2 \frac{1}{\beta} \sum_{\boldsymbol{k},n} \mathrm{Tr}\Bigg[\left(\frac{|\boldsymbol{k}+\rangle \langle \boldsymbol{k}+|}{-i\hbar\omega_n + \epsilon_+(\boldsymbol{k})} + \frac{|\boldsymbol{k}-\rangle \langle \boldsymbol{k}-|}{-i\hbar\omega_n + \epsilon_-(\boldsymbol{k})} \right) \\ \times \left(\frac{|\boldsymbol{k}+\rangle \langle \boldsymbol{k}+|}{-i\hbar\omega_n + i\hbar\omega_r + \epsilon_+(\boldsymbol{k})} + \frac{|\boldsymbol{k}-\rangle \langle \boldsymbol{k}-|}{-i\hbar\omega_n + i\hbar\omega_r + \epsilon_-(\boldsymbol{k})} \right) \Bigg]$$

and, similarly,

$$\pi_{i0}(i\omega_r) = \left(\frac{e}{\hbar}\right)^2 \frac{1}{\beta} \sum_{\boldsymbol{k},n} \mathrm{Tr}\Bigg[\left(\frac{|\boldsymbol{k}+\rangle \langle \boldsymbol{k}+|}{-i\hbar\omega_n + \epsilon_+(\boldsymbol{k})} + \frac{|\boldsymbol{k}-\rangle \langle \boldsymbol{k}-|}{-i\hbar\omega_n + \epsilon_-(\boldsymbol{k})} \right) \\ \times \left(\frac{|\boldsymbol{k}+\rangle \langle \boldsymbol{k}+|}{-i\hbar\omega_n + i\hbar\omega_r + \epsilon_+(\boldsymbol{k})} + \frac{|\boldsymbol{k}-\rangle \langle \boldsymbol{k}-|}{-i\hbar\omega_n + i\hbar\omega_r + \epsilon_-(\boldsymbol{k})} \right) \frac{\partial h(\boldsymbol{k})}{\partial k^i} \Bigg].$$

Indeed, from the action Eq. (A.2) we see that a vertex with index i should come with a factor $\frac{e}{\hbar}\frac{\partial h(\boldsymbol{k})}{\partial k^i}$ whilst a vertex with index 0 only comes with a factor $\frac{e}{\hbar}$. Analogously to Eq. (4.19), only mixed terms with one ϵ_+ and one ϵ_- in the denominator survive in the low-temperature limit after working out the Matsubara sums and doing the analytic continuation $i\omega_r \to \omega + i\eta^+$. In the present case, however, terms with mixed ϵ_+ and ϵ_- denominators contain a factor $\langle \boldsymbol{k}+|\boldsymbol{k}-\rangle = 0$ in the numerator. Therefore, $\pi_{\mu 0}(\boldsymbol{p},\omega)$ for $\mu = 0,1,2$ vanishes in the long-wavelength, low-temperature limit. The vanishing of these components means that it is always possible to gauge away A_0 without changing the resulting current J^μ. It follows that the computation of the Hall conductivity in Section 4.3 is independent of A_0. Therefore, the result obtained in the gauge $A_0 = 0$ is completely general.

For completeness sake, let us also establish the complete gauge invariance of the current. In order to show that Eq. (A.1) is satisfied, it remains to show that

$$\lim_{\boldsymbol{p}\to 0} \pi^{ij}(\boldsymbol{p},\omega) p_j = 0. \tag{A.3}$$

As the limit of a product is the product of a limit, it suffices to prove that $\lim_{\boldsymbol{p}\to 0} \pi^{ij}(\boldsymbol{p},\omega)$ exists. In fact, we need to check a slightly stronger requirement than Eq. (A.3) because the conductivity tensor $\sigma_{ij}(\omega) = \frac{1}{-i\omega}\pi_{ij}(\omega)$ needs to be well-defined. We thus need that $\lim_{\boldsymbol{p}\to 0} \pi_{ij}(\boldsymbol{p},\omega) = O(\omega)$ in the long-wavelength, low-temperature limit. In Section 4.3 we have claimed this to be the case because the constant term contributing to $\pi_{ij}(\omega)$ coming from the Kubo diagram is cancelled by the contribution from the fish diagram. Let us prove this claim here.

Recall from Eq. (4.20) that the contribution to $\pi_{ij}(\omega)$ from the Kubo diagram is

$$\pi_{ij}^{\mathrm{Kubo}}(\omega) = -\left(\frac{e}{\hbar}\right)^2 \int \frac{d\boldsymbol{k}}{(2\pi)^2} \frac{(\epsilon_+(\boldsymbol{k}) - \epsilon_-(\boldsymbol{k}))\left(M_{ij} n_{\mathrm{FD}}(\epsilon_-(\boldsymbol{k}) + \hbar\omega) + M_{ji} n_{\mathrm{FD}}(\epsilon_-(\boldsymbol{k}))\right)}{(\epsilon_+(\boldsymbol{k}) - \epsilon_-(\boldsymbol{k}))^2 - (\hbar\omega)^2} \\ -\left(\frac{e}{\hbar}\right)^2 \int \frac{d\boldsymbol{k}}{(2\pi)^2} \frac{\hbar\omega(M_{ij} n_{\mathrm{FD}}(\epsilon_-(\boldsymbol{k}) + \hbar\omega) - M_{ji} n_{\mathrm{FD}}(\epsilon_-(\boldsymbol{k})))}{(\epsilon_+(\boldsymbol{k}) - \epsilon_-(\boldsymbol{k}))^2 - (\hbar\omega)^2}.$$

The second term is $O(\omega)$, the first term is not. In the limit $\omega \to 0$, the first term becomes

$$\pi_{ij}^{\mathrm{Kubo,first}} = -\left(\frac{e}{\hbar}\right)^2 \int \frac{d\boldsymbol{k}}{(2\pi)^2} \frac{M_{ij} + M_{ji}}{\epsilon_+(\boldsymbol{k}) - \epsilon_-(\boldsymbol{k})}, \tag{A.4}$$

where we recall that

$$M_{ij} = \langle \boldsymbol{k} + | \frac{\partial h(\boldsymbol{k})}{\partial k^i} | \boldsymbol{k} - \rangle \, \langle \boldsymbol{k} - | \frac{\partial h(\boldsymbol{k})}{\partial k^j} | \boldsymbol{k} + \rangle \ .$$

On the other hand, the contribution to the polarization coming from the fish diagram $\mathrm{Tr}[G_0 G_2^{-1}]$ is

$$\pi_{ij}^{\mathrm{fish}}(\boldsymbol{p}, i\omega_r) = -\left(\frac{e}{\hbar}\right)^2 \frac{1}{\beta} \sum_n \int \frac{d\boldsymbol{k}}{(2\pi)^2} \left[-i\hbar\omega_n + h(\boldsymbol{k})\right]^{-1}_{ab} \frac{\partial^2 h^{ba}(\boldsymbol{k})}{\partial k^i \partial k^j} \, . \tag{A.5}$$

Note that this term does not depend on $\boldsymbol{p}$ or $i\omega_r$. Diagonalizing the Hamiltonian in the basis $|\boldsymbol{k}\pm\rangle$ and working out the Matsubara sums yields

$$\pi_{ij}^{\mathrm{fish}} = \left(\frac{e}{\hbar}\right)^2 \sum_n \int \frac{d\boldsymbol{k}}{(2\pi)^2} \, \langle \boldsymbol{k} - | \frac{\partial^2 h(\boldsymbol{k})}{\partial k^i \partial k^j} | \boldsymbol{k} - \rangle \tag{A.6}$$

for the contribution in the low-temperature limit. Using $h(\boldsymbol{k}) = [\epsilon_+(\boldsymbol{k}) - \epsilon_-(\boldsymbol{k})] \, |\boldsymbol{k}+\rangle \, \langle \boldsymbol{k}+| + \epsilon_-(\boldsymbol{k})$, a straightforward calculation shows that the matrix element in the integrand can be computed to yield

$$\pi_{ij}^{\mathrm{fish}} = \left(\frac{e}{\hbar}\right)^2 \sum_n \int \frac{d\boldsymbol{k}}{(2\pi)^2} \left[\frac{M_{ij} + M_{ji}}{\epsilon_+(\boldsymbol{k}) - \epsilon_-(\boldsymbol{k})} + \frac{\partial^2}{\partial k^i \partial k^j} \epsilon_-(\boldsymbol{k}) \right] . \tag{A.7}$$

The first term precisely cancels Eq. (A.4). The second term also vanishes. Indeed, $\epsilon_- : X \to \mathbb{R}$ is a scalar function on a space X with no boundary, namely either $\mathbb{T}^2$ or S^2. Therefore, the integral of the total derivative vanishes. This establishes that taking into account both the Kubo and the fish diagram, the polarization tensor satisfies

$$\lim_{\boldsymbol{p} \to 0} \pi_{ij}(\boldsymbol{p}, \omega) = O(\omega)$$

in the low-temperature limit, thus proving the gauge independence of the computations of the conductivity done in Section 4.3.

B Equivalence of Čech and sheaf cohomology

The constructions of Čech cohomology and sheaf cohomology have been reviewed and generalized to an equivariant setting in Section 9. Although their constructions are completely different, it turns out that Čech cohomology and sheaf cohomology are equivalent. More precisely, if we fix a topological space X that is nice enough, say paracompact Hausdorff, then Čech and sheaf cohomology are naturally equivalent as functors $\mathrm{Sh}(X) \to \mathrm{Ab}$ from the category of sheaves over X to the category of abelian groups. The proof is based on the following theorem, which is Corollary 6.49 in Rotman's book on Homological Algebra [40].

Theorem B.1 ([40, Corollary 6.49])**.** Let $(F^n : \mathcal{A} \to \mathcal{B})_{n \geq 0}$, $(G^n)_{n \geq 0}$ be sequences of additive covariant functors, where $\mathcal{A}$ and $\mathcal{B}$ are abelian categories and $\mathcal{A}$ has enough injectives. If,

(i) for every short exact sequence $0 \to A \to B \to C \to 0$, there are long exact sequences with natural connecting homomorphisms,

(ii) F^0 is naturally isomorphic to G^0 ,

(iii) $F^n(E) = 0 = G^n(E)$ for all injective objects E and all $n \geq 1$,

then F^n is naturally isomorphic to G^n for all $n \geq 0$.

The intuition behind this result is roughly the following. For any object $A \in \mathcal{A}$ we can find $E \in \mathcal{A}$ such that A injects into E and we thus get a short exact sequence

$$0 \to A \to E \to C \to 0\,, \tag{B.1}$$

where C is the cokernel of $A \to E$. By (i), the short exact sequence Eq. (B.1) gives rise to a long exact sequences. From (ii) and a Five Lemma type argument one deduces that $F^1(A)$ is isomorphic to $G^1(A)$. This establishes the base case in an inductive proof. Indeed, in positive degree, many terms vanish in the long exact sequences because of (iii) and this will provide the rest of the isomorphisms $F^n(A) \to G^n(A)$. The naturality assumptions are needed for the relevant diagrams to commute.

We will first apply Theorem B.1 to the two sequences $(\check{H}^n(X;-) : \mathrm{Sh}(X) \to \mathrm{Ab})_{n\geq 0}$ and $(H^n(X;-) : \mathrm{Sh}(X) \to \mathrm{Ab})_{n\geq 0}$ of Čech and sheaf cohomology groups. The technical assumption of being an additive functor between abelian categories in this case just means that cohomology behaves well when taking direct sums of sheaves. We have also seen that the category of sheaves of abelian groups has enough injectives. The important thing is thus to check that assumptions (i), (ii) and (iii) in the Theorem are satisfied. We then indicate how to adapt the proofs of assumptions (i), (ii) and (iii) in the equivariant case to also establish the equivalence there.

B.1 Ordinary setting

Let $(\check{H}^n(X;-) : \mathrm{Sh}(X) \to \mathrm{Ab})_{n\geq 0}$ and $(H^n(X;-) : \mathrm{Sh}(X) \to \mathrm{Ab})_{n\geq 0}$ be the Čech and sheaf cohomology functors. To verify assumption (i) of Theorem B.1, it has to be argued that given a short exact sequence of sheaves, there is a long exact sequence in cohomology. For sheaf cohomology, the result can be found in [41] and is based on purely categorical arguments. For Čech cohomology, the result has been proven by Serre in [49] in the case that X is paracompact Hausdorff. We will henceforth restrict to this case. In both cohomology theories, the underlying reason for the existence of the long exact sequence is that the section functor $\Gamma(X,-) : \mathrm{Sh}(X) \to \mathrm{Ab}$ is left exact, but not right exact. In the case of Čech cohomology, given a short exact sequence of sheaves

$$0 \to \mathcal{F} \to \mathcal{G} \to \mathcal{H} \to 0$$

this only gives rise to the exact sequence

$$0 \to \check{C}^*(\mathcal{U};\mathcal{F}) \to \check{C}^*(\mathcal{U};\mathcal{G}) \to \check{C}^*(\mathcal{U};\mathcal{H})$$

on the level of Čech cochains. Indeed, the last map may fail to be surjective because a section of $\mathcal{H}(U_\alpha)$ may fail to have a primitive in $\mathcal{G}(U_\alpha)$. However, by the exactness of the sequence of sheaves, which is an exactness at the level of stalks, primitives do always exist on sufficiently small neighbourhoods. Roughly speaking, Serre showed that for any $(f_{\alpha_0\ldots\alpha_n}) \in \check{C}^n(\mathcal{U},\mathcal{H})$, there always exists a refinement $\mathcal{V} = \{V_\beta\}_{\beta\in\mathcal{B}}$ of $\mathcal{U}$, consisting of small enough opens, so that f has primitives over the V_β. Appropriate use of this fact leads to the long exact sequence in cohomology.

To verify assumption (ii), we note that degree zero Čech cohomology and degree zero sheaf cohomology are both naturally isomorphic to the global sections. For Čech cohomology with coefficients in a sheaf $\mathcal{F}$, we have shown in Section 9 that $\check{H}^0(X;\mathcal{F}) \cong \Gamma(X,\mathcal{F})$.

It is straightforward to verify that this isomorphism is natural, since sheaf morphisms play well with restrictions. To show that $H^0(X;\mathcal{F}) \cong \Gamma(X;\mathcal{F})$, we pick any injective resolution $0 \to \mathcal{F} \to \mathcal{I}^0 \to \mathcal{I}^1 \to \cdots$. The left exactness of the global section functor implies that

$$0 \to \mathcal{F}(X) \to \mathcal{I}^0(X) \to \mathcal{I}^1(X)$$

is exact. Therefore, denoting the inclusion of $\mathcal{F}$ into $\mathcal{I}^0$ by i,

$$H^0(X;\mathcal{F}) = \ker d(X) = \operatorname{im} i(X) \cong \mathcal{F}(X) = \Gamma(X,\mathcal{F})\,.$$

By general properties of injective resolutions, the isomorphism is natural [41]. We thus conclude that there is a chain of natural isomorphisms

$$\check{H}^0(X;\mathcal{F}) \cong \Gamma(X,\mathcal{F}) \cong H^0(X;\mathcal{F}),$$

which verifies (ii).

Let us then consider assumption (iii), which states that the cohomology of injective sheaves vanishes in positive degree cohomology. In the case of sheaf cohomology, this is immediate. If $\mathcal{I}$ is an injective sheaf, then

$$0 \to \mathcal{I} \to \mathcal{I} \to 0$$

is an injective resolution. The sheaf cohomology is thus the cohomology of the complex

$$0 \to \mathcal{I}(X) \to 0\,,$$

which clearly vanishes in positive degree.

To show that Čech cohomology of an injective sheaf vanishes in positive degree requires significant work, however. Let $U \subset X$ be an open set and define the sheaf $\mathbb{Z}_U$ by

$$\mathbb{Z}_U(V) = \begin{cases} \mathbb{Z} & \text{if } V \subseteq U \\ 0 & \text{else} \end{cases}\,.$$

This sheaf can be used to express the sections of $\mathcal{F}$ over U in the form of a Hom-group between sheaves.

Lemma B.2. Let $\mathcal{F}$ be a sheaf over X and let $U \subset X$ be open, then there is an isomorphism of groups

$$\operatorname{Hom}(\mathbb{Z}_U, \mathcal{F}) \cong \mathcal{F}(U)\,.$$

Proof. Given a sheaf morphism $\varphi : \mathbb{Z}_U \to \mathcal{F}$ we can construct a section of $\mathcal{F}$ over U by evaluating $\varphi(U)$ on the constant 1 section of $\mathbb{Z}_U(U) \cong \mathbb{Z}$. This defines a map

$$J : \operatorname{Hom}(\mathbb{Z}_U, \mathcal{F}) \to \mathcal{F}(U) \quad \varphi \mapsto \varphi(U)(1)\,.$$

The claim is that J is an isomorphism of groups. To show injectivity, suppose that $\varphi \in \operatorname{Hom}(\mathbb{Z}_U, \mathcal{F})$ is such that $J(\varphi) = \varphi(U)(1) = 0$. Since $1 \in \mathbb{Z}_U(U)$ is a generator, this implies that $\varphi(U) = 0$. For any $V \subseteq U$, it follows that $\varphi(V) = \rho_V^U \varphi(U) = 0$. Moreover, it also holds that $\varphi(V) = 0$ for any $V \not\subseteq U$, since $\mathbb{Z}_U(V) = 0$. Therefore, $\varphi : \mathbb{Z}_U \to \mathcal{F}$ is the zero morphism. This proves injectivity.

For surjectivity, suppose $s \in \mathcal{F}(U)$ is a section. The requirement $\varphi(U)(1) = s$ defines a morphism $\varphi \in \operatorname{Hom}(\mathbb{Z}_U, \mathcal{F})$. Indeed, if $V \subseteq U$, then compatibility with restriction fixes $\varphi(V) = \varphi(U)\rho_V^U$ and if $V \not\subseteq U$, then the fact that $\mathbb{Z}_U(V) = 0$ fixes $\varphi(V) = 0$. This establishes the isomorphism $\operatorname{Hom}(\mathbb{Z}_U, \mathcal{F}) \to \mathcal{F}(U)$. □

For any sheaf $\mathcal{F}$ and open cover $\mathcal{U} = \{U_\alpha\}_{\alpha\in\mathcal{A}}$, the degree n Čech cochain group can be expressed as

$$\check{C}^n(\mathcal{U};\mathcal{F}) = \prod_{\langle\alpha_0\dots\alpha_n\rangle\in\mathcal{A}_{n+1}} \mathcal{F}(U_{\alpha_0\dots\alpha_n})\,,$$

where $\mathcal{A}_{n+1}$ is the set of all $n+1$-overlaps and the $\langle\dots\rangle$ remind us that cochains need to be antisymmetric in their indices. Using the Lemma above and properties of Hom, we can rewrite this as

$$\check{C}^n(\mathcal{U};\mathcal{F}) \cong \prod_{\langle\alpha_0\dots\alpha_n\rangle} \mathrm{Hom}\left(\mathbb{Z}_{U_{\alpha_0\dots\alpha_n}},\mathcal{F}\right) \cong \mathrm{Hom}\Big(\bigoplus_{\langle\alpha_0\dots\alpha_n\rangle} \mathbb{Z}_{U_{\alpha_0\dots\alpha_n}},\mathcal{F}\Big)\,, \tag{B.2}$$

where the indices $\alpha_0\dots\alpha_n$ run over the $n+1$-fold overlaps $\mathcal{A}_{n+1}$ and elements are still antisymmetric in their indices. Let us briefly explain the antisymmetry in the Hom-group in the case of one overlap $U_1\cap U_2$. If $(\varphi_{12},\varphi_{21})\in\mathrm{Hom}(\bigoplus^2_{\langle\alpha_0\alpha_1\rangle=1}\mathbb{Z}_{U_{\alpha_0\alpha_1}},\mathcal{F})$, then, for $U = U_{12}$, the map $\varphi_{12}(U):\mathbb{Z}=\mathbb{Z}_{U_{12}}(U)\to\mathcal{F}(U)$ is minus the map $\varphi_{21}(U):\mathbb{Z}=\mathbb{Z}_{U_{21}}(U)\to\mathcal{F}(U)$.

Let us define the sheaf $\mathbb{Z}_{\mathcal{U},n} := \bigoplus_{\langle\alpha_0\dots\alpha_n\rangle}\mathbb{Z}_{U_{\alpha_0\dots\alpha_n}}$. We collect all the sheaves $(\mathbb{Z}_{\mathcal{U},n})_{n\geq 0}$ in a chain complex

$$\dots \longrightarrow \mathbb{Z}_{\mathcal{U},2} \xrightarrow{\partial} \mathbb{Z}_{\mathcal{U},1} \xrightarrow{\partial} \mathbb{Z}_{\mathcal{U},0} \longrightarrow 0\ , \tag{B.3}$$

where the differential ∂ is the sum of the maps $\mathbb{Z}_{U_{\alpha_0\dots\alpha_n}}\to\mathbb{Z}_{U_{\alpha_0\dots\hat{\alpha}_i\dots\alpha_n}}$ given by $(-1)^i$ times the canonical map. The Čech cochain complex is recovered from (B.3) by applying $\mathrm{Hom}(-,\mathcal{F})$. The fact that applying $\mathrm{Hom}(-,\mathcal{F})$ gives the right cochain groups follows from the isomorphism in Eq. (B.2). What we still have to argue is that under the $\mathrm{Hom}(-,\mathcal{F})$ isomorphism, the Čech differential δ is given by precomposition with ∂. To this end, consider the following diagram

$$\begin{array}{ccc} \mathrm{Hom}(\mathbb{Z}_{\mathcal{U},n},\mathcal{F}) & \xrightarrow{\partial^*} & \mathrm{Hom}(\mathbb{Z}_{\mathcal{U},n+1},\mathcal{F}) \\ \downarrow\cong & & \downarrow\cong \\ \check{C}^n(\mathcal{U};\mathcal{F}) & \xrightarrow{\delta} & \check{C}^{n+1}(\mathcal{U};\mathcal{F}) \end{array}\ ,$$

where the vertical arrows are the isomorphisms induced by $\varphi\mapsto\varphi(U)(1)$ as in Lemma B.2. We have to show that the diagram commutes. To do this, let $(f_{\alpha_0\dots\alpha_n})\in\check{C}^n(\mathcal{U};\mathcal{F})$ be a cochain. Under the vertical isomorphism, this cochain (f) maps to the sheaf morphism

$$\varphi:\bigoplus_{\langle\alpha_0\dots\alpha_n\rangle}\mathbb{Z}_{U_{\alpha_0\dots\alpha_n}}\to\mathcal{F}\,,$$

which sends $1_{\vec{\alpha}}\in\mathbb{Z}_{\mathcal{U}}(U_{\vec{\alpha}})$ to $f_{\vec{\alpha}}\in\mathcal{F}(U_{\vec{\alpha}})$. Here, $1_{\vec{\alpha}}$ is the section in $\bigoplus_{\langle\vec{\beta}\rangle}\mathbb{Z}_{U_{\vec{\beta}}}(U_{\vec{\alpha}})$ which is 1 in the $\vec{\beta}=\vec{\alpha}$ slot and zero in all the other slots. Then,

$$\begin{aligned}(\partial^*\varphi)(U_{\alpha_0\dots\alpha_{n+1}})(1_{\alpha_0\dots\alpha_{n+1}}) &= \varphi(U_{\alpha_0\dots\alpha_{n+1}})(\partial(U_{\alpha_0\dots\alpha_{n+1}})(1_{\alpha_0\dots\alpha_{n+1}})) \\ &= \varphi(U_{\alpha_0\dots\alpha_{n+1}})\left(\sum_{i=0}^{n+1}(-1)^i 1_{\alpha_0\dots\hat{\alpha}_i\dots\alpha_{n+1}}\right) \\ &= \sum_{i=0}^{n+1}(-1)^i\varphi(U_{\alpha_0\dots\alpha_{n+1}})(1_{\alpha_0\dots\hat{\alpha}_i\dots\alpha_{n+1}}) \\ &= \sum_{i=0}^{n+1}(-1)^i f_{\alpha_0\dots\hat{\alpha}_i\dots\alpha_{n+1}}\,,\end{aligned}$$

where the appropriate restrictions of $1_{\alpha_0\ldots\hat{\alpha}_i\ldots\alpha_n}$ and $f_{\alpha_0\ldots\hat{\alpha}_i\ldots\alpha_n}$ are implicit. This proves that $\mathrm{Hom}(\mathbb{Z}_{\mathcal{U},*},\mathcal{F})$ together with the differential ∂^* is isomorphic to the Čech cochain complex $\check{C}^*(\mathcal{U},\mathcal{F})$. The merit of this reformulation is that the chain complex $\mathbb{Z}_{\mathcal{U},*}$ is exact in positive degrees [42, Lemma 03AT], so that we can now prove the desired result.

Proposition B.3. Let X be a space and let $\mathcal{I}$ be an injective sheaf on X. Then, $\check{H}^n(X;\mathcal{I}) = 0$ for $n > 0$.

Proof. Let $\mathcal{U}$ be an open cover of X. By the discussion above, the Čech cochain complex $\check{C}^*(\mathcal{U};\mathcal{I})$ is isomorphic to $\mathrm{Hom}(\mathbb{Z}_{\mathcal{U},*},\mathcal{I})$. Since $\mathcal{I}$ is injective, the functor $\mathrm{Hom}(-,\mathcal{I})$ preserves both left and right exactness. In fact, left exactness always holds, but the injectivity also provides the right exactness. But then because $\mathbb{Z}_{\mathcal{U},*}$ is exact in positive degrees, it follows that $\mathrm{Hom}(\mathbb{Z}_{\mathcal{U},*},\mathcal{I})$ is exact in positive degrees. Therefore, its cohomology in positive degrees vanishes and so also the Čech cohomology vanishes in positive degree. □

B.2 Equivariant setting

Let (X,τ) be an involutive space and denote by $(\check{H}^n(X;\mathbb{Z}_2,-) : \mathrm{Sh}_{\mathbb{Z}_2}(X,\tau) \to \mathrm{Ab})_{n\geq 0}$ and $(H^n(X;\mathbb{Z}_2,-) : \mathrm{Sh}_{\mathbb{Z}_2}(X,\tau) \to \mathrm{Ab})_{n\geq 0}$ the equivariant Čech and equivariant sheaf cohomology functors. Just like in the ordinary setting, the equivalence of these two functors is established by checking conditions (i), (ii) and (iii) of Theorem B.1.

One can view the equivariant sheaf cohomology $H^*(X;\mathbb{Z}_2,-)$ as the right derived functor of the equivariant section functor $\Gamma(X,-)^{\mathbb{Z}_2}$. The proofs of assumption (i) and (iii) then follow from exactly the same homological algebra arguments as in the ordinary case. Furthermore, the degree zero equivariant sheaf cohomology is seen to be naturally isomorphic to the global equivariant sections. It also follows from very similar arguments to the ordinary case that the degree zero equivariant Čech cohomology is naturally isomorphic to the global equivariant sections too. Hence, we have the chain of natural isomorphisms

$$H^0(X;\mathbb{Z}_2,-) \cong \Gamma(X,-)^{\mathbb{Z}_2} \cong \check{H}^0(X;\mathbb{Z}_2,-)\,,$$

which establishes (ii).

To show that a short exact sequence of equivariant sheaves give rise to a long exact sequence in equivariant Čech cohomology one can straightforwardly generalize the proof by Serre in the ordinary case [49]. The only alteration that has to be made is to change all the covers in Serre's proof into $\mathbb{Z}_2$-covers. This then proves (i) for equivariant Čech cohomology.

Finally, it has to be argued that $\check{H}^n(X;\mathbb{Z}_2,\widetilde{\mathcal{I}}) = 0$ for $n > 0$ and $\widetilde{\mathcal{I}}$ and injective $\mathbb{Z}_2$-sheaf. The key point is to generalize the isomorphism Eq. (B.2) of Čech cochains with Hom-sets. Let $\mathcal{U} = \{U_\alpha\}_{\alpha\in\mathcal{A}}$ be a $\mathbb{Z}_2$-cover of (X,τ). To do so, we have to upgrade the sheaf $\mathbb{Z}_{\mathcal{U},n}$ to $\mathbb{Z}_2$-sheaf. Define the sheaf morphism $\theta_n^{\mathbb{Z}} : \mathbb{Z}_{\mathcal{U},n} \to \tau^*\mathbb{Z}_{\mathcal{U},n}$ through

$$\theta_n^{\mathbb{Z}}(U_{\vec{\alpha}})(1_{\vec{\alpha}}) = 1_{\tau\vec{\alpha}}\,, \tag{B.4}$$

where again $1_{\vec{\alpha}}$ denotes the section in $\mathbb{Z}_{\mathcal{U},n}(U_{\vec{\alpha}}) = \bigoplus_{\langle\vec{\beta}\rangle}\mathbb{Z}_{U_{\vec{\beta}}}(U_{\vec{\alpha}})$ that is the constant 1 section in the $\vec{\alpha}$-component of the direct sum and zero in all the other components. By the same arguments as in Lemma B.2, Eq. (B.4) leads to a well-defined sheaf morphism. The morphism $\theta_n^{\mathbb{Z}}$ defines a $\mathbb{Z}_2$-structure on $\mathbb{Z}_{\mathcal{U},n}$. We denote the resulting $\mathbb{Z}_2$-sheaf by $\widetilde{\mathbb{Z}}_{\mathcal{U},n}$. The differential ∂ of the chain complex Eq. (B.3) is equivariant with respect to the $\mathbb{Z}_2$-structure $\theta^{\mathbb{Z}}$ so that we obtain a chain complex of $\mathbb{Z}_2$-sheaves

$$\ldots \longrightarrow \widetilde{\mathbb{Z}}_{\mathcal{U},2} \xrightarrow{\partial} \widetilde{\mathbb{Z}}_{\mathcal{U},1} \xrightarrow{\partial} \widetilde{\mathbb{Z}}_{\mathcal{U},0} \longrightarrow 0\ . \tag{B.5}$$

Lemma B.4. Let $\widetilde{\mathcal{F}} = (\mathcal{F}, \theta^{\mathcal{F}})$ be a $\mathbb{Z}_2$-sheaf over an involutive space (X, τ) and let $\mathcal{U} = \{U_\alpha\}_{\alpha \in \mathcal{A}}$ be a $\mathbb{Z}_2$-cover. Then, there is an isomorphism of cochain complexes

$$\check{C}^*(\mathcal{U}; \widetilde{\mathcal{F}}) \cong \mathrm{Hom}_{\mathbb{Z}_2}(\widetilde{\mathbb{Z}}_{\mathcal{U},*}, \widetilde{\mathcal{F}}) \,.$$

Proof. Recall that, by definition, $\check{C}^n(\mathcal{U}, \widetilde{\mathcal{F}}) = (\prod_{\langle \vec{\alpha} \rangle} \mathcal{F}(U_{\vec{\alpha}}))^{\mathbb{Z}_2}$, where $(f_{\vec{\alpha}})$ is a $\mathbb{Z}_2$-invariant cochain if $\theta^{\mathcal{F}}(U_{\vec{\alpha}})(f_{\vec{\alpha}}) = f_{\tau\vec{\alpha}}$. Given a $\mathbb{Z}_2$-invariant cochain $(f_{\alpha_0 \dots \alpha_n})$ we can construct an element Φ of $\mathrm{Hom}_{\mathbb{Z}_2}(\widetilde{\mathbb{Z}}_{\mathcal{U},*}, \widetilde{\mathcal{F}})$ by defining Φ on the $\alpha_0 \dots \alpha_n$-component of the direct sum in $\mathbb{Z}_{\mathcal{U},n}$ through

$$\Phi(U_{\vec{\alpha}})(1_{\vec{\alpha}}) = f_{\vec{\alpha}} \,.$$

To see that the element Φ constructed this way is indeed a $\mathbb{Z}_2$-sheaf morphism note that for any $\vec{\alpha}$, we have

$$\theta^{\mathcal{F}}(U_{\vec{\alpha}})\Phi(U_{\vec{\alpha}})(1_{\vec{\alpha}}) = \theta^{\mathcal{F}}(U_{\vec{\alpha}})(f_{\vec{\alpha}}) = f_{\tau\vec{\alpha}} = \Phi(U_{\tau\vec{\alpha}})(1_{\tau\vec{\alpha}}) = \Phi(\tau U_{\vec{\alpha}})\theta^{\mathbb{Z}}(U_{\vec{\alpha}})(1_{\vec{\alpha}}) \,.$$

Therefore, on all generators Φ is compatible with the $\mathbb{Z}_2$-structures $\theta^{\mathbb{Z}}$ and $\theta^{\mathcal{F}}$, which means that Φ is in fact a $\mathbb{Z}_2$-morphism. Define then the map

$$\check{C}^*(\mathcal{U}; \widetilde{\mathcal{F}}) \to \mathrm{Hom}_{\mathbb{Z}_2}(\widetilde{\mathbb{Z}}_{\mathcal{U},*}, \widetilde{\mathcal{F}}), \quad (f_{\vec{\alpha}}) \mapsto \Phi \,,$$

where Φ is as constructed above.

We claim that this map gives the required isomorphism. Injectivity is clear, since Φ is the zero morphism if and only if it vanishes on all the generators $1_{\vec{\alpha}}$ and $\Phi(U)(1_{\vec{\alpha}}) = 0$ for all $U \subset X$ if and only if $f_{\vec{\alpha}} = 0$. Surjectivity is also straightforward, since any $\mathbb{Z}_2$-morphism in $\mathrm{Hom}_{\mathbb{Z}_2}(\widetilde{\mathbb{Z}}_{\mathcal{U},*}, \widetilde{\mathcal{F}})$ is fully determined by what it does on the generators $1_{\vec{\alpha}}$. □

By exactness of $\mathrm{Hom}_{\mathbb{Z}_2}(-, \widetilde{\mathcal{I}})$ for injective sheaves, we then immediately get the analogue of Proposition B.3.

Proposition B.5. Let (X, τ) be an involutive space and let $\widetilde{\mathcal{I}}$ be an injective $\mathbb{Z}_2$-sheaf on (X, τ). Then, $\check{H}^n(X; \mathbb{Z}_2, \widetilde{\mathcal{I}}) = 0$ for $n > 0$.

C Equivariant sheaf cohomology of $\widetilde{S}^d$

In this Appendix, we compute some important equivariant sheaf cohomology groups of the time reversal sphere $\widetilde{S}^d = (S^d, \tau)$ in the case $d \geq 2$. For $d = 1$, the equivariant cohomology groups are already given by Eq. (9.6). Since Čech and sheaf cohomology are equivalent, we will compute whichever one is easiest in the case at hand.

C.1 Proof of $H^2(\widetilde{S}^2; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) = 0$

Consider the time reversal sphere $\widetilde{S}^2$ with its involution $\tau : (x, y, z) \mapsto (-x, -y, z)$. In the same way as we covered $\widetilde{S}^1$ by six contractible opens, we can cover $\widetilde{S}^2$ with eight contractible opens indexed by open hemispheres indexed by

$$\mathcal{A} = \{N_+, N_-, S_+, S_-, x_+, x_-, y_+, y_-\} \,.$$

The involution $\tau : \mathcal{A} \to \mathcal{A}$ acts again on the index set by changing a subindex from $+$ to $-$ and vice versa. The cochain groups are determined by specifying a generating set of indices for the overlaps, i.e. a set of indices such that any index is obtained by permutation

or action of τ. The cover itself is generated by $\{N_+, S_+, x_+, y_+\}$. The double overlaps are generated by

$$\begin{array}{llll} N_+N_- & S_+S_- & x_+y_+ & x_+y_- \\ N_+x_+ & N_+x_- & N_+y_+ & N_+y_- \\ S_+x_+ & S_+x_- & S_+y_+ & S_+y_- \,. \end{array}$$

The triple overlaps are generated by

$$\begin{array}{lll} N_+N_-x_+ & N_+x_+y_+ & S_+x_+y_+ \\ S_+S_-x_+ & N_+x_+y_- & S_+x_+y_- \\ N_+N_-y_+ & N_-x_+y_+ & S_-x_+y_+ \\ S_+S_-y_+ & N_-x_+y_- & S_-x_+y_- \,. \end{array}$$

The quadruple overlaps are generated by

$$\begin{array}{ll} N_+N_-x_+y_+ & N_+N_-x_+y_- \\ S_+S_-x_+y_+ & S_+S_-x_+y_- \,. \end{array}$$

There are no higher order overlaps. The equivariant Čech cochain complex for $\widetilde{S}^2$ thus takes the form

$$0 \longrightarrow \mathbb{Z}^4 \xrightarrow{\delta^0} \mathbb{Z}^{12} \xrightarrow{\delta^1} \mathbb{Z}^{12} \xrightarrow{\delta^2} \mathbb{Z}^4 \longrightarrow 0 \ .$$

Since the cover indexed by $\mathcal{A}$ satisfies the assumptions of Theorem 9.4, we have that

$$\check{H}^2(\widetilde{S}^2; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \cong \ker \delta^2 / \operatorname{im} \delta^1 \,.$$

The kernel of δ^2 is computed as follows. For a cochain $(a) \in \check{C}^2(\mathcal{U}, \widetilde{\mathbb{Z}})^{\mathbb{Z}_2}$, we can look at its image under δ^2:

$$\begin{aligned} (\delta^2 a)_{N_+N_-x_+y_+} &= a_{N_-x_+y_+} - a_{N_+x_+y_+} + a_{N_+N_-y_+} - a_{N_+N_-x_+} \\ (\delta^2 a)_{N_+N_-x_+y_-} &= a_{N_-x_+y_-} - a_{N_+x_+y_-} + a_{N_+N_-y_+} - a_{N_+N_-x_+} \\ (\delta^2 a)_{S_+S_-x_+y_+} &= a_{S_-x_+y_+} - a_{S_+x_+y_+} + a_{S_+S_-y_+} - a_{S_+S_-x_+} \\ (\delta^2 a)_{S_+S_-x_+y_-} &= a_{S_-x_+y_-} - a_{S_+x_+y_-} + a_{S_+S_-y_+} - a_{S_+S_-x_+} \,, \end{aligned}$$

where we have used equivariance and antisymmetry to write $a_{N_+N_-y_-} = a_{N_+N_-y_+}$ and $a_{S_+S_-y_-} = a_{S_+S_-y_+}$. The reason for writing it like this is because we want to use only the triple overlaps from our generating list. Now, $(a) \in \ker \delta^2$ if and only if $(\delta^2 a) = 0$, i.e. the left hand side of the above four equations must vanish. We can thus get rid of four components of (a) by expressing them in terms of the remaining eight using the kernel equations. Let's eliminate $N_-x_+y_+$, $N_-x_+y_-$, $S_-x_+y_+$ and $S_-x_+y_-$, which are the left most terms in each equation. The kernel of δ^2 is thus generated by the remaining eight independent components.

To prove that $\ker \delta^2 / \operatorname{im} \delta^1 = 0$ we have to show that δ^1 is surjective onto $\ker \delta^2$. This can be shown by listing out the eight independent triple overlaps of $\ker \delta^2$ and listing out the double overlaps in some specific order. Then, we write

$$\delta^1 : \check{C}^2 \cong \mathbb{Z}^{12} \to \mathbb{Z}^8 \cong \ker \delta^2$$

as a matrix with respect to the chosen ordering of the double and triple overlaps. This yields an 8×12 matrix. Row reducing this matrix δ^1 by only doing integer operations shows that δ^1 is surjective onto $\ker \delta^2$. We thus conclude that $\check{H}^2(\widetilde{S}^2; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) = 0$.

C.2 Proof of $H^2(\widetilde{S}^d; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) = 0$ for $d \geq 2$

To compute $H^2(\widetilde{S}^d; \mathbb{Z}_2, \widetilde{\mathbb{Z}})$ for $d > 2$ one could try to repeat the Čech cohomology computation performed in the previous section. The amount of sets we need to cover $\widetilde{S}^d$ grows linearly with d. However, the amount of overlaps we have to consider grows in exponential fashion. In fact, the amount of overlaps one has to consider in $d = 3$ is already quite unwieldy. There is a more efficient way to proceed now. We will set up a Mayer-Vietoris-type sequence to inductively prove that $H^2(\widetilde{S}^d; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) = 0$ for $d \geq 2$. The Čech computation in the case $d = 2$ serves as the base case for the induction argument.

In ordinary sheaf cohomology one has a Mayer-Vietoris sequence [42, Section 01E9]. Since the proof is purely categorical, it can be repeated verbatim in the equivariant case. The only subtlety is that one uses that in the ordinary case it is true that injectives sheaves are "flasque", meaning that any local section extends to a global section. It is not immediately obvious that this carries over to the equivariant setting, but it is nonetheless true [50] . This leads to the following statement of the Mayer-Vietoris sequence.

Lemma C.1. Let $\widetilde{X} = (X, \tau)$ be an involutive space and let U and V be two $\mathbb{Z}_2$-invariant opens such that $X = U \cup V$. For any $\mathbb{Z}_2$-sheaf $\widetilde{\mathcal{F}}$, there is a long exact sequence in cohomology

$$\ldots \to H^{n-1}(U \cap V) \longrightarrow H^n(X) \longrightarrow H^n(U) \oplus H^n(V) \longrightarrow H^n(U \cap V) \to \ldots ,$$

where $H^n(-)$ is shorthand for $H^n(-; \mathbb{Z}_2, \widetilde{\mathcal{F}})$.

We will now prove $H^2(\widetilde{S}^d; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) = 0$ for $d \geq 2$ by induction on d. The base case $d = 2$ has already been established in the previous section. Suppose then that $H^2(\widetilde{S}^d; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) = 0$ for some $d \geq 2$. The time reversal sphere $\widetilde{S}^{d+1}$ is obtained from $\widetilde{S}^d$ by attaching two cells $\{-\} \times D^{d+1}$ and $\{+\} \times D^{d+1}$ of dimension $d + 1$ as in Fig. 13. Let's take U to be the

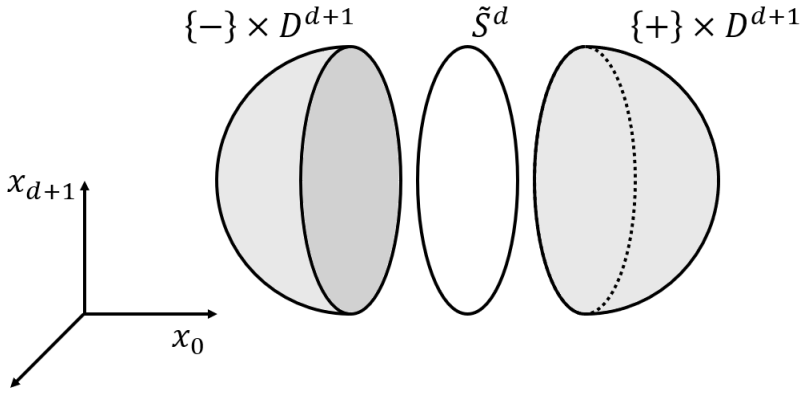

Figure 13: Attachment of free $d+1$-cells to $\widetilde{S}^{d+1}$. The coordinate system in the lower left corner indicates the x_0 direction, the x_{d+1} direction and the unlabelled axis represents the other directions.

interior of the two attached cells, i.e.

$$U = \{(x_0, x_1, \ldots, x_d, x_{d+1}) \in S^{d+1} : x_0 \neq 0\} .$$

The convention is that the $\mathbb{Z}_2$ action is $(x_0, x_1, \ldots, x_d, x_{d+1}) \mapsto (-x_0, -x_1, \ldots, -x_d, x_{d+1})$. We take V to be a fattened up version of the $\widetilde{S}^d$ that we attach to,

$$V = \{(x_0, x_1, \ldots, x_d, x_{d+1}) \in S^{d+1} : x_0 \in (-1/2, 1/2)\} .$$

The opens U and V are $\mathbb{Z}_2$-invariant and cover $\widetilde{S}^{d+1}$. From the Mayer-Vietoris sequence we obtain the exact sequence

$$H^1(U\cap V;\mathbb{Z}_2,\widetilde{\mathbb{Z}}) \longrightarrow H^2(\widetilde{S}^{d+1};\mathbb{Z}_2,\widetilde{\mathbb{Z}}) \longrightarrow H^2(U;\mathbb{Z}_2,\widetilde{\mathbb{Z}})\oplus H^2(V;\mathbb{Z}_2,\widetilde{\mathbb{Z}})\,. \tag{C.1}$$

The claim is that $H^1(U\cap V)$, $H^2(U)$ and $H^2(V)$ are all zero.

Let's start with $H^1(U\cap V;\mathbb{Z}_2,\widetilde{\mathbb{Z}})$. The set $U\cap V$ consists of two disjoint opens. There is one component with $x_0>0$ and one with $x_0<0$. The $\mathbb{Z}_2$-action maps points with $x_0>0$ to points with $x_0<0$ and vice versa. Therefore, an open cover $\mathcal{W}$ of the $x_0>0$ component can be extended by the $\mathbb{Z}_2$-action to a cover $\mathcal{W}\sqcup\tau\mathcal{W}$ of $U\cap V$. In fact, any $\mathbb{Z}_2$-cover of $U\cap V$ will be of this form. Furthermore, there is an isomorphism of cochain complexes

$$\check{C}^*(\mathcal{W}\sqcup\tau\mathcal{W};\widetilde{\mathbb{Z}})^{\mathbb{Z}_2}\cong\check{C}^*(\mathcal{W};\mathbb{Z})\,,$$

which from left to right is given by restriction and from right to left by equivariant extension. We conclude that

$$H^*(U\cap V;\mathbb{Z}_2,\widetilde{\mathbb{Z}})\cong H^*((U\cap V)_{x_0>0};\mathbb{Z})\,,$$

where the right hand side is ordinary sheaf cohomology with coefficients in the constant sheaf $\mathbb{Z}$. The $x_0>0$ component of $U\cap V$ is homotopy equivalent to S^d. Hence, by Theorem 9.2, i.e. homotopy invariance of sheaf cohomology with coefficients in a constant sheaf, it follows that

$$H^1(U\cap V;\mathbb{Z}_2,\widetilde{\mathbb{Z}})\cong H^1(S^d;\mathbb{Z})\cong H^1_{\text{sing}}(S^d;\mathbb{Z})=0\,,$$

where we have also invoked the equivalence of sheaf and singular cohomology for constant sheaves over paracompact spaces [51]. The same argument establishes that

$$H^2(U;\mathbb{Z}_2,\widetilde{\mathbb{Z}})\cong H^2(D^{d+1};\mathbb{Z})\cong H^2_{\text{sing}}(D^{d+1};\mathbb{Z})=0\,.$$

Finally, we have to show that $H^2(V;\mathbb{Z}_2,\widetilde{\mathbb{Z}})=0$. We can obtain a "good" $\mathbb{Z}_2$-cover of V by taking the "good" $\mathbb{Z}_2$-cover of $\widetilde{S}^d$ indexed by

$$\mathcal{A}=\{N_+,N_-,S_+,S_-,(x_1)_+,(x_1)_-,\ldots,(x_d)_+,(x_d)_-\}\,,$$

analogous the $\mathbb{Z}_2$-covers we have used for $\widetilde{S}^1$ and $\widetilde{S}^2$ before, and fattening it in the x_0 direction. In this way, we obtain a "good" cover of V indexed by $\mathcal{A}$. The Čech cochain complex associated to this cover of V with $\widetilde{\mathbb{Z}}$ coefficients is isomorphic to the corresponding one of $\widetilde{S}^d$, because their index sets have the same combinatorics. Hence, the Čech cohomologies of V and $\widetilde{S}^d$ agree. By the induction hypothesis, we conclude that

$$H^2(V;\mathbb{Z}_2,\widetilde{\mathbb{Z}})\cong H^2(\widetilde{S}^d;\mathbb{Z}_2,\widetilde{\mathbb{Z}})=0\,.$$

The portion of the Mayer-Vietoris sequence in Eq. (C.1) thus becomes

$$0\to H^2(\widetilde{S}^{d+1};\mathbb{Z}_2,\widetilde{\mathbb{Z}})\to 0.$$

Exactness of this sequence implies $H^2(\widetilde{S}^{d+1};\mathbb{Z}_2,\widetilde{\mathbb{Z}})=0$. This concludes the inductive proof that $H^2(\widetilde{S}^d;\mathbb{Z}_2,\widetilde{\mathbb{Z}})=0$ for $d\geq 2$.

C.3 Proof of $H^2(\widetilde{S}^d, (\widetilde{S}^d)^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \cong \mathbb{Z}_2$ for $d \geq 2$

We show that $\check{H}^2(\widetilde{S}^d, (\widetilde{S}^d)^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \cong \mathbb{Z}_2$ when $d \geq 2$. Because $\check{H}^2(\widetilde{S}^d; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) = 0$ for $d \geq 2$, it follows from the long exact sequence of the pair $(\widetilde{S}^d, (\widetilde{S}^d)^\tau)$ that

$$\check{H}^2(\widetilde{S}^d, (\widetilde{S}^d)^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \cong H^1((\widetilde{S}^d)^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}})/rH^1(\widetilde{S}^d; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \,,$$

where r denotes the map induced in cohomology by the inclusion $(\widetilde{S}^d)^\tau \to S^d$.

We first give the argument for $d = 2$. The degree one equivariant cohomology of the point with coefficients in $\mathbb{Z}_2$ has been computed to be $\mathbb{Z}_2$. Since $(\widetilde{S}^2)^\tau$ is just the disjoint union of two points, namely, the north and south pole within $\widetilde{S}^2$, it follows that

$$H^1((\widetilde{S}^2)^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2 \,.$$

Computing $H^1(\widetilde{S}^2; \mathbb{Z}_2, \widetilde{\mathbb{Z}})$ requires some more effort. We use the same "good" cover $\mathcal{U}$ indexed by $\mathcal{A}$ as in Appendix C.1. Recall that the cochain complex for this cover is

$$0 \longrightarrow \mathbb{Z}^4 \xrightarrow{\delta^0} \mathbb{Z}^{12} \xrightarrow{\delta^1} \mathbb{Z}^{12} \xrightarrow{\delta^2} \mathbb{Z}^4 \longrightarrow 0 \quad .$$

To shorten the notation, we will represent a cochain $(b) \in \check{C}^1(\mathcal{U}, \widetilde{\mathbb{Z}})^{\mathbb{Z}_2}$ in terms of its indices only. For example, instead of writing $b_{N_+N_-}$, we simply write N_+N_-. We claim that

$$\begin{aligned}\ker \delta^1 = \{(N_+N_-, S_+S_-, N_+x_+, &S_+x_+, N_+y_+, S_+, y_+) : \\ &N_+N_- = S_+S_-,\, N_+x_+ = N_+y_+,\, S_+x_+ = S_+y_+\} \,.\end{aligned}$$

This follows from the twelve kernel equations

$$\begin{array}{lll} N_+x_- = N_+x_+ - N_+N_- & x_+y_+ = N_+y_+ - N_+x_+ & x_+y_+ = S_+y_+ - S_+x_+ \\ N_+y_- = N_+y_+ - N_+N_- & x_+y_- = N_+y_- - N_+x_+ & x_+y_- = S_+y_- - S_+x_+ \\ S_+x_- = S_+x_+ - S_+S_- & x_+y_+ = N_-y_+ - N_-x_+ & x_+y_+ = S_-y_+ - S_-x_+ \\ S_+y_- = S_+y_+ - S_+S_- & x_+y_- = N_-y_- - N_-x_+ & x_+y_- = S_-y_- - S_-x_+ \,. \end{array}$$

From the first column, we see that we can express N_+x_-, N_+y_-, S_+x_- and S_+y_- in terms of indices we claim to be in the kernel. This already reduces the kernel from $\mathbb{Z}^{12}$ to $\mathbb{Z}^8$. We can also completely get rid of x_+y_+ and x_+y_-, since these too can be expressed in terms of the indices we claim to be in $\ker \delta^1$. We now prove the relations between the variables of $\ker \delta^1$. From the second and third column of equations above, it follows that

$$N_+y_+ - N_+x_+ = S_+y_+ - S_+x_+ \,, \tag{C.2}$$

$$N_+y_- - N_+x_+ = S_+y_- - S_+x_+ \,. \tag{C.3}$$

Using the first column, we can express Eq. (C.3) as

$$N_+y_+ - N_+N_- - N_+x_+ = S_+y_+ - S_+S_- - S_+x_+ \,.$$

Invoking Eq. (C.2), it follows that

$$N_+N_- = S_+S_- \,.$$

From the second and third column we also get

$$N_+y_+ - N_+x_+ = N_-y_+ - N_-x_+ \,.$$

By equivariance and the first column, the right hand side becomes

$$N_-y_+ - N_-x_+ = -N_+y_- + N_+x_- = -N_+y_+ + N_+N_- + N_+x_+ - N_+N_- = -(N_+y_+ - N_+x_+)\,.$$

Hence, $N_+y_+ = N_+x_+$ and similarly we prove $S_+y_+ = S_+x_+$. This establishes the claimed form of the kernel. The image of δ^0 is fully determined by

$$(\delta^0 a)_{N_+N_-} = -2N_+\,, \quad (\delta^0 a)_{N_+x_+} = x_+ - N_+ \quad \text{and} \quad (\delta^0 a)_{S_+y_+} = y_+ - S_+\,.$$

From the N_+N_- and S_+S_- entries, we get a $\mathbb{Z}/2$ and all other entries cancel when taking the quotient because we can use x_+ to cancel N_+x_+ and y_+ to cancel S_+y_+. Therefore,

$$\begin{aligned} H^1(\widetilde{S}^2; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) &= \ker \delta^1 / \operatorname{im}\, \delta^0 \\ &\cong \operatorname{diag}(\mathbb{Z} \oplus \mathbb{Z})^3 / (\operatorname{diag}(2\mathbb{Z} \oplus 2\mathbb{Z}) \oplus \operatorname{diag}(\mathbb{Z} \oplus \mathbb{Z})^2) \\ &\cong \operatorname{diag}(\mathbb{Z}_2 \oplus \mathbb{Z}_2)\,, \end{aligned}$$

where the remaining direct sum pertains to the N_+N_- and S_+S_- components.

The relative cohomology $H^1(S^2, (S^2)^\tau; G, \widetilde{\mathbb{Z}})$ can now be computed. Recall that

$$H^1((\widetilde{S}^2)^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \cong \mathbb{Z}/2 \oplus \mathbb{Z}/2\,,$$

where the first $\mathbb{Z}/2$ corresponds to N_+N_- and the second to S_+S_-. The restriction of $H^1(\widetilde{S}^2; \mathbb{Z}_2, \widetilde{\mathbb{Z}})$ to the fixed points is precisely the diagonal in $\mathbb{Z}/2 \oplus \mathbb{Z}/2$. Hence,

$$\begin{aligned} H^2(\widetilde{S}^2, (\widetilde{S}^2)^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) &\cong H^1((\widetilde{S}^2)^\tau; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) / r H^1(\widetilde{S}^2; \mathbb{Z}_2, \widetilde{\mathbb{Z}}) \\ &\cong (\mathbb{Z}/2 \oplus \mathbb{Z}/2) / \operatorname{diag}(\mathbb{Z}/2 \oplus \mathbb{Z}/2) \\ &\cong \mathbb{Z}/2\,. \end{aligned}$$

This establishes the result for $d = 2$. However, the computation for S^2 can be immediately extended to S^d for $d > 2$. Taking the same cover, with more directions x_i's, we get by considering just x_1 and x_2 already the relation $N_+N_- = S_+S_-$ from exactly the same equations as above. This then deals already with all things we need, since we are only interested in the restriction to the fixed points.

This combinatoric argument for the $\mathbb{Z}_2$ valuedness very much has the same structure as the homotopy argument given in Section 8.3. For the sphere $\widetilde{S}^2$ we argue the $\mathbb{Z}_2$ valuedness from the fact that behaviour at the north pole is linked to behaviour at the south pole. In the homotopy argument, the relation between north and south pole is reflected by the fact that global continuous equivariant sections need to have the same signs at the north and south pole. In the Čech cohomology argument, the relation is that for an equivariant cochain (b) in the kernel of δ^1 it must be that $b_{N_+N_-} = b_{S_+S_-}$. The argument for higher dimensional spheres $\widetilde{S}^d$ follows by just restricting to a copy of $\widetilde{S}^2$.

References

[1] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*. Holt-Saunders, 1976.

[2] J. R. Hook and H. E. Hall, *Solid state physics*. John Wiley & Sons, 2013.

[3] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, "Quantum spin hall effect and topological phase transition in hgte quantum wells," *science*, vol. 314, no. 5806, pp. 1757–1761, 2006.

[4] X.-L. Qi, T. L. Hughes, and S.-C. Zhang, "Topological field theory of time-reversal invariant insulators," *Phys. Rev. B*, vol. 78, p. 195424, Nov 2008.

[5] X.-L. Qi and S.-C. Zhang, "Topological insulators and superconductors," *Reviews of Modern Physics*, vol. 83, no. 4, p. 1057, 2011.

[6] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, "Quantized hall conductance in a two-dimensional periodic potential," *Physical review letters*, vol. 49, no. 6, p. 405, 1982.

[7] G. E. Volovik, *The Universe in a Helium Droplet*. Oxford University Press, 02 2009.

[8] A. M. Essin and V. Gurarie, "Bulk-boundary correspondence of topological insulators from their respective green's functions," *Phys. Rev. B*, vol. 84, p. 125132, Sep 2011.

[9] J. Fröhlich, "Chiral anomaly, topological field theory, and novel states of matter," *Reviews in Mathematical Physics*, vol. 30, no. 06, p. 1840007, 2018.

[10] J. C. Budich and B. Trauzettel, "From the adiabatic theorem of quantum mechanics to topological states of matter," 2013.

[11] G. Panati, "Triviality of bloch and bloch–dirac bundles," *Annales Henri Poincaré*, vol. 8, pp. 995–1011, jul 2007.

[12] C. L. Kane and E. J. Mele, "Z_2 topological order and the quantum spin hall effect," *Phys. Rev. Lett.*, vol. 95, p. 146802, Sep 2005.

[13] C. L. Kane and E. J. Mele, "Quantum spin hall effect in graphene," *Phys. Rev. Lett.*, vol. 95, p. 226801, Nov 2005.

[14] L. Fu and C. L. Kane, "Time reversal polarization and a Z_2 adiabatic spin pump," *Phys. Rev. B*, vol. 74, p. 195312, Nov 2006.

[15] L. Fu and C. L. Kane, "Topological insulators with inversion symmetry," *Phys. Rev. B*, vol. 76, p. 045302, Jul 2007.

[16] S. Ryu, A. P. Schnyder, A. Furusaki, and A. W. Ludwig, "Topological insulators and superconductors: tenfold way and dimensional hierarchy," *New Journal of Physics*, vol. 12, no. 6, p. 065010, 2010.

[17] G. De Nittis and K. Gomi, "Classification of "real" bloch-bundles: topological quantum systems of type ai," *Journal of Geometry and Physics*, vol. 86, pp. 303–338, 2014.

[18] G. De Nittis and K. Gomi, "Classification of "quaternionic" bloch-bundles," *Communications in Mathematical Physics*, vol. 339, no. 1, pp. 1–55, 2015.

[19] M. F. Atiyah, "K-theory and reality," *The Quarterly Journal of Mathematics*, vol. 17, no. 1, pp. 367–386, 1966.

[20] M. Sigrist, "Solid state theory," 2014.

[21] C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, "Quantum anomalous hall effect in $Hg_{1-y}Mn_yTe$ quantum wells," *Physical Review Letters*, vol. 101, oct 2008.

[22] "Unitary group in nlab." Available at `https://ncatlab.org/nlab/show/unitary+group#HomotopyGroups` (2023/07/07).

[23] R. Bott, "The stable homotopy of the classical groups," *Annals of Mathematics*, vol. 70, no. 2, pp. 313–337, 1959.

[24] M. Nakahara, *Geometry, topology and physics*. CRC press, 2003.

[25] L. Tu, *Differential Geometry: Connections, Curvature, and Characteristic Classes*. Graduate Texts in Mathematics, Springer International Publishing, 2017.

[26] J. M. Lee, *Smooth manifolds*. Springer, 2012.

[27] G. E. Bredon, *Topology and geometry*, vol. 139. Springer Science & Business Media, 2013.

[28] T. Kato, "On the adiabatic theorem of quantum mechanics," *Journal of the Physical Society of Japan*, vol. 5, no. 6, pp. 435–439, 1950.

[29] B. Simon, "Holonomy, the quantum adiabatic theorem, and berry's phase," *Physical Review Letters*, vol. 51, no. 24, p. 2167, 1983.

[30] M. V. Berry, "Quantal phase factors accompanying adiabatic changes," *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, vol. 392, no. 1802, pp. 45–57, 1984.

[31] B. Simon, "Tosio kato's work on non–relativistic quantum mechanics," *arXiv preprint arXiv:1711.00528*, 2017.

[32] H. T. Stoof, K. B. Gubbels, and D. Dickerscheid, *Ultracold quantum fields*. Springer, 2009.

[33] A. Hatcher, *Vector Bundles and K-Theory*. 2017.

[34] G. Heuts and S. Sagave, *Lecture Notes for Algebraic Topology 2*. 2021.

[35] J. Domingos, "Time reversal in classical and quantum mechanics," *International Journal of Theoretical Physics*, vol. 18, pp. 213–230, 1979.

[36] A. Zee, *Quantum field theory in a nutshell*, vol. 7. Princeton university press, 2010.

[37] H. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, "Topological insulators at room temperature," *arXiv preprint arXiv:0812.1622*, 2008.

[38] M. Furuta, Y. Kametani, H. Matsue, and N. Minami, "Stable-homotopy seiberg-witten invariants and pin bordisms," *preprint*, 2000.

[39] S. Eilenberg and N. Steenrod, *Foundations of algebraic topology*, vol. 2193. Princeton University Press, 2015.

[40] J. J. Rotman and J. J. Rotman, *An introduction to homological algebra*, vol. 2. Springer, 2009.

[41] B. Iversen, *Cohomology of sheaves*. Springer Science & Business Media, 2012.

[42] T. Stacks project authors, "The stacks project." `https://stacks.math.columbia.edu`, 2023.

[43] R. O. Wells, *Differential Analysis on Complex Manifolds*. Graduate Texts in Mathematics, Springer New York, NY, 2010.

[44] V. A. Krasnov, "Characteristic classes of vector bundles on a real algebraic variety," *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, vol. 55, no. 4, pp. 716–746, 1991.

[45] A. Grothendieck, "Sur quelques points d'algèbre homologique," *Tohoku Mathematical Journal, Second Series*, vol. 9, no. 2, pp. 119–183, 1957.

[46] R. Bott, L. W. Tu, *et al.*, *Differential forms in algebraic topology*, vol. 82. Springer, 1982.

[47] G. De Nittis and K. Gomi, "The cohomological nature of the fu-kane-mele invariant," 2016.

[48] M. R. Zirnbauer, "Symmetry classes in random matrix theory," 2004.

[49] J.-P. Serre, "Faisceaux algébriques cohérents," *Annals of Mathematics*, pp. 197–278, 1955.

[50] A. Stieglitz, "Equivariant sheaf cohomology," *manuscripta mathematica*, vol. 26, no. 1-2, pp. 201–221, 1978.

[51] Y. Sella, "Comparison of sheaf cohomology and singular cohomology," 2016.