Topological and geometrical aspects of band theory

J Cayssol and J N Fuchs

1 Univ. Bordeaux, CNRS, LOMA, UMR 5798, F-33405 Talence, France
2 Sorbonne Université, CNRS, Laboratoire de Physique Théorique de la Matière Condensée, LPTMC, F-75005 Paris, France
3 Laboratoire de Physique des Solides, Université Paris Saclay, CNRS, F-91405 Orsay, France

Abstract

This paper provides a pedagogical introduction to recent developments in geometrical and topological band theory following the discovery of graphene and topological insulators. Amusingly, many of these developments have a connection to contributions in high-energy physics by Dirac. The review starts by a presentation of the Dirac magnetic monopole, goes on with the Berry phase in a two-level system and the geometrical/topological band theory for Bloch electrons in crystals. Next, specific examples of tight-binding models giving rise to lattice versions of the Dirac equation in various space dimension are presented: in 1D (Su–Schrieffer–Heeger (SSH) and Rice–Mele models), 2D (graphene, boron nitride, Haldane model) and 3D (Weyl semi-metals). The focus is on topological insulators and topological semi-metals. The latter have a Fermi surface that is characterized as a topological defect. For topological insulators, the two alternative view points of twisted fiber bundles and of topological textures are developed. The minimal mathematical background in topology (essentially on homotopy groups and fiber bundles) is provided when needed. Topics rarely reviewed include: periodic versus canonical Bloch Hamiltonian (basis I/II issue), Zak versus Berry phase, the vanishing electric polarization of the SSH model and Dirac insulators.
Contents

1. Introduction
2. Dirac magnetic monopole in real space
   2.1. Obstruction, Dirac string and quantization argument
   2.2. Wu and Yang’s construction with two patches
   2.3. Dirac monopole as a fiber bundle
3. Emergent Berry monopole for a two-level system
   3.1. Two-level system
   3.2. Gauge freedom and Berry connections
   3.3. Berry curvature and Chern number
   3.4. Berry flux monopole in parameter space
4. Geometrical and topological band theory
   4.1. Introduction to fiber bundles
   4.2. Band structure and Bloch fiber bundle
      4.2.1. Bloch Hamiltonian and cell-periodic Bloch states
      4.2.2. Berry curvature and quantum metric
      4.2.3. Berry gauge transformation
      4.2.4. Berry curvature and virtual transitions
      4.2.5. Periodic versus canonical Bloch Hamiltonian (basis I/II issue)
   4.2.6. Conclusion
   4.3. Geometric phases and consequences
      4.3.1. Berry phase
      4.3.2. Zak phase
      4.3.3. Wannier functions
      4.3.4. Chern number
   4.4. Semi-classical equations of motion of a Bloch electron
      4.4.1. Standard equations of motion
      4.4.2. Equations of motion including Berry phase effects
      4.4.3. Quantized Hall conductivity of a filled band
   4.4.4. Berry phase effects in doped semiconductors and metals
   4.5. Non-abelian Berry phases
   4.6. Conclusion: Berry phase effects in solids, gauge fields and fiber bundles
5. Lattice Dirac fermions in 1D
   5.1. Su–Schrieffer–Heeger model
   5.2. Band structure and Berryology of the SSH model
      5.2.1. Bloch Hamiltonians (two bases), band structure and spinors
      5.2.2. Symmetries
      5.2.3. Winding number and edge modes
      5.2.4. Zak phase, position operator, Wannier center and electric polarization
   5.3. Linearization and effective Dirac Hamiltonian
   5.4. Jackiw–Rebbi mechanism and charge fractionalization in Dirac insulators
   5.5. Inversion-protected topological insulator
      5.5.1. Rice–Mele model with inversion symmetry
      5.5.2. Bulk electric polarization
      5.5.3. Which phase is trivial and which is topological?
      5.5.4. Topological phase transition
   5.5.5. Conclusion
6. Dirac fermions on 2D honeycomb lattice
   6.1. Pristine graphene and massless Dirac fermion
      6.1.1. Hamiltonian and band structure in the whole BZ
      6.1.2. Effective Hamiltonian near the Dirac points and symmetries
      6.1.3. Protection of the Dirac points by symmetry
   6.2. Hexagonal boron nitride and massive Dirac fermions
   6.3. Haldane model of graphene and topology
   6.4. Chern number of two-band models in 2D
      6.4.1. Chern number as a wrapping number
      6.4.2. Practical evaluation of the Chern number with a degree formula
   6.5. Topological insulators: a bigger picture
7.1. Bulk quantized electromagnetic response and gapless boundary states 50
  7.1.1. Intrinsic topological insulators 51
  7.1.2. Symmetry-protected topological insulators 52
  7.1.3. Topological crystalline insulators 53
  7.1.4. Conclusion: first definition of a topological insulator 53
7.2. Obstruction to exponentially-localized and symmetric Wannier functions 53
  7.2.1. Intrinsic topological insulators 54
  7.2.2. Symmetry-protected topological insulators 54
  7.2.3. Topological crystalline insulators 54
  7.2.4. Conclusion: second definition of a topological insulator 55
8. Topological metals: Fermi surface as a topological defect 55
  8.1. Introduction to homotopy groups for topological defects and textures 56
  8.2. 3D Weyl semi-metal as topological defect 57
  8.3. 2D Chern insulator as topological texture 59
  8.4. 2D Dirac point as symmetry-protected topological defect 60
  8.5. Classification of topological metals 60
9. Conclusion 60
Appendix. Topological insulator versus topological order 61
References 63
1. Introduction

Band theory was created in the 1930s right after the invention of quantum mechanics [1]. It relies mainly on Bloch's theorem, which rules the behavior of an electron in the periodic potential created by ions, and on Fermi–Dirac statistics, which governs the filling of energy bands by electrons. The coronation of band theory was the classification of crystals by Wilson [2] into insulators and metals depending on whether the band structure contains or not a partially filled band (i.e. a Fermi surface). The equations describing the semiclassical motion of a Bloch electron restricted to a single band were obtained by Bloch [3], Peierls [4], Jones and Zener [5]. These equations have a form similar to that for a classical non-relativistic particle in the vacuum, except for the velocity which is replaced by the band’s group velocity. Even at the quantum level, electrons in a crystal were believed to behave almost like electrons in the vacuum upon replacing the dispersion relation \( E = \frac{(\hbar k)^2}{2m} \) by the band dispersion \( E_n(k) \). This initial version of band theory could account for the electronic behavior of very many crystals [6].

However, in the 1940s, 1950s and 1960s researchers started to realize that there may be more to band theory than simply individual energy bands. We should mention the work of pioneers such as Adams, Blount, Kohn, Luttinger, Roth, Slater, Wannier and others (see the review by Blount [7]), who understood the importance of inter-band effects. For example, Karplus and Luttinger realized that in some materials, corrections to the group velocity could appear in the form of an anomalous transverse velocity [8] which may explain the anomalous Hall effect. Also Kohn found that the effective Hamiltonian for a single band in a magnetic field was not only given by the band dispersion \( E_n(k) \) but was shifted by an orbital magnetic moment (in addition to the Zeeman effect) [9]. But the situation remained obscure and, apart from band theory aficionados, nobody really paid attention. In hindsight, reading the review by Blount [7], one recognizes already a lot of the ‘modern’ concepts of geometrical band theory such as ‘Berry’ connection, curvature, virtual inter-band transitions, analogy to electromagnetism but in reciprocal space, relation to the Dirac equation, etc., that we will encounter in this review. However, topological considerations were absent.

Everything changed with the experimental discovery of the integer quantum Hall effect (IQHE) [10] and its subsequent understanding as a topological effect by Thouless et al [11, 12] building on concepts that had emerged in the context of the Hofstadter butterfly [13]. It was soon recognized that this was a particular instance of a Berry phase effect [14, 15]. It took some more years for Haldane [16] to clearly spell out that the QHE (now known as the quantum anomalous Hall effect (QAHE)) could be understood as a pure band theory effect provided time-reversal symmetry was broken, but that one could dispense with Landau levels (no uniform magnetic field) and maintain the periodicity of the crystal. In particular, this showed that a filled band could indeed conduct electricity despite what had been written in solid-state physics textbooks for 50 years. Simultaneously, Volovik proposed a similar kind of topological effect in the framework of Bogoliubov-de Gennes (mean-field) description of superconductors and applied it to a film of superfluid helium 3 [17]. It also involved coupling between bands but the origin of bands is in particle-hole coupling via the superfluid pairing and not in Bloch’s theorem. This constitutes a first example of what is now called a topological superconductor [18]. Around the same time, Zak realized that another type of Berry phase—an open-path Berry phase along a non-contractible loop—could be defined in the Brillouin zone (BZ) using the torus geometry [19]. A few years later, King-Smith, Vanderbilt and Resta understood that the Zak phase was related to the position operator and the key to settling the problematic issue of the proper definition of a bulk electric polarization for a crystal [20–22]. In addition to this ‘modern theory of electric polarization’ [23], a parallel ‘modern theory of orbital magnetization’ was developed and is reviewed in [24, 25]. The semi-classical equations of motion for a Bloch electron restricted to a given band, and modified by Berry phase terms, were also obtained in final form by Chang, Sundaram and Niu [26, 27] using a wave-packet approach, see [24] for review. In a few years, many long-standing and annoying problems of solid-state physics were solved by realizing that certain measurable quantities explicitly depend on the phase of the Bloch wave functions.

The next major step was taken by Kane and Mele, who realized that topology could also be present in systems that do not break time-reversal symmetry [28, 29]. The original proposal was with graphene and could not be realized because of a too small intrinsic spin–orbit coupling (SOC) of carbon but Bernevig, Hughes and Zhang proposed another system—a HgTe/CdTe quantum well—in which one could obtain a quantum spin Hall insulator (QSHI) [30]. This was realized experimentally in the group of Molenkamp [31]. This first example of a symmetry-protected topological (SPT) insulator in two dimensions was soon followed by its generalization to three dimensions (something not possible for the IQHE) [32–34] and the subsequent experimental discovery, see [35] for review.

The proposal of Kane and Mele was concomitant with a major experimental discovery, that of graphene [36, 37]. This two-dimensional honeycomb carbon crystal has low-energy electrons that obey a massless Dirac equation rather than an effective single-band Schrödinger equation [38, 39]. This is another
extension of band theory. A system that is neither a metal nor an insulator—it is gapless but the density of states vanishes at the Fermi surface reduced to two points—and whose description involves two bands that are strongly coupled. The vicinity of each contact point resembles a diabolo [40] and is now called a Dirac cone. One characterization of these Dirac fermions is that they carry a $\pi$ Berry phase. Graphene is now considered as an example of a (symmetry-protected) topological semi-metal. The culmination in this modern version of band theory—that may be summarized as Berry phase effects + graphene + topological insulators—was in the periodic table (or ten-fold way) classification of topological insulators and superconductors by Schnyder et al [41] and by Kitaev [42]. This classification was extended in several directions including topological semi-metals [43, 44], of which Volovik should be mentioned as an early pioneer [45].

Amusingly, in many of the above modern developments in band theory, one may see shadows of contributions by Dirac in high-energy physics. A first instance is the simultaneous invention of anti-matter (the positron) by Dirac and its solid-state version (the hole) by Peierls (this story is beautifully told in [1]). Obviously the Dirac equation [46] plays an important role as the simplest Hamiltonian describing the coupling between two (or four) bands. It was invented for relativistic electrons in the 3D vacuum but now serves to describe various crystals in the long-wavelength limit in 1D, 2D and 3D [47–49]. The most prominent example is the honeycomb lattice of graphene, which gives rise at long wavelength to the emergence of two massless Dirac equations in 2D. Note also that in the review by Blount [7], it was already recognized that the Dirac equation was some sort of model for band-coupling effects. At that time, it was mainly used as an effective description of bulk 3D bismuth. Another contribution of Dirac that has descendants in solid-state physics is the magnetic monopole [50]. It may be seen as a forerunner of the Aharonov–Bohm phase and more generally of Berry phases. Although elusive as magnetic charge in real space, the Dirac monopole actually exists in reciprocal space as a source of Berry flux and is related to band contact points. The well-known quantization of the magnetic strength of the Dirac monopole has a counterpart in the Chern numbers characterizing the bands. Later, Wu and Yang have shown that the Dirac monopole has topological significance and is related to the mathematical notion of fiber bundles [51].

In the present paper, we provide a pedagogical review of modern band theory focusing on geometrical and topological effects. These effects are all due to coupling between bands. The latter modify the effective description of an electron restricted to a given band, by the appearance of an emergent gauge field (also known as the Berry connection), that takes into account the possibility of virtual transitions to other bands. These are geometrical effects in band theory, i.e. physical effects in solid-state physics that do not only depend on the energy bands in the absence of external fields but also involve the cell-periodic Bloch eigenfunctions. In addition, and because the BZ is a compact manifold (a torus in $D$ dimensions), some of these geometrical effects turn topological. For geometrical band theory and Berry phase effects in solids, we recommend the reviews by Xiao et al [24], Resta [52, 53] and the book by Vanderbilt [54]. On the topic of topological insulators, see [35, 55–59] and the books by Bernevig [18] and by Asbóth et al [60]. On the subject of topological semi-metals and the classification of Fermi surfaces as topological defect, see the book by Volovik [45]. For topological superconductors, we recommend the chapters written by Hughes in [18]. For the extension of these ideas to cold atoms or to photonics, see [61, 62].

The structure of our review is the following. In section 2, we study the Dirac magnetic monopole. Then in section 3, we consider a quantum two-level system (TLS) and show the appearance of a Berry phase, i.e. and emergent gauge structure in parameter space. Next, we turn to periodic crystals in section 4 and present geometrical and topological band theory. In particular, we give an introduction to the mathematical notion of fiber bundles. In section 5, we review the physics of one-dimensional non-interacting electrons on dimerized (or diatomic) chains, using the Su–Schrieffer–Heeger (SSH) and Rice–Mele (RM) models as examples. The following section 6 deals with two-dimensional band structures on the honeycomb lattice; we discuss the geometrical and topological aspects of graphene (section 6.1), boron nitride (section 6.2) and the Haldane model of a Chern insulator (section 6.3). In section 7, we sketch a bigger picture of the notion of topological insulators. In section 8, we consider topological semi-metals (especially 3D Weyl semi-metals) in which the Fermi surface is treated as a topological defect and describe a connection between topological metals and insulators via the relation between topological defects and textures. Here, the mathematical notion of homotopy groups is outlined. In the general conclusion (section 9), we summarize the most important points and in an appendix, we make a distinction between topological insulators (covered in this review) and topological order (TO) (not covered).

2. Dirac magnetic monopole in real space

In 1931, Dirac investigated the compatibility of quantum mechanics with the presence of point-like magnetic charges [50]. At the level of classical electrodynamics, such a magnetic monopole is forbidden as a point
charge but may exist as the termination of a semi-infinite solenoid; it is not only a point-like singularity in the magnetic field, but it also implies a line singularity in the vector potential $\mathbf{A}(r)$ along the solenoid. Those latter singularities, called Dirac strings, are semi-infinite lines emanating from the monopole and extending to infinity. At first sight, such extended singularities in $\mathbf{A}(r)$ look pretty harmful to quantum mechanics since $\mathbf{A}(r)$ enters directly the Schrödinger equation. But Dirac realized that the framework of quantum mechanics can be kept perfectly coherent provided that the wave functions vanish along such Dirac strings (the Dirac veto). This condition leads to a relation between electrical charge and the monopole strength. We start by presenting a quick argument for the Dirac quantization condition. Then we present a derivation or reinterpretation of this quantization due to Wu and Yang in the seventies [51], that allows one to avoid the concept of Dirac strings (and the related Dirac veto) and is important to realize the global topological nature of the Dirac monopole. As a general reference on the Dirac monopole, we recommend a chapter in the book by Ryder [63].

### 2.1. Obstruction, Dirac string and quantization argument

A magnetic monopole of strength $g$, located at the origin of real space $\mathbb{R}^3$, would produce a radial magnetic field given by:

$$
\mathbf{B} = \frac{g}{r^2} \mathbf{e}_r,
$$

which is the solution of:

$$
\nabla \cdot \mathbf{B} = 4\pi g \delta(r).
$$

The total magnetic flux piercing a closed surface (e.g. a sphere $S^2$) surrounding the origin is therefore:

$$
\oint_{S^2} \mathbf{B} \cdot d\mathbf{S} = 4\pi g.
$$

It is natural to search for a vector potential $\mathbf{A}$ associated to the monopole radial magnetic field equation (1) as $\nabla \cdot \mathbf{B} = 0$ in $\mathbb{R}^3 - \{0\} \simeq S^2 \times \mathbb{R}^+ \sim S^2$. It turns out that it is impossible to find a single regular/smooth vector potential expression (electromagnetic gauge) covering the whole space $\mathbb{R}^3 - \{0\}$ [64]. To understand this fact, let us assume the existence of such a smooth gauge and show some contradiction. If $\mathbf{A}$ were smooth on the whole sphere $S^2$, for any closed line $C$ on $S^2$, separating the sphere in two regions, it would be possible to apply Stokes’ theorem for each region and get:

$$
\oint_{C} \mathbf{B} \cdot d\mathbf{l} = \oint_{C} d\mathbf{l} \cdot \mathbf{A} = 0.
$$

Then the flux of $\mathbf{B} = \nabla \times \mathbf{A}$ through $S^2$ (or any closed surface) would always be zero in contradiction with equation (3). The conclusion is therefore that a magnetic monopole is only possible if there is an obstruction to finding such a smooth global gauge for the vector potential. Therefore there should be at least a point on $S^2$ where the potential $\mathbf{A}$ is singular. By connecting this singularity for continuously varying radius of the sphere, one gets a line singularity called a Dirac string, which is a not necessarily straight, starts from the monopole and extends to infinity (see figure 1).

The Dirac string can be used to partially (in a restricted space) bypass the obstruction discussed above. The idea is to see the monopole as the free end of an infinite line of magnetic dipoles (or a semi-infinite thin solenoid), the other end of the line being sent to infinity. This is equivalent to attaching a thin solenoid to the magnetic pole. The role of this solenoid is to feed some flux ($-g$) into the sphere to compensate the radial flux $g$ created by the monopole itself. In this way the total magnetic flux through the sphere is exactly zero and the total magnetic field (of the system pole + solenoid) can be written as the curl of a vector potential. Then the magnetic field of the monopole alone reads $\nabla \times \mathbf{A} - \mathbf{B}_{\text{sol}}$, where $\mathbf{B}_{\text{sol}}$ is the highly singular magnetic field located inside the thin solenoid.

By solving the problem of a scalar wave function in the background of the monopole, Dirac has shown that the wave function must vanish along the Dirac string (i.e. a nodal line), which implies a quantization condition. A heuristic argument to get this quantization condition is to anticipate that the obstruction originates from the finite magnetic flux $4\pi g$. One may suspect that when this magnetic flux is a multiple of the quantum of flux $\hbar/e$, its effect is undetectable, and the phases of the wave functions can be defined in a consistent way. This lead to:

$$
4\pi g = \frac{n\hbar}{e} \quad \text{with} \quad n \in \mathbb{Z},
$$

which is Dirac’s quantization condition.
Figure 1. The Dirac monopole is represented as a red point. (a) The monopole alone creates a radial magnetic field (blue arrows) with a total flux $4\pi g$ through the sphere $S^2$ oriented outward if $g > 0$. This magnetic field cannot be written as the curl of a vector potential. (b) The magnetic monopole plus an attached semi-infinite thin solenoid carrying a magnetic flux (solenoid in red) $-g$ that compensates the flux from the monopole itself. The magnetic field of this composite system ‘monopole+solenoid’ can be expressed as the curl of a vector potential $A(n)$. The latter is regular everywhere on the north cap except along the negative $z$ axis. (c) The location of the radial Dirac string/solenoid is changed (with respect to (b)) and therefore the magnetic field of the ‘monopole+solenoid’ is the curl of the vector potential $A(s)$ defined everywhere except along the north radial line (positive $z$ axis). The analytic expressions of $A(n)$ and $A(s)$ are given in the text, see equations (8) and (10).

2.2. Wu and Yang’s construction with two patches

In 1975, Wu and Yang proposed an alternative way to handle the Dirac monopole problem which circumvents the use of the Dirac strings [51]. The idea consists in working with two-distinct gauges, $A^{(n)}$ and $A^{(s)}$, each of them well-defined in a restricted subset of space, respectively $R_n$ and $R_s$ (see figure 1). The two regions $R_n$ and $R_s$ are such that their union covers all space, and that their intersection is not empty. Using this procedure, the magnetic flux:

$$\oint_{S^2} B \cdot dS = \oint_C dl \cdot (A^{(n)} - A^{(s)}) \neq 0,$$

is not necessary zero because $A^{(n)}$ and $A^{(s)}$ are distinct (they differ by the gradient of a scalar function).

Let us now find expressions for the different vector potentials $A^{(n,s)}$ using a specific path $C = C_{\theta}$, which is the parallel-type circle defined by the constant value $\theta$ of the polar angle, separating the sphere in a north cap $N$ and a south cap $S$. Applying Stokes’ theorem to the north cap:

$$\int_N B \cdot dSe = 2\pi(1 - \cos \theta)g = \int A^{(n)} \cdot dl = 2\pi r \sin \theta A^{(n)} \cdot e_\phi,$$

yields the following expression for the electromagnetic vector potential:

$$A^{(n)} = \frac{g(1 - \cos \theta)}{r \sin \theta} e_\phi,$$

which is singular when $\theta = \pi$. The negative $Oz^-$ axis corresponds to the Dirac string singularity. The potential $A^{(n)}$ is well-defined everywhere else, namely in the set $R_n = \mathbb{R}^3 - (Oz^-)$. Note that any closed surface within $R_n$ does not contain the monopole, so the total flux is zero.

Applying similarly Stokes’ theorem to the south cap:

$$\int_S B \cdot dSe = 2\pi(1 + \cos \theta)g = -\int A^{(s)} \cdot dl = -2\pi r \sin \theta A^{(s)} \cdot e_\phi,$$

provides another expression for the vector potential:

$$A^{(s)} = -\frac{g(1 + \cos \theta)}{r \sin \theta} e_\phi,$$

which is singular along the positive $Oz$ axis ($\theta \to 0$).
Along any parallel $C_\phi$, both vector potentials are well-defined and we can therefore compare them. The difference of the two vector potentials is given by a gradient:

$$A^{(n)} - A^{(s)} = \frac{2g}{r \sin \theta} e^\varphi \nabla (2g \varphi). \quad (11)$$

Up to this point, everything was derived within classical electrodynamics. Quantum mechanics enters via the concept of gauge invariance/connection which states that the wave functions along a parallel $C_\phi$ in the gauge $(n)$ and $(s)$ differ by a phase factor as:

$$\Psi^{(n)}(r) = \Psi^{(s)}(r) e^{i\int (A^{(n)} - A^{(s)}) \cdot dl} = \Psi^{(s)}(r) e^{i\int 2g \varphi}, \quad (12)$$

where $\varphi$ is the azimuthal angle. Hence, in order to ensure the single-valuedness of the wave function when comparing $\varphi$ and $\varphi + 2\pi$, one must impose that:

$$\frac{e}{\hbar} 2g2\pi = n2\pi, \quad (13)$$

where $n$ is an integer. This can be rewritten in the following way:

$$4\pi g = n\frac{\hbar}{e} \quad \text{with} \quad n \in \mathbb{Z}, \quad (14)$$

meaning that the total flux of the monopole has to be a multiple of the flux quantum $\hbar/e$, or that the monopole strength $g$ has to be an integer multiple of $g_0 \equiv \hbar(2e)$. The latter plays the role of a quantum of magnetic strength, i.e. of the ‘smallest magnetic pole’. The number of such quanta (or the ‘charge’ of the magnetic monopole) is therefore:

$$\frac{g}{g_0} = \frac{1}{2\pi} \oint_S \frac{e}{\hbar} B \cdot ds = n \in \mathbb{Z}. \quad (15)$$

The above ‘quantization’ of the magnetic strength is unusual and is known as topological quantization (see the book by Thouless [65]). Its origin is different from usual ‘quantum numbers’ that arise because of the eigen-spectrum of a Hermitian operator (an observable). Here it is the result of a topological constraint, i.e. that the gluing condition for the wave function on the equator of the sphere is related to mappings from the circle to the circle and therefore to the winding number $n$ as the fundamental group of the circle is $\pi_1(S^1) = \mathbb{Z}$ (an introduction to homotopy groups is given in section 8.1). Dirac noticed that if one magnetic monopole is present in the Universe, then all charges have to be quantized to preserve the single-valued character of wave functions.

In summary, Dirac extends the realm of electromagnetism. In classical Maxwell electromagnetism, magnetic monopoles do not exist $\nabla \cdot B = 0$. With quantum mechanics, magnetic monopoles can exist $\nabla \cdot B = 4\pi g(r)$ but only if their flux $4\pi g$ is a multiple of the flux quantum $\hbar/e$.

### 2.3. Dirac monopole as a fiber bundle

(This section may be omitted in a first reading by readers not familiar with the mathematical notion of a fiber bundle, to which we give a brief introduction later in section 4.1.)

In 1931, Dirac started its seminal paper on monopoles [50] by a philosophical discussion about the evolution of mathematics that occurs in parallel to physics and shifts towards always more abstract concepts, citing examples such as Riemannian geometry and non-commutative algebra. Amusingly, the mathematician Hopf published the very same year his work on the higher homotopy groups of the 3-sphere $S^3$ which is somewhat related to the Dirac monopole issue. The Hopf fibration relies on the fact that $S^3$ can be seen as being a nontrivial fiber bundle with base space $S^2$ and fiber $S^1$, ‘non trivial’ meaning that globally $S^3 \neq S^2 \times S^1$ although the equality holds true locally. Dirac was probably not aware of this work and it took more than 40 years to physicists and mathematicians to realize that the mathematical structure behind the Dirac monopole (with unit charge) is indeed the Hopf fiber bundle [66, 67]. Wu and Yang are actually the ones who realized that the mathematical structure behind the Dirac monopole was that of fiber bundles. Here, the base space is the total space minus the position of the monopole i.e. $\mathbb{R}^3 - \{0\} \sim S^2 \times \mathbb{R}^+ \sim S^2$ and the fiber corresponds to the phase of the wave function i.e. $U(1) \sim S^1$. As the total space $S^3$ is not globally the direct product of the base space $S^2$ and the fiber $S^1$, the fiber bundle is said to be non-trivial or twisted. A twisted fiber bundle can be characterized by a topological invariant. When the fiber is a complex vector space (here a one-dimensional Hilbert space), this invariant is known as the first Chern number and reads

$$n = \frac{1}{\hbar/e} \oint_{S^2} B \cdot ds. \quad (16)$$
which we recognize again as the ‘charge’ of the magnetic monopole or the number of flux quanta piercing the sphere. Actually, for a twisted fiber bundle, the ‘wave function’ is no longer a function but becomes a more general object (as understood by Dirac) and now known as a ‘wave-section’, building on the notion of a section of a fiber bundle. The main difference with an ordinary function is that a wave-section or generalized wave function can have a non-integrable phase.

3. Emergent Berry monopole for a two-level system

In this section, we consider a TLS driven by some external parameters, a typical example being a spin coupled to an external magnetic field. We use this fundamental system to introduce the concept of Berry phase, also called geometric phase [14, 18, 24, 54, 68]. The Berry phase is a phase angle (a number defined modulo $2\pi$) that quantifies the global phase evolution of a quantum state when this state is transported along a closed loop in the external parameter space. The related concepts of Berry connection, curvature and flux are also introduced in this minimal context. Finally, we describe the analogy between two apparently unrelated situations: the Berry connection of a TLS driven by two external parameters on the one hand [14, 24, 68], and the electromagnetic vector potential of a charge moving in the background field of a Dirac monopole on the other hand [50]. The main idea is the existence of a topological number which is the flux of the Berry curvature in the TLS case, and the flux of magnetic field for the Dirac monopole.

3.1. Two-level system

Before discussing $D$-dimensional lattice systems ($D = 1, 2, 3$), we start by introducing briefly the fundamental topological aspects of the 0-dimensional quantum TLS, whose Hamiltonian generically reads:

$$H = d \cdot \sigma = \begin{pmatrix} d_x & d_y - id_z \\ d_y + id_x & -d_x \end{pmatrix}. \quad (17)$$

The TLS consists in some isospin degree of freedom described by standard Pauli matrices $\sigma = (\sigma_x, \sigma_y, \sigma_z)$, coupled to its environment via the external parameters $d = (d_x, d_y, d_z)$. In the example of a spin $1/2$ in a magnetic field, the vector $d$ corresponds to the external magnetic field. Alternatively, equation (17) may also describe a superconducting qubit, like the Cooper pair box, the fluxonium, or the transmon. In this latter example, the isospin $\sigma$ would describe some charge, phase or flux degrees of freedom, and the vector $d = (d_x, d_y, d_z)$ would be a set of control parameters depending on gate voltages, bias fluxes, etc. Whatever the physical implementation is, the energy levels are generically given by:

$$E_{\pm} = \pm |d|, \quad (18)$$

meaning that the spectrum is determined by the norm of the vector of external parameters $d$ solely. In contrast, the corresponding spinor wave functions of the excited state $|\Psi_+^{(n)}\rangle$ and ground state $|\Psi_-^{(n)}\rangle$ depend solely on the direction of the vector $d$, and can be written:

$$|\Psi_+^{(n)}\rangle = \begin{pmatrix} \cos \frac{\theta}{2} e^{-i\varphi} \\ \sin \frac{\theta}{2} \end{pmatrix} \quad \text{and} \quad |\Psi_-^{(n)}\rangle = \begin{pmatrix} \sin \frac{\theta}{2} e^{-i\varphi} \\ -\cos \frac{\theta}{2} \end{pmatrix}, \quad (19)$$

where $\theta$ and $\varphi$ are respectively the polar (colatitude) and azimuthal (longitude) angles of the vector $d$. The explicit forms of the spinors given in equation (19) are not unique because they are defined up to a global phase. The choice made in equation (19) defines a gauge where the spinor $|\Psi_-^{(n)}\rangle$ is not well-defined when $\theta = \pi$. Indeed at the south pole of the Bloch sphere this spinor reads:

$$|\Psi_-^{(n)}\rangle \rightarrow \begin{pmatrix} e^{-i\varphi} \\ 0 \end{pmatrix}, \quad (20)$$

and $\varphi$ is not defined at the poles. In contrast, $|\Psi_-^{(n)}\rangle \rightarrow (0, -1)^T$ is well-defined at the north pole, and in fact everywhere except at the south pole, hence the superscript $(n)$. This gauge is characterized by the fact that the phase $\varphi$ disappears from the spinors $|\Psi_{\pm}^{(n)}\rangle$ at the north pole, where $\theta = 0$, but not at the south pole.

In order to cure the fact that the ground state spinor is not defined unambiguously at the south pole of $S^2$, it is possible to choose a different gauge simply by multiplying the spinors equation (19) by an overall $e^{\pm i\varphi}$ factor, leading to:

$$|\Psi_+^{(n)}\rangle = \begin{pmatrix} \cos \frac{\theta}{2} e^{-i\varphi} \\ \sin \frac{\theta}{2} \end{pmatrix} \quad \text{and} \quad |\Psi_-^{(n)}\rangle = \begin{pmatrix} \sin \frac{\theta}{2} e^{-i\varphi} \\ -\cos \frac{\theta}{2} e^{+i\varphi} \end{pmatrix}. \quad (21)$$
Note that the phase factor $e^{\pm i\varphi}$ is now multiplied by $\cos \frac{\theta}{2}$ instead of $\sin \frac{\theta}{2}$, compare with equation (19). Within this new gauge, denoted by the superscript $(s)$, the ground state wave function is now well-defined at the south pole, but at the expense of being ill-defined at the north pole because when $\theta \to 0$, then $|\Psi^s_+\rangle \to (0, e^{i\varphi})^T$ [69]. There is always a singularity remaining; changing from gauge $(n)$ towards gauge $(s)$ only moves away this singularity but cannot remove it completely. We will see that this is related to the fact that the total Berry flux is non zero.

We have exhibited two distinct gauges, namely $(s)$ and $(n)$, and we will keep on comparing them below for pedagogical purposes, but there are of course an infinity of other possible gauges. They are all perfectly equivalent and useful to describe the system at given values of the parameters. In a time-independent problem, this huge gauge freedom is merely a matter of fixing a working convention for representing states at once, and of course this initial choice will not alter the final results. The stationary states of the Hamiltonian are fixed up to a global phase. Once this phase is chosen, it is possible to keep this fixed basis to study the unitary evolution of the state of the system.

In a time-dependent problem, the situation is more subtle. Even in the adiabatic limit, the parameters $d$ of the Hamiltonian change and thus one needs to diagonalize a different Hamiltonian at each value of the parameters, and therefore pick a different choice of global phase for each values of these parameters. There is a huge amount of gauge freedom, and clearly the physical observables cannot depend on this arbitrariness.

### 3.2. Gauge freedom and Berry connections

We now consider a driven-TLS described by a parameter-dependent Hamiltonian $H(\theta, \varphi)$. If we are interested in the variation of the spinors, it is convenient to use a parametrization in terms of the spherical angles $\theta$ and $\varphi$ of the vector $d$. A spinor state is attached to each point of the unit sphere $S^2$, which is called the Bloch sphere in this context (the Riemann sphere for mathematicians). The Bloch sphere representation is extensively used to monitor the evolution of a spin state in nuclear magnetic resonance, or a qubit state in the Bloch sphere in this context (the Riemann sphere for mathematicians). The Bloch sphere representation is the stereographic projection from the south pole, but at the expense of being ill-defined at the north pole because when $\theta \to 0$, $|\Psi\rangle$ only moves away this singularity but cannot remove it completely. We will see that this is related to the fact that the total Berry flux is non zero.

The next step is to focus on the evolution of the groundstate $|\Psi\rangle = |\Psi_-\rangle$ as the angular parameters $\theta$ and $\varphi$ are varied, so we drop the subscript in the following. This is essentially the idea of the adiabatic following of a single level. At this point, we have projected on a single band (the lowest level). We call it a band because we consider that the Hamiltonian depends on two parameters $(\theta, \varphi)$. As long as only the specific properties of the spinorial wave functions are investigated, the specific dispersion of the band $E_-(\theta, \varphi) = -d(\theta, \varphi)$ upon $(\theta, \varphi)$ is not relevant. Provided the level do not cross, i.e. $|d(\theta, \varphi)| \neq 0$ for all parameter values, it is even possible to do a band flattening procedure which leads to a constant groundstate energy $E_-$. To follow the evolution of an eigenstate $|\Psi\rangle$, it is natural to compute overlaps such as:

$$\langle \Psi(\theta, \varphi) | \Psi(\theta + d\theta, \varphi) \rangle \quad \text{and} \quad \langle \Psi(\theta, \varphi) | \Psi(\theta, \varphi + d\varphi) \rangle.$$  \hspace{1cm} (22)

Extracting the phase of these inner products leads to define the Berry connection of a ket $|\Psi\rangle$ as the inner products [24]:

$$A_\theta = i\langle \Psi | \partial_\theta \Psi \rangle \quad \text{and} \quad A_\varphi = i\langle \Psi | \partial_\varphi \Psi \rangle,$$ \hspace{1cm} (23)

where $\partial_\theta = \partial / \partial \theta$ and $\partial_\varphi = \partial / \partial \varphi$. The components $A_\theta$ and $A_\varphi$ are real quantities, because $\langle \Psi | \partial_\theta \Psi \rangle$ and $\langle \Psi | \partial_\varphi \Psi \rangle$ are purely imaginary. Indeed $\partial_\theta (\langle \Psi | \Psi \rangle) = 0 = \langle \partial_\theta \Psi | \Psi \rangle + \langle \Psi | \partial_\theta \Psi \rangle$ and therefore $(\langle \partial_\theta \Psi | \Psi \rangle)^* = \langle \Psi | \partial_\theta \Psi \rangle = -\langle \partial_\theta \Psi | \Psi \rangle$.

The Berry connections are gauge-dependent objects. For instance, the Berry connection associated to the ground state $|\Psi_-\rangle$ is then given by:

$$A_\theta^{(n)} = 0 \quad \text{and} \quad A_\varphi^{(n)} = \sin^2 \frac{\theta}{2},$$ \hspace{1cm} (24)

within the $(n)$-gauge, and by:

$$A_\theta^{(s)} = 0 \quad \text{and} \quad A_\varphi^{(s)} = -\cos^2 \frac{\theta}{2},$$ \hspace{1cm} (25)

within the $(s)$-gauge. These two different expressions for $A_\varphi^{(s)}$ differ by a constant one, which is the gradient $\partial_\varphi \varphi = 1$. 

10
Berry noticed that integrals of these connections around closed loops in parameter space are gauge-independent. Let us consider the circulations of $A_{\varphi}^{(i)}$ and $A_{\varphi}^{(n)}$ along a specific path $C_0$, which is defined as the parallel-type circle at constant $\theta$, and oriented from $\varphi = 0$ to $\varphi = 2\pi$. Those two circulations read:

$$\Phi^{(i)} = \oint_{C_0} d\varphi A_{\varphi}^{(i)} = 2\pi \sin^2 \frac{\theta}{2} \quad \text{and} \quad \Phi^{(n)} = \oint_{C_0} d\varphi A_{\varphi}^{(n)} = -2\pi \cos^2 \frac{\theta}{2}.$$  \hspace{1cm} (26)

Clearly these circulations differ by $\Phi^{N} - \Phi^{S} = 2\pi$, and therefore describe the same phase. The relevant gauge-invariant quantity is not the Berry phase (except in the modulo $2\pi$ sense), but rather its exponential, i.e. the Berry phase factor also known as an abelian Wilson loop (more on Wilson loops in section 4.5):\[ W(C_0) = e^{i\Phi^{(i)}} = e^{i\Phi^{(n)}}. \hspace{1cm} (27)\]

In conclusion, the Berry phase accumulated along a closed path is gauge independent modulo $2\pi$, and therefore may be observable in some interference experiments. In contrast, the Berry phase accumulated by a quantum state along an open path of the parameter space typically/usually depends on the gauge, except if one takes special care by defining a closing procedure, see [52]. We will see one such example of open-path Berry phase when discussing the Zak phase, see section 4.3.2.

The Berry phase is an example of anholonomy, i.e. the failure to come back to the exact same initial state after performing parallel transport along a closed path in a curved parameter space. It is actually a quantum version of a well-known geometrical effect. An elementary example, not in quantum mechanics, is that of the parallel transport of a stick on the surface of earth (the globe). Imagine a walker starting from the north pole and holding a stick in a given direction. The walker now moves to the south along a meridian trying to maintain the stick parallel at each moment (that's the notion of parallel transport). The walker next reaches the equator, makes a left turn and walks along the equator for a quarter of its length, before turning left again and holding a stick in a given direction. The walker now moves to the south along a meridian trying to parallel transport a stick, the surprise of the walker is that the final direction of the stick makes an angle (90° in our example) with the original direction. The angle between the initial and final direction is equal to the solid angle covered on the globe (namely 1/8 of the total solid angle 4π in our example). The Berry phase is a quantum version of such a classical anholonomy.

3.3. Berry curvature and Chern number

It is important to define physical quantities that are independent of the gauge choice. By taking the curl of the Berry connection equation (23), it is possible to get rid of the gradients and obtain such a gauge-invariant quantity, the so-called Berry curvature. In a 2D parameter space, the curl has only one component which is a pseudo-scalar:

$$F_{\theta \varphi} = \partial_{\theta} A_{\varphi} - \partial_{\varphi} A_{\theta} = i(\partial_{\theta} \Psi | \partial_{\varphi} \Psi) - i(\partial_{\varphi} \Psi | \partial_{\theta} \Psi) = i(\partial_{\theta} \Psi | \partial_{\varphi} \Psi) + c.c.$$  \hspace{1cm} (28)

For the TLS, the Berry curvature reads:

$$F_{\theta \varphi} = \frac{1}{2} \sin \theta,$$  \hspace{1cm} (29)

and its total flux integrated over the whole parameter space is finite:

$$\int_{S^2} d\varphi d\theta F_{\theta \varphi} = \frac{1}{2} \int_{0}^{2\pi} d\varphi \int_{0}^{\pi} d\theta \sin \theta = 2\pi.$$  \hspace{1cm} (30)

This can be seen as (i) the integral of the function $(\sin \theta)/2$ over the square $[0, 2\pi] \times [0, \pi]$, or alternatively as (ii) the flux of a radial vector of constant length 1/2 through the unit sphere whose surface element is $d\theta \ d\varphi \sin \theta$. The nonzero value of the total flux through the parameter sphere signals a topological feature, that we will interpret in the next section as originating from the presence of a monopole of unit strength at the origin.

One can prove that the total Berry flux is always a multiple of $2\pi$ for a single band, using Stokes’ theorem on appropriate domains of the sphere. Let us define two submanifolds realizing a partition of the sphere: the cap ($N$) gathering the regions located at the north of the parallel $C_0$, and the cap ($S$) at the south of $C_0$ (figure 2).

Within the northern cap ($N$), one may safely apply Stokes’ theorem using the gauge ($n$):

$$\int_{N} d\varphi d\theta F_{\theta \varphi} = \oint_{C_0} d\varphi A_{\varphi}^{(n)} = 2\pi \sin^2 \frac{\theta}{2},$$  \hspace{1cm} (31)

because $A_{\varphi}^{(n)}$ is well-defined over ($N$).
Within the southern cap \((S)\), one may similarly apply Stokes’ theorem but using the gauge \((s)\):
\[
\int_S d\varphi d\theta F_{\theta \varphi} = -\oint_{C_n} d\varphi A^{(s)}_{\varphi} = 2\pi \cos^2 \frac{\theta}{2},
\]
where the minus sign is due to the orientation of the circle \(C_n\), which should be reversed to apply Stokes’ theorem to the south cap \(S\). Finally, the total flux through the 2-sphere is:
\[
\int_S d\varphi d\theta F_{\theta \varphi} = \int_{S'} d\varphi d\theta F_{\theta \varphi} + \int_S d\varphi d\theta F_{\theta \varphi} = \oint_{C_n} d\varphi (A^{(s)}_{\varphi} - A^{(n)}_{\varphi}) = 2\pi,
\]
where the last equality is consistent with the direct evaluation equation (30). More generally, this shows that the total Berry flux is the circulation of the difference between Berry connections written in two distinct gauges and therefore it is the circulation of a gradient over a closed path, which has to be a multiple of \(2\pi\), here simply \(2\pi\). This allows one to define the Chern number as the total flux of the Berry curvature in units of \(2\pi\). This quantization of the total Berry flux is also valid for any closed 2D-manifold, like a torus \(T^2\), because the same demonstration can be done by covering the parameter manifold by patches. A smooth connection is defined over each patch by a proper choice of gauge, the Berry phase accumulation along the closed path separating the patches has to be unique modulo \(2\pi\).

In conclusion, the Berry curvature is a local gauge-independent field. The Berry flux through the whole parameter space is a global gauge-independent quantity, which is a multiple of \(2\pi\). The Chern number is an integer, which is the total Berry flux in units of \(2\pi\). Recently, this quantization of the Berry flux has been measured for an individual superconducting qubit [70]. It is actually possible to simulate the physics of a complete band structure with a single TLS that is driven in time. For example, the physics of the Haldane model on the honeycomb lattice, that we discuss below in section 6.3, was simulated in [71].

In lattice systems of space dimensions two (see section 6), this Chern number is very important because it is related to physical observables and to their topological robustness.

### 3.4. Berry flux monopole in parameter space

In the previous paragraph, we have presented a justification of the integer character of the Chern number which is very reminiscent of the Wu–Yang construction of the flux quantization for a Dirac magnetic monopole. There is indeed a strong analogy between the structure of the two problems although they might seem very different at first sight.

To facilitate the analogy, let us perform a change of parameter space from the spherical angular parameters \((\theta, \varphi)\) to the Euclidian space \(R^3\) spanned by the cartesian parameters \((d_x, d_y, d_z)\) of the driven TLS. The Berry connection introduced previously appears in a new guise, because it is defined here with respect to the cartesian components of the parameter field \(d = (d_x, d_y, d_z)\), rather than in terms of the spherical angles \((\theta, \varphi)\) of its direction:
\[
\tilde{A} = i\langle \Psi \mid \nabla_d \Psi \rangle = i\langle \Psi \mid \left( e_\theta \frac{1}{d} \frac{\partial}{\partial \theta} + e_\varphi \frac{1}{d \sin \theta} \frac{\partial}{\partial \varphi} \right) \rangle \langle \Psi \rangle.
\]
Here \(\tilde{A} = (\tilde{A}_x, \tilde{A}_y, \tilde{A}_z)\) instead of \(A = (A_\theta, A_\varphi)\). The radial component of the gradient could have been written, but it is actually vanishing because the eigenkets are independent of the norm \(d = |\mathbf{d}|\). This ‘change of variables’ leads to the following expressions. In the ‘north-gauge’ \((n)\), from equation (19), we immediately obtain for the new vector potential:
\[
\tilde{A}^{(n)} = \frac{\sin^2(\theta/2)}{d \sin \theta} e_\varphi = \frac{1 - \cos \theta}{2d \sin \theta} e_\varphi,
\]
for \(\theta = 0\) and \(\pi\), as expected. This result is very reminiscent of the Wu–Yang construction of the flux quantization for a Dirac magnetic monopole.
corresponding to equation (24). Similarly, in the ‘south gauge’ (S) the new connection reads:

$$\tilde{A}^{(s)} = \frac{-\cos^2(\theta/2)}{d\sin^2 \theta} e_\varphi = -\frac{1 + \cos \theta}{2d\sin \theta} e_\varphi,$$

replacing equation (25). In this representation, the singularities at the poles can be seen even more explicitly in the expressions of the connections. For instance when $\theta \to 0$, the norm of $\tilde{A}^{(n)}$ is regular, while the norm of $\tilde{A}^{(s)}$ diverges. We recognize that the Berry connections equations (35) and (36) map exactly to the electromagnetic vector potentials equations (8) and (10) provided one sets $g = 1/2$ (i.e. $n = 1$ with $h = 1$, $e = 1$) and $d \to r$. The parameter space, spanned by the components of $d$, replaces the real space $r$ of the original Dirac monopole problem. The location of the Berry monopole at $d = 0$ corresponds to a level degeneracy as $E_{\pm} = \pm |d| = 0$.

The Berry curvature is obtained by taking the curl of the connection, in either gauge (so we drop the $(n), (s)$ indexes below):

$$F = \nabla_d \times \tilde{A} = \frac{1}{d\sin \theta} \frac{\partial}{\partial \theta} (\sin \theta \tilde{A}_\varphi) e_r = \frac{1}{2d^2} e_r,$$

which is a radial vector field. It is worth noticing that the Berry curvature is a 3-component vector in this definition, while it was a pseudoscalar in section 3.1. This Berry curvature vector is a local and gauge-invariant quantity, and it is therefore observable in principle. It is important to notice than the dimension of the parameter space is not related to the dimension of real space. As in the previous section, one can built a global gauge-invariant quantity by evaluating the flux of the Berry curvature through a surface. For instance, the flux of $F$ through a sphere, with radius $d$, surrounding the origin:

$$\oint F \cdot dS = 4\pi d^2 / 2d^2 = 2\pi.$$

The geometric (or Berry) phase structure of the TLS is in fact related to the existence of a monopole in parameter space $(d_x, d_y, d_z)$. Within an electromagnetic analogy, the Berry phases can be interpreted as quantum mechanical phases accumulated by a charge coupled to a fictitious vector potential. This analogy between the driven-TLS and the Dirac monopole can be spelled out in detail. In the first case we have a quantum TLS described by a spinor wave function and no magnetic field. If we describe it in an approximate manner as a scalar—one-component instead of two for the spinor—wave function upon projection on a single band, i.e. adiabatic following, then we are forced to introduce an emergent gauge field (the Berry connection). The latter corresponds to a magnetic monopole in parameter space and accounts for the effect of virtual transitions to the other band (the band that we got rid of upon projection). Afterwards, we realize that we study a scalar wave function in the field of a magnetic monopole in parameter space. This is nothing but the situation considered by Dirac in real space. Hence we see that the problem of a scalar wave function in the field of a monopole is an adiabatic approximation to the quantum evolution of a spinor wave function. We also note that both the Dirac monopole and the TLS share the mathematical structure of the Hopf fibration [72].

4. Geometrical and topological band theory

In the previous section, we outlined the topological features of a driven TLS controlled by two independent external parameters, essentially the angles $\theta$ and $\varphi$ that determine its spinor ground state. We now move to electrons on $D$-dimensional lattices. These particles may carry spin and/or some other internal isospin (orbital, sublattice, etc). We neglect electron-electron interactions and concentrate on the geometrical and topological features of the band structure. Independent electrons (or cold atoms [61]) in a periodic $D$-dimensional system occupy bands of Bloch states separated by gaps. In the absence of disorder, these Bloch states are labeled by a crystal momentum (or Bloch wave vector) $k$ living in $D$-dimensional periodic BZ, which is a torus $T^D$. Inter-band effects may occur when at least two bands are coupled. Then for each value of the crystal momentum $k$ one has essentially an effective TLS whose two states are generically described by the spinors equation (21). In this context, the angles $\theta$ and $\varphi$ become functions of the crystal momentum $k$, and it is not necessary to drive externally the parameters. Indeed, the physical quantities are naturally expressed as sums over the occupied Bloch states. The physics depends on the dimension $D$ and also on how the point $(\theta, \varphi)$ covers the Bloch sphere as $k$ spans the whole BZ $T^D$, which is constrained by the symmetries of the Bloch Hamiltonian. The Berry phase, connection, curvature and Chern number concepts (defined in the previous section on a 0D TLS) have been extended very fruitfully to electrons in crystals, or atoms in optical lattices.
In this section, we first introduce the mathematical concept of fiber bundle, and then show its implementation in band theory. An important difference with the previous sections (Dirac monopole and TLS) is that the parameter space in band theory is a torus (the BZ) instead of a sphere. When relevant, we will point out the consequences of this difference. Next, we study geometrical phases and review the semi-classical formalism describing the dynamics of a particle in a given band in the presence of the Berry curvature field which accounts for the influence of inter-band transitions on the intra-band motion. This leads to the anomalous velocity concept and the related Hall effects.

4.1. Introduction to fiber bundles

The mathematical objects hiding behind the geometrization of band theory are fiber bundles (see Nakahara for a reference accessible to physicists [57, 73]). A fiber bundle \( E \) is a geometrical object made of a base manifold \( B \), at each point of which a fiber \( F \) is attached. A fiber is itself a manifold that may be, e.g., a real or a complex vector space. Locally, a fiber bundle \( E \) resembles the direct product \( B \times F \). A very simple example is \( E = \mathbb{R}^3 \) that may be described as a fiber bundle of base \( B = \mathbb{R}^2 \) and fiber \( F = \mathbb{R}^2 \) (alternatively it can also be described as a fiber bundle with base \( B = \mathbb{R} \) and fiber \( F = \mathbb{R}^2 \)). Even if locally, a fiber bundle \( E \) resembles the direct product \( B \times F \), this needs not be the case globally over the complete base space. When a fiber bundle is simply the direct product of a base space and a fiber \( E = B \times F \), it is said to be topologically trivial. This is the case of the above example, \( \mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R} \). When it is not, it is said to be non-trivial or twisted. The (local) geometry of a fiber bundle is described by objects such as connections and curvature, whereas its (global) topology is characterized by topological invariants called characteristic classes (for example, a Chern number).

One also defines a map \( p \) that projects from \( E \) to \( B \) and a structure group \( G \) that acts on the fiber. A standard notation for a fiber bundle is:

\[
F \rightarrow E \xrightarrow{p} B. \tag{39}
\]

An important notion about fiber bundles is that of a section \( s \). It is a continuous map from \( B \) to \( E \). Naively, it could be thought as being the inverse of the projection map \( p \). This is the case for a trivial fiber bundle but not for a twisted fiber bundle. Actually, the existence of a global and non-vanishing section is equivalent to the fiber bundle being trivial. A section can also be thought of as a generalization of a function defined over the base space. If the fiber bundle is trivial, a global section is simply an ordinary function. If it is twisted, the section is a function that is defined over different patches that together cover the base space. In other words, a section appears as a multi-valued (and hence ill-defined) function. Alternatively, a section can be seen as an extension of the notion of a function. One definition of a twisted fiber bundle is that there is an obstruction in finding a global section that is non-vanishing.

When the fiber is itself the structure group \( G \), then the fiber bundle is called a principal \( G \)-bundle. This will turn out to be the relevant type of fiber bundles in band theory.

The simplest example of a non-trivial fiber bundle is the well-known Möbius strip. In this case, the base space is a circle \( B = S^1 \) and the fiber is a line segment \( F = [-1, 1] \). The structure group is \( G = \mathbb{Z}_2 = \{1, -1\} \), whose non-trivial element \(-1\) flips the fiber \([1, -1] \) into \([-1, 1] \). Locally, at each point of the circle \( B \), one places a line segment \( F \) perpendicularly. The global subtlety is in how these fibers are arranged around the circle. Let us call \( \theta \) the parameter that spans the base space \( 0 \leq \theta \leq 2\pi \). If the gluing of the last fiber (the one at \( \theta = 2\pi^- \)) to the first one (at \( \theta = 0^+ \)) is done naively, then one ends up with a regular cylinder, which is a trivial fiber bundle: the cylinder \( S^1 \times [-1, 1] \) (see figure 3 left). If the gluing is done after twisting or flipping the last fiber (i.e. by gluing the upper end of the fiber at \( \theta = 2\pi^- \) to the lower end of the fiber at \( \theta = 0^+ \)), then one ends up with a Möbius strip (see figure 3 middle). The latter is a non-trivial fiber bundle: the Möbius strip \( S^1 \times [-1, 1] \). The two fiber bundles (cylinder and Möbius strip) are locally identical but are globally very different. For example, the Möbius strip has a single edge and a single side, whereas the cylinder has two different edges and two different sides (an inside and an outside). For a twisted bundle, there is no global non-vanishing section, which translates into the following funny fact for the Möbius strip: when cutting it in its middle (i.e. along the value 0 in each fiber \( F = [-1, 1] \)), one still obtains a connected object (made of a single piece but twice longer) that is neither a cylinder nor a Möbius strip (see figure 3 right). It is similar to a Möbius strip but with two twists instead of a single.

Another example of a fiber bundle is that of the torus \( T^2 \) that can be described as a trivial fiber bundle of base space \( S^1 \) and fiber \( S^1 \) (very similar to the cylinder with \( F = [-1, 1] \) replaced by \( F = S^1 \)): \( T^2 = S^1 \times S^1 \). The structure group is now \( G = U(1) \) instead of \( \mathbb{Z}_2 \). When twisting such a fiber bundle in the same way as the Möbius strip, one obtains the Klein bottle. Locally the torus and the Klein bottle are very similar, but globally one is a trivial and the other one a twisted fiber bundle.

As a last example of a fiber bundle, we go back to our favorite Dirac monopole (of unit charge) which corresponds to the Hopf fiber bundle \( E = S^3 \) with base space \( B = S^2 \), fibers \( F = S^1 \) and structure group
$G = U(1)$. Why $S^3$? Because $S^3$ is the Hilbert space for a spinor wave function $|\psi\rangle = (z_1, z_2)^T$. Normalization of the spinor means that its two complex components $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ satisfy

$$1 = |z_1|^2 + |z_2|^2 = x_1^2 + y_1^2 + x_2^2 + y_2^2$$

which is indeed the equation of a unit sphere in 4-dimensional space, i.e. $S^3$. In the previous section on the Berry monopole, we have seen that there is a TLS hiding in the back of the Dirac monopole. The topology of such a fiber bundle is characterized by the Chern number, which can also be interpreted as the number of monopoles enclosed by the base space. The projection $p$ from $E = S^3$ to $B = S^2$ is known as the Hopf map.

4.2. Band structure and Bloch fiber bundle

In the context of band theory for a single electron in a periodic potential (a crystal), the relevant fiber bundles are called Bloch bundles. The base (parameter) space is the BZ torus ($k \in T^3$) and the fibers are complex vectors spaces (Hilbert spaces) corresponding to band space (with band index $n$), the dimension of which depends on the number of bands $N = 1, \ldots, N_B$. For example, in a two-band system, one may define a Bloch bundle with a two-dimensional Hilbert space as a fiber. In this case, the complete Hilbert space is seen as a fiber bundle and the latter is not very interesting as it can be shown to be always trivial, see e.g. [57]. A more interesting fiber bundle is obtained when one only takes a sub-set of bands, e.g. a fiber obtained by keeping only the one-dimensional Hilbert space corresponding to the lower band (imagine the case of a two-band insulator: the lower band is filled and the upper band empty and they are separated by a band gap). For a nice review on fiber bundles in the context of band theory, see [57].

4.2.1. Bloch Hamiltonian and cell-periodic Bloch states

Let us consider a single-particle Hamiltonian $H$ having translation invariance under a Bravais lattice (e.g. a tight-binding Hamiltonian). Most of this section is valid in any space dimension $D$, but at the end we will focus on space dimension $D = 2$ which is special for topology as we have seen in the TLS section. A parameter-dependent Bloch Hamiltonian is defined via a unitary transformation as:

$$H(k) = e^{-i k \cdot r} H e^{i k \cdot r},$$

(40)

where $r$ is the position operator and $k$ is a parameter with the dimension of a wave vector. In the following, we will refer to it as the canonical Bloch Hamiltonian or simply the Bloch Hamiltonian. Bloch’s theorem allows us to diagonalize the Hamiltonian as:

$$H|\psi_{n,k}\rangle = E_{n,k} |\psi_{n,k}\rangle,$$

(41)

where $k$ is the crystal momentum and $n$ is a discrete band index. The Bloch eigenvectors are:

$$|\psi_{n,k}\rangle = e^{i k \cdot r} |u_{n,k}\rangle,$$

(42)

where in coordinate representation the ‘cell-periodic Bloch state’ $u_{n,k}(r) = \langle r | u_{n,k}(k) \rangle$ obeys:

$$u_{n,k}(r + R) = u_{n,k}(r),$$

(43)

namely it is periodic in real space with the periodicity of the unit cell ($R$ belongs to the Bravais lattice). In the following, in order to emphasize the fact that the Bloch wavevector $k$ now plays the role of a parameter (unlike the band index $n$ which remains a quantum number), we have chosen to write the cell-periodic Bloch eigenvector as $|u_{n,k}(k)\rangle$ rather than $|u_{n,k}\rangle$. Let us now examine the $k$-dependence of $|u_{n,k}(k)\rangle$ in the reciprocal
Another reason for choosing \(54\) modified by the integration over the fast (or 'light') degrees of freedom is well explained in 

\[
\text{much larger than the second. The idea that the emergent (Berry) gauge structure generally appears via a}
\]

\[
\Delta
\]

\[
\text{description for the dynamics of an electron restricted to a single band (this will be done by projecting on a single band) nevertheless taking into account the coupling to other bands (this will occur via an emergent gauge field). To make this discussion more concrete, let us consider a one-dimensional tight-binding model (but with several bands) with hopping amplitude \(t\) and crystal size \(L\). The timescale for the intra-band dynamics can be estimated as \(\sim \hbar L/ta\) (because the spacing between allowed \(k\) values is \(\Delta k = 2\pi/\hbar L\) and the typical band velocity is \(ta/\hbar\) so that the typical energy change \(\Delta E \sim ta\Delta k \sim taL\), and that for the inter-band dynamics as \(\sim h/\text{bandgap} \sim \hbar t\). For a macroscopic crystal \(L \gg a\), the first timescale is much larger than the second. The idea that the emergent (Berry) gauge structure generally appears via a separation of two timescales such that the effective dynamics of the slow (or 'heavy') degrees of freedom gets modified by the integration over the fast (or 'light') degrees of freedom is well explained in [74].}

Paraphrasing Michael Berry, the reaction of the fast degrees of freedom ('light system') onto the slow degrees of freedom ('heavy system') occurs via the appearance of an emergent gauge field, as we will see.

### 4.2.2. Berry curvature and quantum metric

The \(|u_n(k)\rangle\)'s are the parameter-dependent eigenvectors of the Bloch Hamiltonian:

\[
H(k)|u_n(k)\rangle = E_n(k)|u_n(k)\rangle
\]

This last equation appears very similar to (41) but is actually quite different. Whereas the set \(|\psi_{n,k}\rangle\) forms an orthonormal basis in Hilbert space, it is not so for \(|u_n(k)\rangle\). Indeed \(|u_n(k)\rangle\) and \(|u_n(k')\rangle\) are eigenvectors of two different Hamiltonians \(H(k)\) and \(H(k')\) and therefore need not be orthogonal. However \(|u_n(k)\rangle |u_{n'}(k')\rangle = \delta_{n,n'}\) as \(|u_n(k)\rangle\) and \(|u_{n'}(k')\rangle\) are eigenvectors of the same Hamiltonian \(H(k)\). Their overlap

\[
(u_n(k)\rangle|u_{n'}(k')\rangle) \neq \delta_{k,k'}
\]

is a non-zero complex number in general. Note that, precisely at this point, we have restricted the discussion to a single band (the \(n\)th band). This is the moment where we stop discussing the dynamics in the full Hilbert space and project on a single band of interest. The corresponding fiber \(F\) is a one dimensional complex vector space, i.e. essentially a \(U(1)\) phase, while the first BZ torus \(T^D\) plays the role of the base space \(B\).

When \(k' = k + \Delta k\) is close to \(k\), one may study the deviation of this overlap (46) from unity and define the evolution (i) of its phase and (ii) of its norm:

(i) The evolution of its phase, by expanding at first order in \(\Delta k\),

\[
(u_n(k)\rangle|u_n(k + \Delta k)\rangle \approx (u_n(k)\rangle(1 + \Delta k \cdot \nabla_k)|u_n(k)\rangle \approx e^{-i\Delta k A_n(k)}
\]

defines the Berry connection:

\[
A_n(k) = i \langle u_n(k)|\nabla_k u_n(k)\rangle,
\]

which is the Bloch version of the TLS formula equation (23). This quantity is called \(X_{nn}(k)\) in [7] and is related to the projected position operator. The Berry curvature is given by the curl of the Berry connection:

\[
F_{nn}^B(k) = i \langle \partial_k u_n|\partial_k u_n\rangle + \text{c.c.}
\]

which is the Bloch counterpart of equation (28) for the TLS. The Berry curvature \(F_n(k) = A_n(k)\) is called \(\Omega_n(k)\) in [7]. The geometry of the fiber bundle is described by the Berry connection, curvature and phase.

---

4 Another reason for choosing \(u_{n,k}(x)\) over \(\psi_{n,k}(x)\) is that the \(u_{n,k}(x)\)'s all belong to the same Hilbert space as they have the same boundary conditions in \(x = 0\) and \(x = a\) [24, 54].
while its topology is characterized by the Chern number (see below). In the context of band theory, this was first recognized by Thouless et al. [12, 75] and Simon and coworkers [15, 76] who underlined the relation with Berry’s contribution.

(ii) The evolution of the norm of this overlap (46) defines another geometric quantity, known as the quantum metric, obtained by introducing a distance (squared) in the Hilbert space:

$$ds^2 = 1 - |\langle u_n(k)|u_n(k + dk)\rangle|^2 \approx \sum_{i,j} g_{ij}^{(n)}(k) dk_i dk_j.$$  (50)

This distance in projective Hilbert space was introduced by Provost and Vallée [77]. Expanding at second order in $dk$, the quantum metric is obtained as:

$$g_{ij}^{(n)}(k) = \text{Re}(\partial_i u_n)(|\mathbb{1} - |u_n\rangle\langle u_n|)|\partial_j u_n|,$$  (51)

where $\partial_i$ is a short-hand notation for $\partial_k$ and $\mathbb{1}$ is the identity. The quantum metric and the Berry curvature are the real and imaginary part of a more general object called the quantum geometric tensor:

$$T_{ij}^{(n)}(k) = (\partial_i u_n)(|\mathbb{1} - |u_n\rangle\langle u_n|)|\partial_j u_n|.$$  (52)

Indeed $g_{ij}^{(n)} = \text{Re}T_{ij}^{(n)}$ and $F_{ij}^{(n)} = -2\text{Im}T_{ij}^{(n)}$. For more information on the quantum metric, see the article by Berry in [68]. The quantum metric appears in some physical quantities such as the magnetic orbital susceptibility [78] or the superfluid weight [79]. It is also useful to define localization in an insulator [53] and gives a measure of the minimal wavepacket spreading for a Bloch electron [80].

4.2.3. Berry gauge transformation

There is a gauge freedom in the cell-periodic Bloch states seen as functions of the parameter $k$. Indeed, provided the phase $\varphi(k)$ is a smooth enough function of $k$,

$$|\tilde{u}_n(k)| = e^{i\varphi(k)}|u_n(k)|,$$  (53)

is also a valid choice for the cell-periodic Bloch states. We refer to it as a Berry gauge transformation to distinguish it from a real-space electromagnetic gauge transformation. In this gauge transformation $|u_n(k)\rangle \rightarrow |\tilde{u}_n(k)\rangle$, the Berry connection gets modified as:

$$\tilde{A}_n = i\langle \tilde{u}_n|\nabla_k \tilde{u}_n\rangle = A_n - \nabla_k \varphi,$$  (54)

and is therefore not gauge-invariant just like the vector potential in electromagnetism. However, the Berry curvature is gauge-invariant since:

$$\tilde{F}_n = \text{curl } \tilde{A}_n = \text{curl } A_n + 0 = F_n,$$  (55)

just like the magnetic field. More generally, the quantum geometric tensor, i.e. both the Berry curvature and the quantum metric, is gauge-invariant and therefore measurable. A map of the Berry curvature and of the quantum metric in the whole BZ has been measured in artificial crystals, see e.g. [81, 82].

4.2.4. Berry curvature and virtual transitions

The above expression may give the impression that the Berry curvature in the $n$th band only depends on a single band. This is actually not the case. To show that the Berry curvature is related to virtual transitions between bands (at fixed $k$), it is useful to rewrite it using perturbation theory to express $|\langle \partial_i u_n\rangle|$ as a function of $|u_{n'}\rangle$ with $n' \neq n$:

$$F_{ij}^{(n)}(k) = i \sum_{n' \neq n} \frac{|\langle u_n|\partial_i H(k)|u_{n'}\rangle\langle u_{n'}|\partial_j H(k)|u_n\rangle}{(E_{n'} - E_n)^2} + \text{c.c.}$$  (56)

This expression has the flavor of second-order perturbation theory and clearly shows that the Berry curvature in the $n$th band is the result of virtual transitions (see figure 4) to other bands and that the inter-band coupling is related to the velocity operator $\partial_i H(k)$. As a consequence, it is obvious that if there is a single band in the model, these effects are totally absent. Also, the Berry curvature is well-defined only when there is a gap between bands (at fixed $k$, i.e. $E_{n'}(k) \neq E_n(k)$) and becomes large when this gap becomes small.

The above expression makes it obvious that the Berry curvature is gauge-independent (as for each bra $\langle u_{n'}|$), it involves the corresponding ket $|u_{n'}\rangle$), in contrast to the Berry connection which is gauge-dependent (as it involves a bra $|u_n\rangle$ but a different ket $|\nabla_k u_n\rangle$). The Berry curvature also has the periodicity of the reciprocal
Bravais lattice $F_{ij}^n(k + G) = F_{ij}^n(k)$, although this is not the case of the $|u_n(k)\rangle$'s. A consequence of equation (56) is that the sum over all bands of the Berry curvature at a given $k$ point vanishes:

$$\sum_n F_{ij}^n(k) = 0.$$  \hspace{1cm} (57)

Time-reversal symmetry implies:

$$F_{ij}^n(-k) = -F_{ij}^n(k),$$  \hspace{1cm} (58)

while inversion symmetry imposes:

$$F_{ij}^n(-k) = F_{ij}^n(k).$$  \hspace{1cm} (59)

Hence, if both symmetries are present, the Berry curvature vanishes everywhere in the BZ.

4.2.5. Periodic versus canonical Bloch Hamiltonian (basis I/II issue)

In general, the Bloch Hamiltonian $H(k)$ does not have the periodicity of the reciprocal lattice, even if the spectrum $E_n(k)$ always has it. The reason is that the definition of the Bloch Hamiltonian (40) involves the Hamiltonian $H$ and the position operator $r$. The latter depends on the position of every site in the crystal, including sites within the unit cell (e.g. for a lattice with a basis). The distance between sites within the unit cell need not have a special relationship (i.e. need not be commensurable) with the Bravais lattice vectors. Therefore, in general, the Bloch Hamiltonian need not be a periodic function of $k$ [83]. In section 5, we will see on the example of the SSH chain that the Bloch Hamiltonian is periodic but with a double periodicity compared to the reciprocal lattice. In section 6, we will see another example: the honeycomb lattice, for which the Bloch Hamiltonian has a triple periodicity compared to the reciprocal lattice [83].

An alternative ‘periodic Bloch Hamiltonian’ is often defined as follows:

$$\mathcal{H}(k) = e^{-ikR}He^{ikR}.$$  \hspace{1cm} (60)

It depends only on the position operator for the unit cell $R$ (i.e. the position on the Bravais lattice) and not on the full position operator $r$. We will write $r = R + \delta$, where $\delta$ is the position operator within the unit cell. The canonical and the periodic Bloch Hamiltonian have the same energy bands $E_n(k)$. This alternative Bloch Hamiltonian is periodic with the reciprocal lattice $\mathcal{H}(k + G) = \mathcal{H}(k)$ unlike the Bloch Hamiltonian $H(k + G) = e^{-ik\delta}H(k)e^{ik\delta}$, which is not, in general. Note also that the canonical Bloch Hamiltonian is unique, while the periodic Bloch Hamiltonian depends on the choice of the unit cell. One should therefore better speak of a, rather than the, periodic Bloch Hamiltonian. This issue of canonical versus periodic Bloch Hamiltonian is crucial in the case of a lattice with a basis (e.g. the honeycomb lattice or the SSH chain). In the literature, it is sometimes known as basis I versus basis II, see [83–86]. The unique (canonical) Bloch Hamiltonian $H(k)$ being that in basis II and the various possible periodic Bloch Hamiltonians $\mathcal{H}(k)$ belonging to basis I.

The eigenvectors of a periodic Bloch Hamiltonian are not the cell-periodic Bloch eigenvectors $|u_n(k)\rangle$. In order to distinguish them, we call them $|v_n(k)\rangle = e^{-ikR}|\psi_n(k)\rangle = e^{ik\delta}|\psi_n(k)\rangle$, where $\delta = r - R$ is the position operator within the unit cell. They also satisfy:

$$\mathcal{H}(k)|v_n(k)\rangle = E_n(k)|v_n(k)\rangle,$$  \hspace{1cm} (61)
and the $v_{n\mathbf{k}}(r)$’s have the periodicity of the Bravais lattice. The $|v_{n\mathbf{k}}(\mathbf{k})\rangle$’s have the periodicity of the reciprocal lattice (at least under the ‘periodic gauge choice’): $|v_{n\mathbf{k}}(\mathbf{k}+\mathbf{G})\rangle = |v_{n\mathbf{k}}(\mathbf{k})\rangle$.

Let us restrict for a moment to a tight-binding model with Hamiltonian $H = \sum_{ij} t_{ij} |i\rangle \langle j|$ and position operator $\mathbf{r} = \sum_{i} r_{i} |i\rangle \langle i|$. The Hamiltonian stores the information about the connectivity between orbitals, that are assumed to form a complete orthogonal set: $\langle i|j\rangle = \delta_{ij}$ and $\sum_{i} |i\rangle \langle i| = \mathbf{1}$. The position operator contains the information on the position of the orbitals in space. The canonical Bloch Hamiltonian should be used to compute geometrical quantities that crucially depend on the spatial location (or spatial embedding, as Haldane calls it) of the orbitals defining the model (i.e. Berry curvature, quantum metric, etc). The periodic Bloch Hamiltonian is more convenient in computing topological invariants such as winding numbers, Chern numbers, symmetry-based indicators (such as that of Fu and Kane etc). The periodic Bloch Hamiltonian incorporates more information about the spatial location of orbitals than the periodic Bloch Hamiltonian. Colloquially speaking, the canonical Bloch Hamiltonian $H(\mathbf{k})$ knows the connectivity contained in the tight-binding Hamiltonian $H$ and the complete position operator $\mathbf{r} = \mathbf{R} + \mathbf{\delta}$. In contrast, the periodic Bloch Hamiltonian $\tilde{H}(\mathbf{k})$ only knows $H$ and the Bravais lattice position operator $\mathbf{R}$ but is unaware of the position operator within the unit cell $\mathbf{\delta}$.

4.2.6. Conclusion

In conclusion of this section, we wish to emphasize that there is no need of introducing Berry phases. One could study the complete quantum mechanical problem of an electron in a band structure in the presence of external fields, without ever projecting on a given subset of bands. The appearance of Berry phase effects is only related to the approximate treatment of restricting to a subset of bands and asking for an effective description within this subspace. In practice, solving the full quantum mechanical problem is often un-doable analytically (but may be done numerically) and one ends up asking for an analytically-tractable effective description restricted to a subset of bands. In such a case, Berry phase effects necessarily appear. For example, the orbital magnetic susceptibility or the electric polarization, which are usually discussed in terms of Berry phases, can be studied using only the energy spectrum numerically-computed in a finite magnetic field (Hofstadter butterfly) [87] or in a finite electric field (Wannier–Stark ladder) [88].

4.3. Geometric phases and consequences

4.3.1. Berry phase

If an electron in the $n$th band performs a closed (and contractible) orbit $C$ in $\mathbf{k}$-space under the influence of a force, it will acquire a geometric phase due to the encircled Berry flux in addition to the dynamical phase. This extra phase is known as the Berry phase [14]:

$$\Gamma_n(C) = \oint_C d\mathbf{k} \cdot \mathbf{A}_n(\mathbf{k}) = \int_S d^2 k F_{xy}^n(\mathbf{k}) [2\pi],$$

(62)

where $\partial S = C$ (here we assumed that the electron is moving in the $xy$ plane). The Berry phase is defined modulo $2\pi$. Using Stokes’ theorem in order to go from the expression involving the connection to that involving the curvature, one needs to assume that the connection is well-defined over the whole patch $S$. When expressed in terms of the Berry curvature, it is obvious that the Berry phase is gauge-invariant. This Berry phase is a dual of the Aharonov–Bohm phase [89] in the sense that it is acquired in $\mathbf{k}$-space (rather than real space) and due to the Berry curvature (rather than to the magnetic field). In the case of a closed cyclotron orbit performed under an applied magnetic field, the electron wave function actually acquires both an Aharonov–Bohm phase (in real space) and a Berry phase (in reciprocal space). See, for example, the discussion of semi-classical quantization of cyclotron orbits for Bloch electrons in [85]. The Berry phase factor $\mathcal{W}(C) = \exp(i\Gamma_n(C))$ is sometimes called an abelian Wilson loop (see section 4.3).

4.3.2. Zak phase

If the Berry phase is computed over a non-contractible loop $P$ of the BZ torus, then it is known as a Zak phase [19]:

$$Z_n(P) = \int_P d\mathbf{k} \cdot \mathbf{A}_n(\mathbf{k}) [2\pi].$$

(63)

Stokes’ theorem can no longer be used to relate it to a Berry curvature and it is not obvious that the Zak phase is gauge-invariant. Here there is no smooth gauge for the Berry connection over the complete path $P$. 
Actually, the Zak phase is only gauge-invariant provided the 'periodic gauge choice' condition is imposed \([19, 52]\). This is an example of open-path geometrical phase, as clearly discussed by Resta \([52]\).

Indeed, the final ket \(|u_n(k_f)\rangle\) in the path is not the same as the initial one \(|u_n(k_i)\rangle\) because the canonical Bloch Hamiltonian is not in general periodic with the first BZ (see the discussion about the canonical Bloch Hamiltonian versus a periodic Bloch Hamiltonian). It is possible to impose a definite phase relation between \(|u_n(k_f)\rangle\) and \(|u_n(k_i)\rangle\) in order to render the Zak phase gauge independent. This phase relation involves the position operator and is known as the 'periodic gauge choice':

\[
|u_n(k_f)\rangle = |u_n(k_i + G)\rangle = e^{-iG \cdot \delta_r} |u_n(k_i)\rangle.
\]

(64)

It does not completely fix the gauge, but only restricts possible gauge choices. As a consequence of this periodic gauge choice, the Zak phase depends explicitly on the position operator \(r\) and therefore on the choice of position origin \([90]\). The Zak phase is therefore best thought as being a particular position within the unit cell known as the Wannier center. The Zak phase is especially relevant in one dimension, where the BZ is a circle (see section 5). If the base space (here the BZ) was not a torus but a sphere or a manifold with a trivial first homotopy group (i.e. no non-contractible loop), then there would be no sense in defining a Zak phase and only the Berry phase would be defined. The Zak phase factor \(W(P) = \exp(\ii \zak(P))\) is sometimes called an abelian Wilson-Zak loop.

The Zak phase should be clearly distinguished from the Berry phase. The former can not be expressed in terms of the Berry curvature and is closely related to the position operator. A convenient habit is to think of the Zak phase as the Wannier center. In contrast, the Berry phase measures the Berry flux across a patch in BZ, has a minor dependence on the position operator (see the above discussion about basis I versus basis II) and does not depend on the choice of position origin. A word of caution to the reader: in many papers, a winding number is mistakenly called a Zak phase.

### 4.3.3. Wannier functions

Here we give a minimal introduction to Wannier functions, which is a whole subject in its own, see \([54, 80]\). They were introduced long ago \([91]\) as the Fourier transform of Bloch states in a given band:

\[
w_{n,R}(r) = \frac{1}{(2\pi)^2} \int \int_{\text{BZ}} e^{-i \mathbf{k} \cdot \mathbf{r}} \psi_{n,k}(r) = \int_{\text{BZ}} \frac{dk}{(2\pi)^2} e^{-i \mathbf{k} \cdot (\mathbf{r} - \mathbf{R})} u_{n,k}(r).
\]

(65)

In 1D, they can also be defined as eigenvectors of the projected position operator onto a given band \([92]\). There is one Wannier function per unit cell and per band. The Wannier states form an orthonormalized basis \(|w_{n,R}\rangle = \delta_{n,n'} \delta(R - R')\). In a given band, one may concentrate on \(w_n(r) = w_{n,0}(r)\) as other Wannier functions in the same band are obtained by translation by a Bravais lattice vector \(w_{n,R}(r) = w_n(r - R)\).

Physically, Wannier functions are the solid-state equivalent of atomic or molecular orbitals and are sometimes called Wannier orbitals. As compared to Bloch states, they are better localized in real space but they are not energy eigenvectors. There is a certain gauge-freedom in their definition (related to the Berry-gauge freedom in the definition of the cell-periodic Bloch states). Two important characterization of Wannier functions are:

- Their average position defined as:

\[
\langle r_{n,R} \rangle = \int dr r |w_{n,R}(r)|^2 = \langle r_n \rangle + R,
\]

(66)

which is gauge-invariant and known as the Wannier (or band) center. One is usually mainly interested in \(\langle r_n \rangle\) which is the Wannier center modulo a Bravais lattice vector \(R\). The Wannier center is related to the electric polarization \([20–22]\) (see section 5.2.4).

- Their localization, which depends on the chosen gauge. The issue of exponential localization (or not) of Wannier functions has a long history starting with Kohn \([93]\) (see e.g. \([94]\) for a discussion of its relation to singularities in Bloch states). Roughly speaking, in a trivial insulator, Wannier functions can be exponentially localized \([95]\); in a Chern insulator, they are only algebraically localized \([96, 97]\); and in a metal, they are delocalized. In particular, Thouless has shown that a non-zero Chern number implies an obstruction in finding an exponentially-localized Wannier function \([96]\). One may characterize the localization by defining the spread or extension \(\langle r_n^2 \rangle - \langle r_n \rangle^2\) of a Wannier function. Playing with the gauge freedom, it is possible to define so-called maximally-localized Wannier functions (MLWF) \([80]\). The gauge-invariant part of the spread of a MLWF is related to the quantum metric \([53, 54]\). The obstruction in finding exponential-localized Wannier functions respecting a given symmetry will turn out to play an important role in the definition of SPT insulators.
4.3.4. Chern number
In space dimension two, the integral of the Berry curvature over the whole BZ torus $T^2$ is quantized:

$$C_n = \frac{1}{2\pi} \int_{T^2} dt^2 k F_{\psi_n}(k) \in \mathbb{Z}.$$  \hspace{1cm} (67)

This integer is called the Chern number and tells whether the corresponding Bloch bundle is twisted or not. For time-reversal-invariant materials, the Chern number is always zero, being the integral of an odd function of $k$ over the BZ. Therefore breaking time-reversal symmetry is a necessary condition to obtain a band with non-zero Chern number, but it is not sufficient, as we will see in the Haldane model. The Chern number can be seen as an obstruction to having a well-defined connection over the whole BZ. Indeed, if a well-defined Berry connection exists over the whole BZ, then the Berry phase computed over the null path should be equal to the total Berry flux across the BZ via Stokes’ theorem and should therefore vanish, i.e. $C_n = 0$. Therefore $C_n \neq 0$ means that there is no well-defined connection over the whole BZ.

An alternative view of the Chern number is the following. A band contact point (or degeneracy) acts as a magnetic monopole in parameter space (a Berry monopole) \cite{14}. Berry calls it a diabolical point because of its diabolo shape \cite{40}. It is also known as a conical intersection. The Chern number counts the number of such degeneracies which are enclosed by the BZ torus, i.e. which are inside the torus. The precise meaning of inside is the following. The band structure has no band degeneracy on the surface of the BZ torus, otherwise the bands would not be well separated, there would be no gap, and the Chern number would not be well-defined. Therefore the band degeneracies that we are talking about actually occur not on the surface of the BZ torus, but really inside a toroid or solid torus \cite{15}. It means that one should imagine extending the 2D $(k_x, k_y)$ model with a third dimension, that we call $k_z$ even it does not correspond to a spatial direction but to some parameter of the Bloch Hamiltonian (such as a hopping amplitude or an on-site energy) that upon tuning creates such a degeneracy. The torus spanned by $(k_x, k_y)$ is now filled inside into a toroid spanned by $(k_x, k_y, k_z)$. We will give a concrete example later in this review when discussing 3D Weyl points in section 8.2. The latter is a contact point between two bands in 3D reciprocal space which is very analogous to the Dirac magnetic monopole.

An equivalent of the Chern number can also be defined at finite temperature \cite{98}. It requires extending the notion of a geometric phase to mixed (i.e. not pure) states and is known as the Uhlmann phase.

4.4. Semi-classical equations of motion of a Bloch electron
4.4.1. Standard equations of motion
In order to discuss the Berry phase effects in the semi-classical equations of motion for a Bloch electron, we first recall the standard equations (see, e.g. \cite{6,99}) which were obtained in the 1930s by Bloch \cite{3}, Peierls \cite{4}, Jones and Zener \cite{5}. For simplicity, we consider the motion of a spinless electron restricted to a non-degenerate band and under the influence of external electromagnetic fields (they can be slightly inhomogeneous but here independent of time). The latter are responsible for the dynamics and also for possible transitions between bands. We assume that the electron stays in a given band (adiabatic following, inhomogeneous but here independent of time). The coupled equations of motion for an electron with average (or center of mass) wave vector $k$ and position $r_c$ in the $n$th band are:

$$\hbar \dot{k} = -\nabla_k \widetilde{E}_n - \dot{r}_c \times \mathbf{B}(r_c) = -e[E(r_c) + \dot{r}_c \times \mathbf{B}(r_c)]$$

$$\dot{r}_c = \frac{1}{\hbar} \nabla_k \widetilde{E}_n = \frac{1}{\hbar} \nabla_k E_n,$$  \hspace{1cm} (68)

where the semi-classical energy:

$$\widetilde{E}_n(k, r_c) = E_n(k) - eA_0(r_c),$$  \hspace{1cm} (69)

is the sum of the band energy and the potential energy of a charge $-e$ in the external electrostatic potential $A_0(r)$. The first equation of the system (68) looks like Newton’s equation for a particle with gauge-invariant momentum $\hbar k$ and electric charge $-e$ under the influence of the Coulomb and Lorentz forces in an electric field $E = -\nabla r A_0(r)$ and in a magnetic field $\mathbf{B} = \nabla \times \mathbf{A}(r)$. The second equation in (68) is the statement that the electron velocity $\dot{r}_c$ is given by the group velocity of the band dispersion relation $E_n(k)$.

It is possible to requantize the above equations and obtain an effective one-band quantum Hamiltonian describing the electron in the $n$th band:

$$H_{\text{eff}} = \widetilde{E}_n(k, r_c) = E_n(k) - eA_0(r_c),$$  \hspace{1cm} (70)
The semi-classical energy is:

\[ \mathbf{k} = \mathbf{q} + \frac{\mathbf{e}}{\hbar} \mathbf{A}(\mathbf{r}_c) \rightarrow -i\nabla \mathbf{r}_c + \frac{\mathbf{e}}{\hbar} \mathbf{A}(\mathbf{r}_c) \]
\[ \mathbf{r}_c = \mathbf{x}, \]  

(71)

where \( \hbar \mathbf{q} \) and \( \mathbf{x} \) are the canonical momentum and position operators, and \( \hbar \mathbf{k} \) is the (electromagnetic) gauge-invariant momentum operator. This is known as the Peierls substitution \cite{100}. To summarize the 'Peierls strategy', one diagonalizes the Bloch Hamiltonian \( H(k) \) in the absence of external fields, projects on a given band to obtain an effective Hamiltonian \( H_{\text{eff}} = E_n(k) \), then introduces external fields in the effective Hamiltonian \( H_{\text{eff}} = E_n(q + \frac{\mathbf{e}}{\hbar} \mathbf{A}) = eA_0(\mathbf{r}_c) \) and eventually requantizes \( q \rightarrow -i\nabla \mathbf{r}_c \). It is then obvious that inter-band transitions induced by the external fields are neglected in this process.

These equations were able to explain and predict many transport phenomena occurring in crystals (e.g. Bloch oscillations, Hall effect, conduction by holes, cyclotron motion). However, in the 1950s and 1960s, it became clear that these equations were not complete because the electron dynamics is actually influenced by the possibility of virtual transitions to other bands driven by the external fields \cite{7}. In other words, while it is possible to render real (Landau–Zener) inter-band transitions vanishingly small (by having small external fields and large gaps), it is not possible to forbid virtual inter-band transitions. In the 'Peierls strategy', in order to describe the effective behavior in a given band, one relies only on the band’s dispersion relation \( E_n(\mathbf{k}) \) (obtained in the absence of external fields) and on no other information (e.g. such as the cell-periodic Bloch states \( |\psi_n(k)\rangle \)). Berry phase effects are essentially the statement that cell-periodic Bloch states do play a role in the effective dynamics, as we now discuss.

4.4.2. Equations of motion including Berry phase effects

The complete (at first order in the external fields) equations of motion were obtained in final form by Qian Niu and coworkers \cite{4, 24, 26, 27} using a wave packet approach. The derivation is quite tedious and we do not reproduce it here. For an alternative Hamiltonian approach (i.e. without wave packets), see \cite{101}. The semi-classical equations of motion for an electron wave packet built from states in the \( n \)th band and having average wave vector \( \mathbf{k} \) and position \( \mathbf{r}_c \) are:

\[ \dot{\mathbf{k}} = -\frac{\mathbf{e}}{\hbar} \nabla \mathbf{E}_n - \dot{\mathbf{r}}_c \times \frac{\mathbf{e}}{\hbar} \mathbf{B}(\mathbf{r}_c) \]
\[ \dot{\mathbf{r}}_c = \frac{\mathbf{e}}{\hbar} \nabla \mathbf{E}_n - \dot{\mathbf{k}} \times \mathbf{F}_n(\mathbf{k}), \]

(72)

where the semi-classical energy is:

\[ \tilde{E}_n(\mathbf{k}, \mathbf{r}_c) = E_n(\mathbf{k}) - eA_0(\mathbf{r}_c) - m_n(\mathbf{k}) \cdot \mathbf{B}(\mathbf{r}_c). \]

(73)

At second order, extra terms appear, some of which are discussed in \cite{102}. Compared to the standard equations (68) and (69), equations (72) and (73) contain two extra contributions (there is also a third hidden modification, related to the phase-space measure and that we discuss below):

One \(-\dot{\mathbf{k}} \times \mathbf{F}_n \) is called the anomalous velocity and depends on the Berry curvature \( \mathbf{F}_n(\mathbf{k}) = \nabla \mathbf{k} \times \mathbf{A}_n \). It was first found by Karplus and Luttinger in a particular case \cite{8}. It is a dual of the Lorentz magnetic force (in \( k \)-space the Berry curvature is the analogous of a magnetic field and \( \dot{\mathbf{k}} \) is the analogous of a velocity) and is at the origin, for example, of the IQHE (see below).

The other extra term \(-m_n \cdot \mathbf{B}\), first obtained by Kohn \cite{9}, is a kind of Zeeman effect and involves another geometrical object (not previously introduced) called the orbital magnetic moment \( m_n(\mathbf{k}) = -\mathbf{\hat{z}} (\mathbf{r} \times (\mathbf{v} - \langle \mathbf{v} \rangle)) \), where \( \mathbf{v} \) is the velocity operator and the average is taken over a wave packet restricted to the \( n \)th band \cite{24}. It has an expression similar to (56) for the Berry curvature:

\[ m_n(\mathbf{k}) = \frac{\mathbf{e}}{2\hbar} \sum_{n' \neq n} \langle u_n | \nabla_k H(\mathbf{k}) | u_{n'} \rangle \times \langle u_n | \nabla_k H(\mathbf{k}) | u_{n'} \rangle \frac{E_{n'} - E_n}{E_{n'} - E_n}. \]

(74)

This Zeeman-like effect is emergent as the electron considered here is spinless. Its emergence is similar to that of the Zeeman effect in the Pauli equation (with the famous \( g = 2 \) factor) when taking the low-energy limit of the 3D Dirac equation and projecting on the positive energy states \cite{7}. Niu and co-workers \cite{24} provide the following picture of this emerging orbital magnetic moment. When restricted to a finite energy band, an electron wavepacket has a minimum size due to the uncertainty principle. This minimum size is similar to the Compton wavelength for the Dirac electron as discussed by Blount \cite{7}. This means that the electron charge is now spread over a finite volume and there is the possibility of self-rotation as in the Uhlenbeck and
Goudsmit picture of the rotating electron as a mechanism for the appearance of the intrinsic magnetic moment. The orbital magnetic moment is a crucial ingredient in the 'modern theory of orbital magnetism', see [24, 25] for review.

It is also possible to generalize the Peierls substitution in order to obtain an effective one-band quantum Hamiltonian including the Berry phase corrections [24]. As before the effective Hamiltonian is given by the semi-classical energy:

\[ H_{\text{eff}} = \tilde{E}_n(k, r_c) = E_n(k) - eA_0(r_c) - m_n(k) \cdot B(r_c), \]  

upon identifying \( k \) and \( r_c \) with operators. However, the canonical quantization is not easy, because \( r_c \) and \( \hbar k \) are not canonical position and momentum: their Poisson bracket is not standard. We consider three cases.

(a) In the particular case in which the Berry curvature vanishes, one uses the standard Peierls substitution (71).

(b) In the particular case where the magnetic field vanishes, these operators are given by a kind of dual of the Peierls substitution:

\[ k = q \]
\[ r_c = x + A_n(k) \to i \nabla k + A_n(k), \]

where \( \hbar q \) and \( x \) are the canonical momentum and position operators, and \( r_c \) is the (Berry) gauge-invariant position operator. It can be shown that \( r_c = A_n(k) + x \) is also the position operator projected onto the \( n \)th band. (In the crystal momentum representation [7], the complete position operator \( r \) also has matrix elements between different bands, that are given by: \( A_{n,n'}(k) = \langle \delta(k) | \nabla k | \delta(k) \rangle \) so that \( r_{n,n'} = \delta_{n,n'} \nabla k + A_{n,n'}(k) = \delta_{n,n'} r_c + (1 - \delta_{n,n'}) A_{n,n'}(k) \). This is the reason for distinguishing between the position operator \( r \) and the projected position operator \( r_c \).) Its average over all wave vectors \( k \) in the BZ, \( r_c = \langle r_c \rangle + R \) is equal to the Wannier center \( \langle A_n \rangle \) modulo a Bravais lattice vector \( R \). Roughly speaking, the Wannier center \( \langle r_c \rangle \) is the electron position within the unit cell and \( R \) is the position of the unit cell. The Wannier center plays an important role in the ‘modern theory of electric polarization’ [20–22], as we discuss below.

(c) In the general case, when both the magnetic field and the Berry curvature are non-zero, it can be shown that the operator identification [24, 101] is:

\[ k = q + \frac{e}{\hbar} A(r_c) + eB(r_c) \times A_n(k) \]
\[ r_c = x + A_n(k), \]

where \( (\hbar q, x) \) are the canonical momentum and position operators. Note that this is not simply (71) together with (76).

There is a third modification of the electron dynamics that is not apparent in the above equations of motion ((72) and (73)), but which is important. It is a consequence of the fact that the wave packet momentum \( \hbar k \) and position \( r_c \) are gauge-invariant (both under an electromagnetic gauge transformation and under a Berry gauge transformation) but are not canonical. Their Poisson bracket (and the corresponding quantum commutator) is not the usual one but is modified by the Berry curvature and the magnetic field. Therefore the volume occupied by a state in the \((r_c, k)\) phase-space is no longer \((2\pi)^D\). In order to take this fact into account, one should modify the phase-space measure as follows [24, 103]:

\[ \frac{d\nu_{dk}}{(2\pi)^D} \to \frac{d\nu_{dk}}{(2\pi)^D} \left( 1 + \frac{e}{\hbar} B \cdot F \right). \]

This modified phase-space density simply comes from the Jacobian in the transformation from the canonical \((x, \hbar q)\) to the gauge-invariant \((r_c, \hbar k)\) position and momentum variables [101] as given in equation (77):

\[ dx dq = d\nu_{dk} \left( 1 + \frac{e}{\hbar} B \cdot F \right). \]

The modified phase-space density affects thermodynamic quantities such as the orbital magnetization and susceptibility [24, 25].

As a side remark, it is easy to recover the behavior of the Berry curvature under time-reversal (58) and space inversion (59) from the anomalous velocity in equation (72). Under time reversal, the velocity, momentum and time-derivative change sign so that \( \mathbf{k} \) does not change, which means that \( \mathbf{F}_n \) must change.
sign. Therefore \( F_n(k) \rightarrow - F_n(-k) \) under time-reversal and if it is a symmetry then \( F_n(k) = - F_n(-k) \). Under space inversion, the velocity and momentum change sign so that \( \dot{k} \) changes sign as well and the curvature must remain. Therefore \( F_n(k) \rightarrow F_n(-k) \) under inversion and if it is a symmetry then \( F_n(k) = F_n(-k) \). From the Zeeman-like term \( - m_n \cdot B \), the same considerations apply to the orbital magnetic moment (74): time-reversal symmetry makes it odd \( m_n(-k) = - m_n(k) \) and inversion symmetry makes it even \( m_n(-k) = m_n(k) \). If both symmetries are present, the Berry curvature and the orbital magnetic moment vanish at every \( k \).

In summary, at first order in the external fields, the equations of motion of a spinless Bloch electron restricted to a single band are given by equation (72) with the effective energy (73) and the relation (77) between gauge-invariant and canonical momentum and position. The extra terms (as compared to the standard equations) vanish in the presence of both time-reversal and inversion symmetries. Compared to the ‘Peierls scheme’, here the restriction to an effective single-band description is done in the presence of external fields and virtual transitions to other bands are allowed, which lead to geometric forces of reaction (74). Only real (Landau–Zener) inter-band transitions are neglected.

4.4.3. Quantized Hall conductivity of a filled band

Here, we specialize to space dimension \( D = 2 \) and time-reversal breaking systems. In their landmark paper, Thouless–Kohmoto–Nightingale–den Nijs (TKNN) have shown that the Hall conductivity of a band insulator is quantized because it is related to a topological number: the sum of the Chern numbers of the occupied bands [12]. They originally considered the case of a two-dimensional electron gas in an applied perpendicular magnetic field. Here, we will consider a slightly different version—the so-called QAHE—by assuming that time-reversal symmetry is broken but no external magnetic field is applied. We therefore consider both translation invariance under a Bravais lattice and time-reversal breaking. We will use the semi-classical equations of motion for an electron restricted to a given band and show that a filled band may nevertheless carry a quantized Hall current.

Consider the semi-classical equations of motion (72) for a single electron in a given band \( n \) in 2D in the presence of an applied electric field but no applied magnetic field. They read:

\[
\dot{k} = - \frac{e}{\hbar} E \\
\dot{r}_c = \frac{1}{\hbar} \nabla_k E_n(k) - \hat{k} \times F_n = \frac{1}{\hbar} \nabla_k E_n(k) + \frac{e}{\hbar} E \times F_n.
\]  
(80)

The first term in the above velocity is the familiar group velocity, while the second term is the anomalous velocity of Karplus and Luttinger [8]. We can easily obtain the average velocity of an electron and deduce the electric current carried by a filled band as:

\[
j_n = (-e) \int_{\mathbb{T}^2} \frac{d^2k}{(2\pi)^2} \dot{r}_c = (-e) \int_{\mathbb{T}^2} \frac{d^2k}{(2\pi)^2} \left( \frac{1}{\hbar} \nabla_k E_n(k) + \frac{e}{\hbar} E \times F_n \right).
\]  
(81)

The first term (group velocity) vanishes after integration over a filled band as is well known (‘a filled band does not conduct electricity’). However, the second term (anomalous velocity) need not vanish and involves the Chern number \( C_n \) of the \( n \)th band:

\[
j_n = - \frac{e^2}{\hbar} E \times \int_{\mathbb{T}^2} \frac{d^2k}{2\pi} F_n = - \frac{e^2}{\hbar} C_n E \times e_z.
\]  
(82)

This means that, provided the Chern number is non-zero, a filled band can nevertheless have a Hall current, which is quantized in units of \( e^2/\hbar \), as the Chern number is an integer. For a band insulator with several filled bands (those with \( n \leq n_f \)), one finds that the Hall conductivity is given by:

\[
\sigma_{xy} = - \frac{e^2}{\hbar} \sum_{n \leq n_f} C_n = - \frac{e^2}{\hbar} n_f.
\]  
(83)

The topological invariant \( n_f \) characterizing the band insulator is called the Hall (or TKNN) number and is given by the sum of the Chern numbers of the filled bands. The Chern number \( C_n \) characterizes a band \( n \), whereas the Hall number \( n_f \) is attached to a band gap and characterizes a band insulator (gap labeling). At this point, we have not yet shown that a band insulator with a non-zero Hall number exists. From symmetry arguments, we know that if we do not break time-reversal symmetry, then \( C_n = 0 \) for all bands and therefore \( \sigma_{xy} = 0 \). Below, we will discuss the Haldane model, which is precisely a band model that breaks time-reversal symmetry and is able to produce a non-zero Chern number for a band (note that breaking time-reversal is a necessary but not sufficient condition).
Figure 5. Quantized Hall effects in insulators. H designates an externally applied magnetic field, M indicates a spontaneous magnetization and SOC stands for spin–orbit coupling. The chirality (clockwise or counter-clockwise) of edge states and their spin projections are also indicated. Time-reversal symmetry is present only in the quantum spin Hall insulator. For a comparison to the associated unquantized Hall effects in metals or semiconductors (Hall effect, anomalous Hall effect and spin Hall effect), see [105].

Thouless has also shown how the Chern number quantizes adiabatic pumping in a 1D band insulator [75]. Actually the 2D quantum Hall system can exactly be mapped on a time-periodic 1D system. Another point to note is that from the perspective of the bulk, the Hall current is carried by all the filled bands. However, in a finite sample with edges, the quantized Hall current is carried by gapless and chiral edge modes as in the usual QHE [104]. In the familiar QHE, the bulk invariant $n_H$ is just the number of filled Landau levels. And the number of chiral gapless modes per edge is also equal to $n_H$. This is a first example of bulk-edge correspondence between a topological invariant in the bulk (number of filled Landau levels) and a topological invariant on an edge (the number of chiral gapless edge modes).

When they occur in insulators, Berry curvature effects are quantized/topological because one integrates over the whole BZ (topological numbers). That is for example the case of the QAHE that occurs in Chern insulators. A typical representative is the Haldane model of graphene [16], that we will treat in detail in section 6.3. Another example is the quantum spin Hall effect (QSHE) that occurs in time-reversal invariant topological insulators with strong intrinsic SOC, see e.g. Kane and Mele’s model [28, 29]. Figure 5 summarizes three kinds of quantized Hall effects.

4.4.4. Berry phase effects in doped semiconductors and metals

We have seen that Berry phase effects, such as the existence of the anomalous velocity, lead to robust quantized responses in insulators. However, the anomalous velocity also exists in doped semiconductors and metals where it induces observable but unquantized effects, such as the anomalous Hall effect [106] in time-reversal breaking compounds (ferromagnetic semiconductors or metals), or the spin Hall effect [107] in time-reversal invariant materials with strong-SOC. These effects are still due to geometric quantities such as the Berry curvature but which are not integrated over the complete BZ, but over a finite portion of the BZ delimited by the Fermi surface, and are therefore not topological but merely geometrical. The anomalous Hall effect [106] and the spin Hall effect [107] are fundamental in the field of spintronics and have important applications. The inverse spin Hall effect is a very convenient method to measure a spin current by a charge signal.

Even more recently the nonlinear electromagnetic responses of non-centrosymmetric crystals have been reinterpreted in the perspective of Berry curvature properties [108–113]. The photogalvanic effect, the second-harmonic generation or the frequency difference generation are captured by various frequency dependencies of the nonlinear second-order susceptibility tensor. At low frequency of the driving field, the rectified current is determined by the scattering time and the intrinsic Berry curvature dipole, which is a measure of the average gradient of Berry curvature of the occupied states [109, 112, 113]. Provided inversion symmetry is broken, this intraband effect is present even in time-reversal invariant materials and has been coined nonlinear Hall effect (or nonlinear anomalous Hall effect depending on the authors). Nonlinear Hall or optical effects are also present in 3D materials, and are especially strong in recently discovered Weyl...
semimetals. Experimentally the nonlinear Hall effect has been measured in the few-layer transition metal dichalcogenides (TMDC) WTe$_2$ [114, 115] and also in 3D Dirac or Weyl semimetals [116]. At higher frequencies, the DC rectified current is an inter-band effect known as the shift current. Interestingly, the frequency integral of the rectified second order conductivity is a purely geometrical quantity that depends neither on the scattering time nor on the band dispersion, but is solely determined by the Berry curvature dipole [113]. In principle, those nonlinear response susceptibilities are not quantized in the metallic regime, but in some particular circumstances they also might be quantized [117].

4.5. Non-abelian Berry phases
If instead of projecting on a single band, one projects on a group of $N$ bands (say $N = 2$ bands) in a model containing more bands (for example because they are degenerate), then one ends up with an emergent SU(2) static gauge field. The main difference is that the structure group is no longer the abelian $U(1)$ but becomes the non-abelian SU(2). The main object in this description is the Wilczek-Zee (or non-abelian Berry) connection:

$$A_{mn}(k) = i(u_n \nabla_k u_m),$$  \hspace{1cm} (84)

which is a matrix-valued version of the (abelian) Berry connection $A_n(k) = i(u_n \nabla_k u_n)$ [118]. The same quantity is called $X_{mn}(k)$ by Blount [7] and $R_{mn}(k)$ by Xiao et al [103]. The matrix corresponding to $A_{mn}(k)$ is noted $\mathcal{A}$ which is both a $D$-component vector (hence in bold) and a $N \times N$ matrix (hence the notation with a curly letter). The corresponding curvature,

$$\mathcal{F} = \nabla_q \times \mathcal{A} - i \mathcal{A} \times \mathcal{A},$$  \hspace{1cm} (85)

becomes a matrix. For a presentation of the semi-classical equations of motion in the case of degenerate bands see [24].

The Berry phase factor becomes a matrix known as the Wilson loop [119]:

$$W(C) = P \exp \left( i \oint_C dk \cdot \mathcal{A} \right),$$  \hspace{1cm} (86)

where $P$ is the path-ordering operator. Its eigenvalues are the non-abelian Berry phase factors. If the path is along a non-contractible loop $P$ in the BZ, then the Zak phase factor of the abelian case gets promoted to a matrix known as the Wilson–Zak loop or large Wilson loop [120]:

$$W(P) = P \exp \left( i \oint_P dk \cdot \mathcal{A} \right).$$  \hspace{1cm} (87)

Its eigenvalues are the non-abelian Zak phase factors.

4.6. Conclusion: Berry phase effects in solids, gauge fields and fiber bundles
To conclude this section, it is interesting to draw an analogy between geometrical band theory, gauge theories such as electromagnetism, and fiber bundles. A useful analogy exists between the geometry of a single band (i.e. of the restriction of band theory to a single band by projection) on the one hand and electromagnetism (gauge theory) on the other hand, see the two first columns in table 1. Furthermore, as understood by Wu and Yang [51], what physicists call gauge structure or gauge theory is what mathematicians call fiber bundles, see the last two columns in table 1. In the present case, the gauge structure is emergent and exists in parameter space. Its origin comes from projecting on a sub-set of bands (typically one band) and reflects the effect of virtual transitions to the other bands. The general mechanism for the appearance of such emergent gauge structure lies in the separation between slow (heavy) and fast (light) degrees of freedom. In the context of band structure, this is the separation between the ‘external’ dynamics of the electron (wave vector $k$) within a band, i.e. the slow motion from unit cell to unit cell, and the ‘internal’ dynamics between bands (band index $n$) and corresponding to the intra-unit cell fast motion. When one wishes to have an effective description for the heavy system only, the reaction of the light system on the heavy system happens through an emergent gauge field. See the general discussion in [68]. In the above effective description, the Bloch wave-vector $k$ plays the role of a parameter, whereas the band index $n$ that of a quantum number.

Inter-band effects are known in the literature under various names, such as Berry phase effects, band coupling, emergent gauge field, Bloch bundle, etc. An important message of the present section is that they give rise not only to global (in parameter space) topological effects but also to interesting local geometrical effects. In other words, there can be interesting geometrical effects already in insulators that are topologically trivial, such as boron nitride. Another important message is that these effects are only present if the bands are
Table 1. Analogy between geometrical band theory (projection on a single band) and electromagnetism in the case of a two-dimensional parameter space (a torus in the band theory case and a sphere in the Dirac monopole case). The last column makes a relation to the mathematical language of fiber bundles. In the left column, $P$ refers to a non-contractible loop over the BZ torus and $C$ to a contractible loop (the surface $S$ is such that $C = \partial S$).

| Geometrical band theory | Electromagnetism (gauge fields) | Fiber bundles |
|-------------------------|---------------------------------|--------------|
| $k$-space ($BZ$ torus $T^2$) | $r$-space (sphere $S^2$ around the monopole $\subset \mathbb{R}^3$) | Base space |
| Phase of $|u_\alpha(k)|$ | Phase of $\psi(r)$ | Fiber |
| $U(1)_{\text{Berry}}$ | $U(1)_{\text{electric charge}}$ | Structure group |
| Berry connection $A_\alpha(k) = \langle u_\alpha| \nabla_k u_\alpha$ | Vector potential $A(r)$ | Connection |
| Projected position $r_\alpha = i\nabla_k + A_\alpha(k)$ | Gauge-invariant momentum | Covariant derivative |
| Berry curvature $F_\alpha^\mu(k) = \nabla_k \times A_\alpha(k)$ | Magnetic field strength $B(r) = \nabla \times A(r)$ | Curvature (Chern class) |
| Berry phase $\Gamma_\alpha(C) = \frac{1}{2\pi} \oint_C dk \cdot A_\alpha(k) [2\pi]$ | Aharonov–Bohm phase $-e \oint \vec{A} \cdot d\vec{r} / h [2\pi]$ | (an)holonomy |
| Zak phase $Z_\alpha = \frac{1}{2\pi} \oint \vec{k} \cdot A_\alpha(k) [2\pi]$ | Does not apply: no non-contractible loops | |
| Band contact point (Berry monopole) | Dirac's magnetic monopole | Singular curvature source |
| TKNN number $C_\alpha = \oint_{\partial S} dk F_\alpha^\mu / (2\pi) \in \mathbb{Z}$ | Monopole charge $-e \oint_{\partial S} \vec{B} / h \in \mathbb{Z}$ | First Chern number |
| Orbital magnetic moment $m_\alpha(k)$ | Generalized wave function (Dirac) | Section |
| Quantum metric $g_\mu^\nu(k)$ | Spin magnetic moment (Dirac equation $\rightarrow$ Pauli equation) | |

coupled: they are absent in a single-band model, and also in models containing several bands that are totally decoupled. Coupling means hopping terms between the corresponding orbitals in a tight-binding model.

5. Lattice Dirac fermions in 1D

In the previous section, we have introduced the key concepts of modern band structure theory, including geometrical (Berry curvature) and topological aspects (Chern number) from a rather general point of view. In this section and the following ones, we provide examples of such effects and emphasize the specificity of each space dimension, starting with 1D systems (section 5) before going on with 2D honeycomb lattices (section 6), topological insulators (section 7) and 3D Weyl semimetals (section 8). Here we discuss the famous SSH model for polyacetylene chains [121, 122], and its generalization to diatomic polymers, the RM model [123], which proved also to be of interest in the study of ferroelectricity [21, 124]. The SSH and RM models describe spinless fermions on a bipartite 1D lattice in the presence of several types of symmetries. Interestingly, the polyacetylene chain realizes a solid state implementation of the Jackiw–Rebbi mechanism for the generation of fractionalized excitations [125, 126].

Here we will use the SSH and the RM models as pedagogical guidelines to understand the emergence of Dirac fermions and topology in 1D electronic band structures. Many features of these models enable to learn a lot about Dirac fermions and topological effects in higher dimensions. For instance the zero energy states of the SSH model generalizes into 1D edge states of topological insulators in 2D, and surface states of 3D topological insulators, although the bulk edge correspondence properties may differ. Moreover, the SSH model has also been extended to superconducting systems by Kitaev which led to the discovery of Majorana end states [127]. Recently, the SSH model was experimentally realized on platforms, like cold atoms, photonic crystals or graphene nanoribbons, which allows a much higher tunability of the parameters, thereby allowing an exploration of the phase diagram and the observation of the interesting features of the SSH model. Among reviews on SSH model, we recommend the historic one [122] and the more recent [61] with illustrations in cold atom systems and photonic systems.

5.1. Su–Schrieffer–Heeger model

Polyacetylene is a 1D polymer consisting in a large number of –CH– monomers. The 2$s$, 2$p_x$, 2$p_y$ orbitals of the C atom, arranged in $sp^2$ hybridization, and the 1$s$ orbitals of H atom form the covalent C–C and C–H bonds and provides its planar structure to the 1D polymer. There is a single remaining $2p_z$ orbital on each C atom whose hybridization with neighboring $2p_z$ orbitals leads to $\pi$-bands. Those $\pi$-bands are the highest occupied and lowest unoccupied molecular orbitals of the polymer, and may also be called conduction and valence bands in the language of solid-state physics. Polyacetylene undergoes a dimerization transition where the gain in electronic energy overcompensates the elastic energy cost for creating the lattice distortion (Peierls instability) [99]. Due to this Peierls instability, polyacetylene is stable in a dimerized form consisting
in alternating shorter and longer C–C bonds (figure 6). The unit cells are labelled by the integer $n$, and each unit cell contains two atomic sites, labelled $A$ and $B$ respectively. The SSH tight-binding Hamiltonian for the $\pi$-bands reads [121]:

$$H_{\text{SSH}} = \nu \sum_n c_{A}^\dagger(n) c_A(n) + w \sum_n c_{B}^\dagger(n) c_A(n + 1) + \text{H.c.},$$

where H.c. means hermitian conjugation, and the operator $c_{i}^\dagger(n)$ creates a spinless fermion in the $p_z$ orbital at site $(n, l)$, with $l = A, B$. The parameter $\nu$ is the intra-cell hopping and $w$ the inter-cell hopping amplitude. The definition of the unit cell is a matter of convention, and therefore which coupling is called intra-cell or inter-cell is arbitrary. The hopping parameters $\nu$ and $w$ are real because there is no external magnetic field or internal magnetic flux. Physically, the largest (resp. smallest) coupling among $\nu$ and $w$, in absolute values, corresponds to the shortest (resp. longest) bond. In polyacetylene, the hopping amplitudes $\nu$ and $w$ are fixed and close to 3.5 eV (strong bond) and 2.5 eV (weak bond) [122]. The average hopping amplitude $(\nu + w)/2 = t \sim 3$ eV in polyacetylene is close to that in graphene.

### 5.2. Band structure and Berryology of the SSH model

We now turn to the band structure of the SSH model which can be handled using two distinct representations, a periodic Bloch Hamiltonian $H(k)$ and the (canonical) Bloch Hamiltonian $H(k)$. Both are presented because they are both useful and complementary. The periodic $H(k)$ is convenient for some purposes, but the standard formula for Berry quantities are only correct in their forms given in section 4, when using the canonical Bloch Hamiltonian $H(k)$.

#### 5.2.1. Bloch Hamiltonians (two bases), band structure and spinors

In the absence of disorder, translation invariance allows one to diagonalize the Hamiltonian with respect to the cell index $n$. The field operators $c_l(n)$ in real space are expanded over operators in reciprocal space $c_l(k)$ as:

$$c_l(n) = \frac{1}{\sqrt{N}} \sum_k e^{ika} c_l(k),$$

where $l = A, B$ is the sublattice index, $k$ the crystal momentum, and $N$ the number of unit cells. The unit cell is chosen to have a unit length $a = 1$ and we choose units such that $h = 1$. After substitution of equation (89) in equation (88), the Hamiltonian becomes:

$$H = \sum_k c_l^\dagger(k) H_{lm}(k) c_l(m),$$

where momentum $k$ is restricted to the 1D BZ $[-\pi, \pi]$ and the Einstein summation over repeated sublattice indexes $(l, m)$ is implied. The SSH Hamiltonian which is a $2N \times 2N$ matrix in real space, becomes a $2 \times 2$ matrix $H(k)$ for each value of $k$. Note that the exponential factors in equation (89) contain only the cell number $n_l$ without any mention about the real position of the sites within each cell. As a result, the Bloch Hamiltonian $H(k)$ is $2\pi$-periodic. This periodic Bloch Hamiltonian can be written as a linear combination of the Pauli matrices $\sigma_x$ and $\sigma_y$ only:

$$H(k) = d_x(k) \sigma_x + d_y(k) \sigma_y,$$

with real coefficients:

$$d_x(k) = \nu + w \cos k \quad \text{and} \quad d_y(k) = w \sin k.$$
The Pauli matrices above act on the sublattice index \((A, B)\) and not on real electronic spin (we assume spinless electrons here). We call \(t = (v + w)/2\) the average hopping amplitude, take units such that \(t = 1\) and call \(\delta = v - w\) the difference in hopping amplitudes. As already discussed, a periodic Bloch Hamiltonian depends on the choice of unit cell. Making the other choice is equivalent to \(\delta \rightarrow -\delta\).

For comparison and later use, one defines also the *canonical* Bloch Hamiltonian (or simply the Bloch Hamiltonian) \(H(k)\). It is obtained by using the following unitary transformation for diagonalizing \(H\):

\[
c_l(n) = \frac{1}{\sqrt{N}} \sum_k e^{i k x_{nl}} c_l(k),
\]

where \(x_{nl} (x_{nl} = n + 0 \text{ if } l = A \text{ and } x_{nl} = n + 1/2 \text{ if } l = B)\) represents the exact position of the type-\(l\) site of the cell with number \(n\). Note that one should have used a different notation for the operators \(c_l(k)\) in order to distinguish (89) and (93). After substitution of equation (93) in equation (88), one obtains the Bloch Hamiltonian:

\[
H(k) = (v + w) \cos(k/2) \sigma_x + (v - w) \sin(k/2) \sigma_y = 2 \cos(k/2) \sigma_x + \delta \sin(k/2) \sigma_y.
\]

Since the exact position of sites is involved in the unitary transformation, the Bloch Hamiltonian \(H(k)\) does not depend on the choice of unit cell, but its periodicity in reciprocal space is altered. Indeed, note that \(H(k + 2\pi) = H(k)\) but \(H(k + 2\pi) = -H(k)\) such that \(H(k + 4\pi) = H(k)\). The Bloch Hamiltonian has a doubled periodicity related to the fact that the distance between the two sites within the unit cell is half of the unit cell size. Later, we will encounter a similar effect in graphene. The Bloch Hamiltonian seems to depend on the sign of \(\delta\). Actually, \(H_{-\delta}(k)\) can be mapped back to \(H_\delta(k)\) by a unitary transform which exchanges \(A\) and \(B\) sites: \(\sigma_x H_{-\delta}(k) \sigma_x = H_\delta(k)\). This relabeling transformation is a do-nothing transformation. One could therefore restrict the study of the infinite SSH chain to \(\delta \geq 0\).

In both representations, we have obtained that \(d_x\) is an even and \(d_y\) an odd function of \(k\) by explicit calculation, but this is also a consequence of the symmetries of the SSH model. Like in any two-band model, the physics is encoded in the functions \(d_x(k)\), and \(d_y(k)\), which are similar to the components of the external magnetic field in the TLS case (section 3), and determine the geometry of the spinors. Note the absence of a \(d_z(k)\) term in the SSH model. We will study the effect of such a term within the RM model in section 5.5.

The electronic band structure of the infinite SSH chain is obtained by diagonalizing equation (91) or equation (94), which leads to the dispersion:

\[
E_{\pm}(k) = \pm |d(k)| = \pm \sqrt{v^2 + w^2 + 4 vw \cos(k)} = \pm \sqrt{4 \cos^2(k/2) + \delta^2 \sin^2(k/2)}.
\]

In the general case \(\delta \neq 0\), the conduction band \((E_+(k) > 0)\) and the valence band \((E_-(k) < 0)\) are separated by a gap, and the chain is therefore insulating at half-filling, see figure 7(left). One may also define an azimuthal angle (along the equator of the Bloch sphere) \(\varphi_k\) reflecting the geometrical/topological properties of the Bloch Hamiltonian beyond the energy spectrum:

\[
H(k) = E_+(k) (\cos \varphi_k \sigma_x + \sin \varphi_k \sigma_y) \text{ and } H(k) = E_+(k) (\cos \phi_k \sigma_x + \sin \phi_k \sigma_y).
\]

In order to clearly distinguish them, we call \(\varphi_k\) the azimuthal angle for the canonical Bloch Hamiltonian \(H(k)\) and \(\phi_k\) that for a periodic Bloch Hamiltonian \(H(k)\). They are both plotted in figure 8.

One may wonder if there are exceptional situations where the spectrum equation (95) becomes gapless. The condition for a gap closing between these two bands requires all coefficients \(d_{ik}(k)\) (for \(i = x, y\) in
equation (92) to vanish simultaneously at the same point \( k \) of the BZ. One has two constraints and only one variable \( k \), which is a typical level crossing/repulsion situation in quantum mechanics. Nevertheless, here it leads to \( d_x = \nu + w \cos(k) = 0 \) and \( d_y = w \sin(k) = 0 \). Therefore, considering solely positive \( \nu \) and \( w \), the gap closing arises for:

\[
k = \pi, \quad \delta = \nu - w = 0.
\]

The SSH model becomes gapless when \( \delta = 0 \), which in fact means that the chain is no longer dimerized. Indeed a monoatomic chain is known to lead to a single metallic band \( E(k) = 2\cos(k/2) \). It can be described by a single band in its natural BZ \([-2\pi, 2\pi]\). Here the two bands in figure 7(middle) correspond to the folding of this metallic band in the BZ of the SSH model \([-\pi, \pi]\) (see figure 7(right)).

Going beyond the energy level description, one considers now the evolution of the stationary states of \( H(k) \) when \( k \) runs over the circular BZ. For each \( k \), the Bloch Hamiltonian has exactly the structure of the two-level Hamiltonian of section 3.1 with \( d_z = 0 \) and therefore \( \theta_k = \pi/2 \). The spinors (cell-periodic Bloch states) are located along the equator of the Bloch sphere and read:

\[
|u_+(k)\rangle = \frac{1}{\sqrt{2}} \left( e^{i\phi_+} \right) \quad \text{and} \quad |u_-(k)\rangle = \frac{1}{\sqrt{2}} \left( e^{i\phi_-} \right),
\]

which are well defined for all \( k \). The ket \( |u_+(k)\rangle \) represents the excited state associated with the positive energy \( E(k) \) (conduction band), while the eigenstate \( |u_-(k)\rangle \) is the ground state with negative energy \( -E(k) \) (valence band).

5.2.2. Symmetries
The symmetries of the SSH model translate into constraints fulfilled by its Bloch Hamiltonian. It turns out that, for the SSH model (but this is not a general property), these constraints are the same whether expressed on the canonical or a periodic Bloch Hamiltonian.

The time-reversal symmetry \( T \) for spinless fermions is expressed by:

\[
T : \mathcal{K} \mathcal{H}(k) \mathcal{K} = \mathcal{H}(k)^* = \mathcal{H}(-k),
\]

where \( \mathcal{K} \) means complex conjugation. This invariance is valid because the hopping parameters \( \nu \) and \( w \) are real. The chiral symmetry reads:

\[
S : \sigma_z \mathcal{H}(k) \sigma_z = -\mathcal{H}(k).
\]

The chiral symmetry is equivalent to the absence of \( \sigma_z \) in \( \mathcal{H}(k) \). The chiral symmetry, also called sublattice symmetry, \( S \) involves only \( \mathcal{H}(k) \), and not a relation between periodic Bloch Hamiltonians \( \mathcal{H}(k) \) and \( \mathcal{H}(-k) \) at opposite quasimomenta \( k \) and \(-k\). For each state \( |\Psi_k\rangle \) with energy \( E \), there is a state \( \sigma_z |\Psi_k\rangle \) with opposite energy \(-E\), which explains the electron/hole symmetry between the conduction and valence bands equation (95). The RM term \( \Delta \sigma_z \) would break this chiral symmetry as it does not respect bipartiteness. The charge conjugation appears to be a combination of time-reversal, which involves complex conjugation, and sublattice symmetry which involves flipping the sign of energy \( C = ST \):

\[
C : \sigma_z \mathcal{H}(k)^* \sigma_z = -\mathcal{H}(-k).
\]

The SSH model is also a condensed matter realization of the quantum field theory concept of charge conjugation. In particle physics, charge conjugation is the interchange of particles and antiparticles. In polyacetylene, there are no genuine positrons of course but the electronic states splits into electron and hole states. In a half-filled chain, hole excitations can propagate with the same parameters of the electron excitation, except for the opposite electric charge. Formally it corresponds to a map between the positive energy solutions and the negative energy solutions of the model.
Finally, the SSH model has an additional symmetry which is the invariance under space inversion, which implies the exchange of $A$ and $B$ isospin index, $I : \epsilon_A \to \epsilon_B$ and $I : \epsilon_R \to \epsilon_A$, as the inversion center is mid-bond and not on-site. The inversion symmetry reads:

$$I : \sigma_z \mathcal{H}(k) \sigma_z = \mathcal{H}(-k).$$

(102)

In the ten-fold classification of topological insulators [41, 42, 128, 129], it would therefore appear that the SSH model belongs to the BDI class of symmetry (i.e. time-reversal symmetry TRS with $T^2 = +1$, particle-hole symmetry PHS with $C^2 = +1$ and therefore chiral or sublattice symmetry $SLS = 1$) which is characterized by a $\mathbb{Z}$ topological index (winding number). Actually, we will see that the bulk winding number has no precise meaning here and that the SSH model is best described as an inversion-symmetric 1D band insulator characterized by a $\mathbb{Z}_2$ invariant.

5.2.3. Winding number and edge modes
For each $k$, the periodic Bloch Hamiltonian $\mathcal{H}(k)$ (and hence its two eigenstates) is represented by the unit vector $\mathbf{d}(k) = |\mathbf{d}(k)|$ on the Bloch sphere $S^2$. Besides explaining the electron/hole symmetry between the conduction and valence bands equation (95), the chiral symmetry also constrains the tip of this unit vector $\mathbf{d}(k)$ to stay along the equator of the Bloch sphere when $k$ runs over the BZ. The 1D BZ has the structure of the circle $S^1$ since its extremities $k = \pm \pi$ represent the same state. The mapping from the circle (BZ) to the circle (equator) allows one to define a winding number $n_w$, which is related to the first homotopy group of the circle $\Pi_1(S^1) = \mathbb{Z}$. The relative integer $n_w$ is the winding number of $\mathbf{d}(k)$ around the origin. For our initial choice of unit-cell convention, when $w < v$, the origin is outside the circle traced by $\mathbf{d}(k)$, so $n_w = 0$ (see figure 9(left)). In contrast, when $w > v$, the origin is inside the circle and $\mathbf{d}(k)$ winds exactly once around the origin: $n_w = 1$ (see figure 9(right)). This winding number $n_w$ is protected by the chiral symmetry of the SSH chain, namely $\sigma_z \mathcal{H}(k) \sigma_z = -\mathcal{H}(k)$ (this periodic Bloch Hamiltonian $\mathcal{H}(k)$ anticommutes with $\sigma_z$).

There is an ambiguity in the assignment of $n_w = 0$ or $n_w = 1$ to one phase or the other, namely to a definite sign of the parameter $v - w$. For a genuinely infinite system, one cannot determine whether the system is in a trivial or non-trivial phase based on $n_w$ if one is not aware of the chosen unit cell. The only unambiguous statement is that this winding number changes by $\pm 1$ when going from one dimerization to the other. This ambiguity is lifted for finite chains, because it is then possible to state whether the chain starts (or ends) by a weak or a strong bond and there is a natural choice of unit cell. There is a relation between this winding property and the presence of protected zero modes at an edge or a domain wall [130, 131]. It is an example of the Jackiw and Rebbi mechanism (see below).

In contrast, the staggered on-site potential $\Delta \sigma_z$ of the RM model breaks explicitly the chiral symmetry, and the tip of $\mathbf{d}(k)$ is now free to follow any loop on the whole Bloch sphere. Since all such unconstrained loops can be smoothly deformed and reduced to a point, there is no possibility to define a winding number for the RM model (at least when inversion symmetry is not imposed, see below).
In summary, the winding number is useful in discussing open chains with edges or domain walls and the presence of protected edge modes. However, as a bulk invariant, it is meaningless as it depends on the unit cell choice.

5.2.4. Zak phase, position operator, Wannier center and electric polarization

A more interesting quantity is the Zak phase [19]. It is a kind of Berry phase defined for a given band but computed along a non-contractible path in the BZ thanks to its torus shape. In 1D, it reads:

\[ Z_n = \int_{-\pi}^{\pi} dk A_n(k) = \int_{-\pi}^{\pi} dk \langle u_n(k) | i \partial_k u_n(k) \rangle, \]  

(103)

for the \( n \)-th band. This bulk quantity is measurable (and was actually measured [132]) and related to the electronic contribution to the polarization [23]. The Zak phase should be computed using the canonical Bloch Hamiltonian \( H(k) \) (and not the periodic one \( H(\bar{k}) \)), as it is related to the projected position operator \( x_\bar{k} = i \nabla_\bar{k} + A_\bar{k}(k) \). It is best thought as being an average electron position (known as the Wannier or band center) within the unit cell [19]:

\[ \langle x_- \rangle = \int_0^1 dx |w_n(x)|^2 = \frac{Z_-}{2\pi}, \]  

(104)

modulo \( a = 1 \), where

\[ w_n(x) = \int_{-\pi}^{\pi} dke^{ikx} u_{nk}(x) = \int_{-\pi}^{\pi} d\psi_{nk}(x), \]  

(105)

is the Wannier function of the \( n \)-th band (here positioned in the unit cell at \( R = 0 \) [80]). Because the Wannier center is a position, it continuously depends on the choice of position origin. Therefore the Zak phase continuously depends on the choice of position origin. Also, it is actually an open-path geometrical phase and therefore becomes gauge-invariant only upon imposing a definite phase relation between initial and final states [19, 52]. Indeed, on the non-contractible path from \( k = -\pi \) to \( k = +\pi \), the final state \( |u_n(\pi)\rangle \) is not the same as the initial state \( |u_n(-\pi)\rangle \); the path is open in Hilbert space, although it is closed in parameter space, i.e. in the BZ. In order to give a gauge-invariant quantity, the Zak phase should be computed under the following restriction called the ‘periodic gauge choice’ [19, 52] \( |u_n(k + G)\rangle = e^{-ikG}|u_n(k)\rangle \), where \( x \) is the position operator, i.e. \( |u_n(\pi)\rangle = e^{-i2\pi x}|u_n(-\pi)\rangle \). The Zak phase being position origin-dependent, computed along a non-contractible path and an open-path geometrical phase subject to a periodic gauge condition, it is truly different from a Berry phase. The two should be carefully distinguished.

The 1D Berry connection defined from the Bloch spinor \( |u_- (k)\rangle \) (ground state or valence band) reads:

\[ A_- (k) = i |u_- (k) \rangle \langle \partial_k u_- (k) | = \frac{1}{2} \frac{d\varphi_k}{dk}, \]  

(106)

and can be easily integrated along a non-contractible closed path in the BZ to give the Zak phase:

\[ Z_- = \int_{BZ} A_- (k)dk = \frac{1}{2} \int_{k = -\pi}^{k = +\pi} d\varphi_k = \frac{1}{2} (\varphi_+ - \varphi_-) = \frac{\pi}{2} \text{sign} \frac{w - v}{w + v} = \frac{\pi}{2}, \]  

(107)

The Zak phase is also related to the electronic contribution to the polarization:

\[ P_{el} = -e \langle x_- \rangle = -e \frac{Z_-}{2\pi} = \pm \frac{e}{4}, \]  

(108)

modulo the electron charge \( e \) [23]. In the following, we take units such that \( e = 1 \) and it therefore seems that the polarization quantum [133], i.e. the modulo in the electric polarization, is \( P_{el} = 1 \). But this is not correct. Actually, the only meaningful electric polarization is defined for a charge-neutral crystal and should therefore include the contribution of the ions:

\[ P_{tot} = P_{el} + P_{ions}. \]  

(109)

In order to have a charge-neutral crystal, we assume that each ion (\( A \) or \( B \)) carries a \( +1/2 \) charge. As another consequence of the presence of the ions (or of the rigid lattice of sites), it will turn out (see section 5.5) that the polarization is actually defined modulo \( P_2 = 1/2 \) and not \( P_2 = 1 \).

In conclusion of these short sections on the winding number and the Zak phase, we remark the following. The winding number is very often mistaken for the Zak phase. They should be clearly
distinguished. The winding number is computed using a periodic Bloch Hamiltonian, whereas the Zak phase should be computed using the canonical Bloch Hamiltonian, as it is related to the position operator (see the above discussion). On the one hand, the winding number depends on the choice of unit cell and is therefore only well-defined when the choice of unit cell is fixed, e.g. because of a boundary [132] or at a domain wall between two dimerizations. On the other hand, the Zak phase requires using the periodic gauge choice, it continuously depends on the position origin (it is best thought as being proportional to the Wannier center) and it is evaluated along a non-contractible loop in BZ.

5.3. Linearization and effective Dirac Hamiltonian

In the generic case, $\delta \neq 0$, the SSH model is gapped. The gap is direct, located at $k = \pi$, and its magnitude is $2|\delta|$. Therefore the sign of $\delta$ is unimportant for the energy spectrum. In contrast, this sign, and the related band inversion, is important for the wave functions, and the topological properties.

A first way to understand this consists in analyzing the SSH chain in the continuous limit, namely at length scales exceeding the lattice spacing between cells $a = 1$. We consider the SSH chain with a narrow gap $|\delta| \ll 1$, close to the semi-metallic regime. We can expand this periodic Bloch Hamiltonian $H(k)$ around $k = \pi$ by writing $k = \pi + q$, with $q \ll 1$:

$$H(k = \pi + q) = H_D(q) = [v + w \cos(q)] \sigma_x + w \sin(q) \sigma_y \approx -q \sigma_y + \delta \sigma_x,$$

with $\delta = \nu - w$ and $w = 1 - \delta/2 \approx 1$. This has the form of a 1D Dirac Hamiltonian.

The Dirac equation was originally invented to describe relativistic electrons in three dimensional space [46] but it can be generalized to any space dimension $D$. Generally speaking, the Dirac Hamiltonian reads:

$$H_D = cp \cdot \alpha + mc^2 \beta,$$

and involves $D + 1$ anti-commuting matrices $\beta, \alpha_1, \ldots, \alpha_D$ that square to one (the so-called Clifford algebra). Here $p$ is the momentum operator and $\alpha = (\alpha_1, \alpha_2, \ldots) = (\alpha_x, \alpha_y, \ldots)$. In 3D, the matrices are $4 \times 4$, but in 2D and 1D, $2 \times 2$ matrices are possible. The Dirac Hamiltonian depends on two parameters, which are the velocity of light $c$ and the electron mass $m$. When it emerges in the low-energy limit as an effective description of a lattice model in solid-state physics, the velocity $c$ and mass $m$ are effective parameters that have no simple relation to the velocity of light or the electron mass. In the present context, the two Dirac matrices are $\alpha_x = -\sigma_y$ and $\beta = \sigma_x$, and they satisfy a Clifford algebra as the Pauli matrices anti-commute and square to one. The velocity is $c = tu = 1$ (it is usually called Fermi velocity $v_F$ in solid-state physics) and the mass is $m = \delta$. The band structure is that of a massive Dirac cone:

$$E(q) = \pm \sqrt{q^2 + \delta^2},$$

that becomes massless $E(q) = \pm |q|$ when $\delta \to 0$ (see figure 7).

This effective Dirac Hamiltonian is useful to analyze the band inversion at $k = \pi$. If we sit exactly at $q = 0$, the Hamiltonian is simply the mass term $\delta \sigma_x$, and the stationary states reduce to the eigenstates of the Pauli matrix $\sigma_x$, denoted $|\sigma_x = 1\rangle$ and $|\sigma_x = -1\rangle$ respectively. The ground state is $|\sigma_x = -1\rangle$ if $\delta > 0$, and switches to $|\sigma_x = 1\rangle$ when $\delta < 0$. Therefore this local analysis shows that the parameter $\delta$ drives a band inversion.

In the limit $\delta = 0$, the system is gapless and described by a 1D massless Dirac equation. As both chiralities are present (left and right movers), this is not a 1D Weyl equation despite the masslessness.

5.4. Jackiw–Rebbi mechanism and charge fractionalization in Dirac insulators

In the low-energy effective description in terms of a Dirac Hamiltonian (110), let us consider that the mass $m = \delta$ is $x$-dependent and changes sign at $x = 0$. This is a domain wall in the Dirac mass. In the seminal paper by Jackiw and Rebbi [125], this mass was provided by the coupling of massless fermions to a bosonic background, and such a bosonic field was assumed to have a topological defect, called a kink. In solid-state physics, polyacetylene provides such a situation at any domain wall junction between a $\delta < 0$ and $\delta > 0$ chain (the two possible dimerizations) [121, 132]. This is a topological defect in the dimerization. The mass kink $m(x)$ breaks translation invariance and one should substitute $q$ in equation (110) by $-i\partial_x$, yielding the first-order differential equation (obtained after multiplication by $i\sigma_x$):

$$i\sigma_y \frac{\partial \Psi}{\partial x} + m(x)\sigma_x \Psi = E \Psi,$$

(113)
where \( m(x) \) is a mass profile varying from \( m(x = -\infty) = -m_0 \) to \( m(x = \infty) = m_0 > 0 \). Since, the gap changes sign between \(-\infty\) and \( \infty \) it has to close somewhere and it is natural to investigate the possibility of zero energy states. Such a zero energy state should satisfy the equation:

\[
\frac{\partial \Psi}{\partial x} = m(x) \sigma_2 \Psi,
\]

which is equivalent to two uncoupled first order differential equations for the \( A \) and \( B \) components of the wave function respectively. The fact that the mass has opposite signs at \( x = \pm \infty \) allows one of these equations to have bounded solutions. For a step in \( m(x) \) with no sign inversion, there is no zero-energy bound state because on both sublattices the solution would grow exponentially on one side of the line or the other. Let us first consider a sharp mass profile \( m(x) = m_0 \Theta(x) - m_0 \Theta(-x) \) where \( \Theta(x) \) is the Heaviside step function. For positive \( m_0 \), the zero energy bound state reads:

\[
\Psi(x) = (\Theta(x) e^{-m_0 x} + \Theta(-x) e^{m_0 x}) |\sigma_2 = -1\rangle,
\]

which is a state completely localized (or polarized) on the \( B \) sublattice. For the choice \( m_0 < 0 \), the bounded solution would be polarized on the \( A \) sublattice. Upon increasing \( m_0 \), the bound state gets more and more localized around \( x = 0 \) over a typical length given by \( \xi = 1/m_0 \). For a general kink (smooth kink), the general solution of equation (113) reads:

\[
\Psi(x) = \left(e^{-\int_0^x dx' m(x') \sigma_2}\right) |\sigma_2 = \sigma\rangle \quad \forall x \in \mathbb{R},
\]

where the sublattice polarization is determined by \( \sigma = -\text{sign}(m_0) \). The zero energy state is localized near the location where the mass \( m(x) \) changes sign, and it is stuck between two insulators on both sides.

The zero-energy states are protected by the chiral symmetry, which means that any perturbation anticommuting with \( \sigma_2 \) will preserve the zero energy level. For instance, disorder in the hopping parameters \( v \) and \( w \) will preserve the zero-energy states provided the gap remains. In contrast, switching on a staggered on-site energy \( \pm \Delta \) breaks the chiral symmetry and splits the zero-energy level in two levels at finite energies \( \pm \Delta \).

If the SSH chain is cut on a strong bond, then the edge hosts a dangling bond. This mid-gap state sits exactly at zero energy and can be empty or occupied by an electron. If it is empty, then the edge has an excess charge \(+1/2\) (that of the isolated cation at the edge). If it is occupied, then the charge turns negative as cation + electron = \(+1/2 - 1 = -1/2\). This is the simplest example of a very general mechanism for charge fractionalization discovered by Jackiw and Rebbi [125]. Here, starting with spinless electrons, one obtains emergent excitations at zero energy and with charge \(+1/2\). The general recipe is: take a Dirac equation, make its mass spatially inhomogeneous and change its sign by inserting a topological defect (a domain wall in 1D, a vortex in 2D, etc) and the result is a trapped zero-energy mode localized on the defect. See the short review by Jackiw [135].

In this section, we have seen that the SSH chain has localized zero-energy modes either on a domain wall (mass kink between two dimerizations) or at an edge provided the cut occurs on a strong bond. In solid-state physics, it means that any insulator that has a low-energy description in terms of a Dirac equation (i.e. we could call it a Dirac insulator) may host such a topologically-protected zero-mode trapped on a topological defect in the Dirac mass. Note that this is independent of the fact that this Dirac insulator is, in addition, a trivial or a topological insulator. In other words: topologically protected zero-mode on a defect is not the same thing as topological band. The former is a local property in the vicinity of a real-space defect, the latter is a property involving the whole band, the whole BZ and not only the low effective description in the vicinity of the gap. For example, boron nitride is a Dirac insulator but it is topologically trivial (the two valleys carry opposite Berry curvature, see section 6.2). This is also the case of the SSH chain (see section 5.5). In contrast, the Haldane model in graphene is also a Dirac insulator but this time it can be a topological (Chern) or trivial insulator depending on its parameters (see section 6.3).

5.5. Inversion-protected topological insulator

The SSH chain is often taken as the simplest example of a one-dimensional topological insulator. Indeed, it is a two-band insulator when half-filled. However, its two ‘phases’ characterized by the sign of \( \delta \) are actually the same, as is obvious by realizing that a shift of the infinite chain by half a unit cell makes \( \delta \rightarrow -\delta \). Therefore, in the bulk, there is not a trivial phase (say \( \delta > 0 \)) and a topological phase (say \( \delta < 0 \)) as a naive analysis of the winding number \( n_w \) would suggest (see figure 9). In addition, zero-energy edge states are not present for
any cut of the chain but only when the cut occurs on a strong bond. Also it is impossible to define a genuine bulk topological invariant that would distinguish the two phases (for example, neither the winding number $n_w$ nor the Zak phase qualify as a well-defined topological invariant). These three facts make the SSH model a poor example of a topological insulator. Nonetheless, it is a good example of a Dirac insulator featuring zero-modes trapped on a topological defect (see section 5.4 on the Jackiw–Rebbi mechanism).

5.5.1. Rice–Mele model with inversion symmetry

However, there is one way in defining a proper 1D topological crystalline insulator [136], i.e. a band insulator whose topology is protected by a point group symmetry. We have already noted that the SSH chain has an inversion symmetry and that the inversion center is mid-bond (at equal distance between two sites $A$ and $B$) and not on-site. Another type of 1D two-band insulator with inversion symmetry is a regular tight-binding chain (as the SSH chain with $\delta = 0$) but with staggered on-site energies $E_A = \Delta$ and $E_B = -\Delta$. When half-filled, this gapped two-band model realizes a band insulator of the CDW type. It has inversion symmetry but the inversion center is now on-site (rather than mid-bond).

In order to realize a phase transition between these two types of inversion-symmetric insulators, we consider the RM model [123] having both the SSH dimerization $\delta$ and the CDW staggered on-site energy $\Delta$. The (canonical) Bloch Hamiltonian reads:

$$H(k) = 2\cos(k/2)\sigma_x + \delta \sin(k/2)\sigma_y + \Delta \sigma_z.$$  \hfill (117)

Following [21, 24, 88], we introduce an angle $\theta_{RM}$ such that $\cos \theta_{RM} = \Delta/M$ and $\sin \theta_{RM} = \delta/M$, where $M = \sqrt{\Delta^2 + \delta^2} > 0$, in order to parametrize the RM model. However, instead of considering the full range of parameters, we restrict them such that the model has inversion symmetry. On-site inversion symmetry means:

$$H(k) \rightarrow H(-k) = 2\cos(k/2)\sigma_x - \delta \sin(k/2)\sigma_y + \Delta \sigma_z = H(k),$$  \hfill (118)

which only occurs if $\delta = 0$ and mid-bond inversion symmetry means:

$$H(k) \rightarrow \sigma_z H(-k) \sigma_z = 2\cos(k/2)\sigma_x + \delta \sin(k/2)\sigma_y - \Delta \sigma_z = H(k),$$  \hfill (119)

which is possible only if $\Delta = 0$. Therefore, we choose $(\delta, \Delta)$ to be $(\neq 0, 0)$ i.e. $\theta_{RM} = \pi/2$ or $3\pi/2$ (SSH); or $(0, 0)$ (gapless); or $(0, \neq 0)$ i.e. $\theta_{RM} = 0$ or $\pi$ (CDW). Inversion symmetry imposes that only two out of the three Pauli matrices appear simultaneously in the Bloch Hamiltonian, which means that this model actually has a chiral symmetry. Indeed, when $\Delta = 0$, $H(k)$ anticommutes with $\sigma_y$ and when $\delta = 0$, it anticommutes with $\sigma_x$ (in the latter case, chiral symmetry is no longer related to the two sublattices and bipartiteness). This means that when $k$ spans the BZ, the Bloch Hamiltonian is restricted to move on a great circle of the Bloch sphere. When $\Delta = 0$, this great circle is the equator ($xy$ plane), and when $\delta = 0$, it is a meridian ($xz$ plane).

We also introduce the dimensionless parameter $\lambda$ such that $\Delta = \lambda M \Theta(\lambda)$ and $\delta = -\lambda M \Theta(-\lambda)$, where $\Theta(x)$ is the Heaviside step function. When $\lambda < 0$, the model is an SSH chain with positive $\delta$ and when $\lambda > 0$ it is a CDW chain with positive $\Delta$. When $\lambda = 0$ it is gapless. When $\lambda$ varies from negative to positive, a phase transition occurs between two different types of band insulators having inversion symmetry. At the transition, the gap closes and the low-energy physics is described by a massless Dirac equation (see figure 13).

Both phases (SSH and CDW) are Dirac insulators. Indeed at low energy, they are both described by a massive Dirac equation in the SSH case, the Dirac Hamiltonian is given in (110). In the CDW case, it is:

$$H(k = \pi + q) \approx -q\sigma_y + \Delta \sigma_z.$$  \hfill (120)

Beware that the Dirac Hamiltonian is here obtained from linearizing the periodic (and not the canonical) Bloch Hamiltonian $H(k)$. The Dirac mass is not the same in the two cases: $\delta \sigma_x$ versus $\Delta \sigma_z$. However, a crucial point is that, in each phase, only two Pauli matrices appear, otherwise this would not be Dirac matrices satisfying the Clifford algebra in 1D. It is interesting to note that the Jackiw–Rebbi mechanism therefore applies to both phases as well with either a domain wall in $\delta(x)$ or in $\Delta(x)$. There is a subtlety involved here in the fact that the Jackiw–Rebbi mechanism strictly applies only in the continuum limit in which $\sigma \rightarrow 0$ and $t \rightarrow \infty$ such that the velocity $ta$ remains finite. In the tight-binding model, there may be surprises as found in the corresponding 2D case (boron nitride) [137].
Wannier centers are to the left or to the right of \( P \) on the bonds, half-way between King-Smith, Vanderbilt and Resta \([i.e.]\) in a inversion-symmetric band insulator. This is a inversion is a symmetry it means that is obviously defined modulo 1 as follows. Consider the CDW situation with the electron localized on the electric polarization has no meaning. In addition, the complete polarization ions \( A \) and \( B \) \( \text{and negative charges} \) \( +1/2 \) \( -1/2 \) \( 0 \) \( \pm \) \( A \) \( \text{ions} \) \( \text{carry a charge} \) \( \text{charge} \) \( \text{a half filled two-band chain, it means that there is a single} \) \( \text{electron per unit cell carrying a charge} \) \( -e = -1 \). In order to ensure electric neutrality, we assume that the ions \( A \) and \( B \) carry each a charge \( +1/2 \) (they are cations). It is crucial to consider a neutral system, otherwise the electric polarization has no meaning. In addition, the complete polarization is defined modulo \( P_\text{tot} = 0 \) \( \text{(see e.g. the pedagogical discussion about the polarization quantum} P_\text{q} \) \( \text{in} \) \( \text{[138]} \). This can be seen as follows. Consider the CDW situation with the electron localized on \( A \) sites. The total charge is therefore \( +1/2 - 1 = -1/2 \) on \( A \) \( \text{sites and} \) \( +1/2 \) \( \text{on} \) \( B \) \( \text{sites. Now, if we move every electron by a distance of} \) \( 1/2 \) \( \text{to the right we obtain a state in which every} \) \( A \) \( \text{site has a charge} \) \( +1/2 \) \( \text{and every} \) \( B \) \( \text{site a charge} \) \( -1/2 \). But up to a translation by half a unit cell, these two states are identical (they correspond to the RM model with \( \theta_{\text{RM}} = 0 \) \( \text{and} \) \( \pi \). Because we have translated a single electron of charge \( -1 \) \( \text{by a distance of} \) \( 1/2 \) \( \text{in order to be back on the same state, it means that the polarization is defined modulo a quantum of} \) \( 1 \times 1/2 = 1/2 \). A subtle issue is the fact that the electronic contribution to the polarization reads \( P_\text{q} = -e \langle x_- \rangle = -2 \pi \) \( \text{see below}, \) \( \text{which is obviously defined modulo} \) \( 1 \) \( \text{as} Z_\pi \text{ is a phase. However, the complete polarization} P_\text{tot} \text{is defined modulo} P_\text{q} = 1/2 \). Next, we remark that under space inversion, the total polarization \( P_\text{tot} \rightarrow -P_\text{tot} \) and because inversion is a symmetry it means that \( -P_\text{tot} = P_\text{tot} \text{modulo} P_\text{q} \). Therefore

\[
P_\text{tot} = 0 \text{ or } P_\text{q}/2 \text{ modulo } P_\text{q}.
\]

in a inversion-symmetric band insulator. This is a \( \mathbb{Z}_2 \) invariant.

The total electric polarization \( P_\text{tot} \) has two contributions: that of ions and that of electrons. The former is computed classically as \( P_\text{ions} = \sum_j q_j x_j = (x_A + x_B)/2 = x \), where \( q_j = +1/2 \) is the ion charge. \( P_\text{ions} \) depends on the average position \( x \) of ions within the unit cell. This contribution can be made to vanish by properly choosing the position origin on a bond, half-way between \( A \) and \( B \) (this is an inversion center for the SSH chain), see figure 10. The electronic contribution can be easily computed thanks to the prescription given by King-Smith, Vanderbilt and Resta [23]: do as if the electron charge was entirely localized at the Wannier center of the occupied band \( \langle x_- \rangle = -2 \pi / (2 \pi) \). In the SSH phase, the Wannier centers are localized on the bonds, half-way between \( A \) and \( B \) sites (see figure 11). The sign of \( \delta \) only decides whether these Wannier centers are to the left or to the right of \( A \) sites (they are on the strong bonds). For such a chain, the total electric polarization (including the electrons and the ions) is \( P_\text{tot} = 0 \) modulo 1/2. The Zak phase is 0, \( \text{i.e.} \langle x_- \rangle = 0 \), or \( \pi \), \( \text{i.e.} \langle x_- \rangle = 1/2 \) \( \text{depending on the sign of} \) \( \delta \) \( \text{and when computed from the inversion center, which is on a bond, can be either chosen} \) \( \text{to the left or to the right of an} \) \( A \) \( \text{site} \). In the CDW phase, the Wannier centers are located on-sites. The sign of \( \Delta \) only decides whether they are located on an \( A \) site (if \( \Delta < 0 \)) or a \( B \) site (if \( \Delta > 0 \)). The total electric polarization here is \( P_\text{tot} = 1/4 \) \( \text{modulo} \) \( 1/2 \). The Zak phase is also 0 or \( \pi \) when computed from the inversion center which is now on-site. The value 0 or \( \pi \) depends on the

---

**Figure 10.** Rice–Mele chain showing the dimerized hopping amplitudes \( v = 1 + \delta/2, w = 1 - \delta/2 \) and the staggered on-site energies \( P_{\text{tot}} = \pm \Delta \). The position origin \( x = 0 \) is chosen on a mid-bond and corresponds to an inversion center for the SSH chain \( (\Delta = 0) \). An inversion center for the charge density wave (CDW) case \( (\delta = 0) \) is on site, e.g. at \( x = -1/4 \).

**Figure 11.** The Rice–Mele chain with \( \theta_{\text{RM}} = 0 \) (CDW), \( \pi/2 \) (SSH), \( \pi \) (CDW) and \( 3\pi/2 \) (SSH) has inversion symmetry. The \( A \) (green) and \( B \) (red) ions carry \( +1/2 \) charge. The Wannier function is schematized as a Gaussian wavepacket that shows the position of the electron (of charge \( -1 \)) in the unit cell. In a classical description, the CDW appears as an alternation of positive and negative charges \( +1/2, -1/2, +1/2, -1/2, \ldots \) \( \text{(an ionic crystal, similar to rock salt)}, \) whereas the SSH chain \( +1/2 - 1 + 1/2 = 0, +1/2 - 1 + 1/2 = 0, 1/2 - 1 + 1/2 = 0, \ldots \) is similar to a chain of neutral dimers (a molecular crystal made of non-polar dimers).
Figure 12. Bulk electric polarization for the Rice–Mele model as a function of $\theta_{\text{RM}}$ (see figure 3 in [88]). In this reference, the polarization quantum was wrongly assumed to be $P_q = 1$ instead of $P_q = 1/2$ for a spinless RM chain). The polarization quantum $P_q = 1/2$ and the polarization is therefore plotted between $-1/4$ and $+1/4$. The SSH phase corresponds to $\theta_{\text{RM}} = \pi/2$ and $3\pi/2$ with $P_{\text{tot}} = 0$ and the CDW phase to $\theta_{\text{RM}} = 0$ and $\pi$ with $P_{\text{tot}} = 1/4$. Other values of $\theta_{\text{RM}}$ correspond to the RM model when it breaks inversion symmetry.

sign of $\Delta$ but also on whether one has chosen the inversion center to be on an $A$ or a $B$ site. If one fixes the same position origin to compute the Zak phase in both the SSH and the CDW phases, then it is 0 or $\pi$ in the SSH phase and $\pm \pi/2$ in the CDW phase. Figure 12 presents the electric polarization for the complete RM model. When $\theta_{\text{RM}}$ is not 0, $\pi/2$, $\pi$ or $3\pi/2$, this model breaks inversion symmetry and the polarization is not constrained to 0 or 1/4.

In a one-dimensional crystal with inversion symmetry, there are two special points in each unit cell that are known as Wyckoff positions. In our case, they are either on-site ($A$ or $B$) or at mid-distance between two sites (to the right of $A$ or to its left). We find that the Wannier center can only occupy one of these two positions and as a consequence the electric polarization is quantized: either $P_q/2$ or 0 modulo $P_q$. This is the fundamental reason for having two classes of 1D insulators protected by inversion symmetry: CDW has $P_{\text{tot}} = P_q/2$ and SSH has $P_{\text{tot}} = 0$ modulo $P_q$.

5.5.3. Which phase is trivial and which is topological?
We give two different answers before concluding.

5.5.3.1. First answer
The vacuum is a band insulator with inversion symmetry and vanishing polarization. We take it as a definition of a trivial insulator. On the one hand, the SSH chain has a vanishing bulk polarization and is smoothly connected to a molecular insulator (this is obvious in the limit $|\delta| = 2$ in which it describes uncoupled non-polar dimers). It is therefore a trivial insulator. On the other hand, the CDW chain has a non-vanishing bulk polarization $P_{\text{tot}} = P_q/2[P_q]$ and is smoothly connected to an ionic insulator (a one-dimensional version of rock salt $\text{Na}^+ \text{Cl}^-$ obtained in the limit $t = (v + w)/2 \rightarrow 0$). It is therefore a SPT insulator. The $\mathbb{Z}_2$ invariant $\nu$ can be taken to be:

$$\nu = -\frac{P_{\text{tot}}}{P_q/2} = 4(\langle x_- \rangle - \bar{x}) = \frac{Z_- - 2\pi\bar{x}}{\pi/2} [2].$$

(122)

This characterization by a $\mathbb{Z}_2$ invariant is consistent with that based on a topological $\theta$-term in the effective gauge theory [128, 139], that we here recall. The long-wavelength electromagnetic response of a dielectric material in one spatial dimension is described by the following Lagrangian density:

$$\mathcal{L} = -\frac{\varepsilon}{4} F^{\mu\nu} F_{\mu\nu} - \frac{\varepsilon}{4\pi} e^{\mu\nu} F_{\mu\nu} = \varepsilon \frac{E_x^2}{2} + P_{\text{tot}}E_x,$$

(123)

where $E_x$ is the electric field and $F^{\mu\nu}$ is the field tensor with $\mu = 0, i; x$. There are two material parameters that distinguish the dielectric material (here a 1D band insulator) from the vacuum: the familiar dielectric constant $\varepsilon$—related to the electric susceptibility $\chi$ by $\varepsilon = 1 + \chi$—and the less familiar angle $\theta$. The latter is
related to the total (spontaneous) electric polarization by \( P_{\text{tot}} = -P_q \theta/(2\pi) \) modulo \( P_q \). The \( \theta \) angle has only two possible values \( \theta = 0 \) (trivial) or \( \pi \) (non-trivial) when the band insulator respects both time-reversal and inversion symmetries, which is the case here. It appears that \( P_{\text{tot}}(\text{CDW}) = P_q/2 \neq 0 \) \( (\text{topological}, \theta = \pi) \) and \( P_{\text{tot}}(\text{SSH}) = 0 \) \( (\text{trivial}, \theta = 0) \). The \( \mathbb{Z}_2 \) invariant is:

\[
\nu = \frac{\theta}{\pi} [2].
\]

5.5.3.2. Second answer
We now take a different definition of a trivial insulator. It is an atomic-like insulator: it should have exponentially-localized Wannier orbitals and the Wannier centers should be on the ions. In this case, we would call CDW a trivial insulator because the Wannier centers are on the ions, whereas SSH would be a topological insulator because the Wannier centers are exactly in between two nearest-neighbor ions (see figure 11).

Although we do not prove it here, in both phases (SSH and CDW), there exists exponentially-localized Wannier functions (as is always the case in 1D) that respect inversion symmetry. The existence of an exponentially-localized Wannier function respecting some symmetry is often taken as a definition of an atomic limit [140]. In the present case, it would therefore seem that there are two inequivalent atomic limits. A naive atomic limit with the Wannier centers on the ions (the CDW case) and an obstructed atomic limit [140] with the Wannier centers in between the ions (SSH case). Such an obstructed atomic limit occurs because of bonding and the fact that the natural objects in this insulator are molecules (here non-polar dimers) rather than atoms.

5.5.3.3. Conclusion
In the end, there are two different classes of inversion-symmetric band insulators in one dimension that are distinguished by their electric polarization. We disagree with the commonly-found statement that the two phases of the SSH model with positive or negative \( \delta \) would correspond to the two different values of the polarization (see e.g. [120]). We think that these two values of polarization actually correspond to SSH and to CDW. The difficulty in this interpretation is in the fact that what seems trivial from the polarization perspective seems non-trivial from the Wannier center perspective and vice-versa. We believe that the first answer is better motivated as it corresponds to a physically measurable quantity, the electric polarization. We therefore call SSH trivial and CDW topological, but this may be a matter of convention.

The above peculiarities of the RM model is related to its having two sites per unit cell. Another model of 1D inversion-symmetric two-band insulator but with a single site per unit cell—the coupled \( s \) and \( p \) bands—is studied in [21] and was originally introduced by Shockley [141]. It has a simpler behavior than SSH and CDW (e.g. there is no difference between the periodic and the canonical Bloch Hamiltonian). In this model, the non-trivial (covalent \( sp \) hybridized) insulator corresponds to \( P_{\text{tot}} = P_q/2 \) \( [P_q] \) and having Wannier centers in between ions, whereas the trivial (atomic \( s \) band) insulator has a vanishing polarization and Wannier centers on the ions [21].

5.5.4. Topological phase transition
Imagine tuning a transition between the SSH and the CDW phases. We start with \( \delta > 0 \) and \( \Delta = 0 \) (i.e. \( \lambda < 0 \)), then diminish the dimerization until we reach \( \delta = 0 \) maintaining \( \Delta = 0 \) (i.e. \( \lambda = 0 \)) and then we increase \( \Delta \) while maintaining \( \delta = 0 \) (i.e. \( \lambda > 0 \)). In such a case, the inversion symmetry is preserved all the way. Then, we necessarily have to close the gap at \( \lambda = 0 \), a point at which the electron is delocalized over the whole crystal and the Wannier center is ill defined. Actually, in the process the Wannier center jumps from on-site to mid-bond exactly when \( \lambda = 0 \). See the phase diagram in figure 13.

5.5.5. Conclusion
In summary, the RM model with inversion symmetry connects the SSH model to a CDW model. It allows one to study the phase transition between the two phases of a 1D band insulator with inversion symmetry. The non-trivial phase is a SPT insulator—rather than an intrinsic topological insulator. As the symmetry in question is the space inversion (and not an external symmetry such as time-reversal or particle-hole), it is known as a topological crystalline insulator [136]. It is characterized by a \( \mathbb{Z}_2 \) topological invariant given in (122) and built from the bulk electric polarization. The CDW chain is an ionic insulator (inversion-protected topological insulator), whereas the SSH chain is a molecular insulator (trivial).

If space inversion is broken (which happens in the RM model whenever \( \theta_{\text{RM}} \neq 0 \) modulo \( \pi/2 \)), the system has a single gapped phase and is a trivial band insulator. Indeed, one can continuously go from the CDW to the SSH phase without closing the gap by breaking inversion symmetry. Its electric polarization is unquantized (neither 0 nor \( P_q/2 \), see figure 12) and the Wannier center is not at a special position.
6. Dirac fermions on 2D honeycomb lattice

Graphene is the archetype of a strictly 2D crystal whose electronic excitations obey a Dirac-(or Weyl-)like equation for multicomponent wave functions. The spinor components correspond to the amplitudes of the Bloch waves on the two inequivalent triangular sublattices of the honeycomb structure [142, 143], instead of the real spin projections involved in the historical Dirac equation of high-energy physics [46, 144]. Since graphene was isolated in 2004 [36, 37, 145], the family of 2D crystalline solids has gained many new members exhibiting striking electrical and optical properties, e.g. TMDC [146], twisted graphene bilayers, van der Waals heterostructures. Single layer graphene is still remarkable by the simplicity of its band structure around the Fermi level [38]. The relevant excitations near the Fermi level form two \( \pi \)-bands that are very well isolated from the other energy bands because carbon is a light element. In comparison, the band structure of MoS\(_2\), and more generally of any TMDC, involves much more complicated orbital combinations.

Graphene has been a fertile playground for the development of topological concepts from the Haldane model [16] to the Kane–Mele model [28, 29] to cite only two major milestones. The Haldane model is the first representative of the Chern insulator class which are the time-reversal breaking topological insulators, namely band insulators exhibiting the QHE without the Landau level structure. Like the QHE, Chern insulators are also characterized by chiral edge states. The Kane–Mele model describes graphene with a finite intrinsic spin–orbit (SO) coupling, which turns out to be a time-reversal invariant topological insulator associated to a robust \( \mathbb{Z}_2 \) index, and helical edge states. In graphene, the SO coupling is low and therefore the real spin of electrons in graphene is almost uncoupled to the motion. Nevertheless in TMDCs, the SO coupling is stronger and may lead to quantum spin hall (QSH) topological insulators. The compound WTe\(_2\) has been predicted [147] and experimentally demonstrated [148–150] to be a QSH insulator. We will not cover the time-reversal invariant TIs in this review, see [35, 55, 56].

In this section, we use graphene to exemplify the concepts of geometrical and topological band theory, introduced in section 4. We start with pristine graphene (without any SO coupling) which has a gapless band structure consisting of two bands that cross each other at two isolated points of the BZ. Pristine graphene being both centrosymmetric and time-reversal invariant, its Berry curvature is zero everywhere in the BZ except at the Dirac band touching points, which are local singularities and where it is ill-defined. In the following, we present the spectral and Berry curvature properties of the Semenov insulator obtained by gapping graphene via a staggered on-site potential [151]. This model describes hexagonal boron nitride (h-BN) and has a well-defined finite Berry curvature, but its total flux through the BZ is zero. It is therefore the typical example of a material whose band structure has non trivial local geometrical properties (we call it a Dirac insulator), but still a trivial global topology. Then we introduce the Haldane model where a periodic pattern of magnetic fluxes breaks time-reversal symmetry without breaking the translational invariance of the Bravais lattice. The resulting Haldane insulator has bands that each carries a finite Chern number, thereby leading to QHE in the absence of any net magnetic flux through the sample. We conclude by discussing the band inversion mechanism and topological transitions, using a simple method to evaluate Chern numbers for any two-band model.
The Haldane insulator has been experimentally observed in Bi$_2$Se$_3$ or Bi$_2$Te$_3$ films doped with magnetic impurities [152, 153], and more recently in the intrinsic magnetic insulator MnBi$_2$Te$_4$ [154]. It was also realized with cold atoms trapped in optical lattices using dynamical methods to induce complex hopping amplitudes [155]. Pristine graphene and the Semenov model were also implemented using cold atom vapors [156].

6.1. Pristine graphene and massless Dirac fermion

We introduce here the tight-binding model of graphene [38] and its symmetries. We review its full band structure and the emergence of Dirac fermions. We put an emphasis on the properties of the spinor wave functions.

6.1.1. Hamiltonian and band structure in the whole BZ

Graphene is the one-atom thick layer of carbon atoms arranged with the honeycomb lattice structure, made of two interpenetrating triangular sublattices, respectively denoted $A$ and $B$ (figure 14). Each carbon atom has six electrons: five core electrons filling the inner shells (2 electrons in the 1s orbital and 3 electrons in the covalent $sp^2$ bonds) while a single valence electron fills the $p_z$ orbital perpendicular to the plane. As in polyacetylene, the $p_z$ orbitals lead to $\pi$ bands that are well ‘isolated’ from filled lower bands and empty higher bands. Much of the physics of graphene is related to those two-dimensional $\pi$ bands that are accurately described by the following tight-binding Hamiltonian [38]:

$$H_0 = t \sum_{r_a} \sum_{\alpha=1}^{3} c_{\alpha}^\dagger (r_a + \delta_\alpha) c_\alpha (r_a) + \text{h.c.}, \quad (125)$$

where $t \simeq -2.7$ eV is the hopping amplitude between the $p_z$ orbitals of two adjacent carbon atoms. The operator $c_\alpha (r)$ destroys an electron in the $p_z$ orbital at site $r$, with $\alpha = A, B$ indicating the sublattice. The sum over $r_a$ runs over the $A$-sites which form a triangular Bravais lattice generated by the basis vectors (figure 14):

$$a_1 = \sqrt{3} a e_x, \quad a_2 = \frac{a}{2} \left( \sqrt{3} e_x + 3 e_y \right), \quad (126)$$

where $a = 0.142$ nm is the length of the carbon–carbon bond. The vectors $\delta_\alpha$ defined by:

$$\delta_{1,2} = \frac{a}{2} \left( \pm \sqrt{3} e_x + e_y \right), \quad \delta_3 = -a e_y, \quad (127)$$

connect any $A$-site to its three $B$-type nearest neighbors (figure 14). The hopping matrix elements between next-nearest neighbors (NNNs) can be safely neglected, being roughly ten times smaller than the main hopping $t$ [142, 143]. Sites from distinct sublattices, $A$ and $B$, are crystallographically different by the orientations of the three attached bonds. In graphene, all sites are occupied by identical carbon atoms, while in hexagonal boron nitride ($h$-BN) the boron atoms are distributed on one sublattice while the nitrogen atoms lie on the other one, see section 6.2.

The graphene lattice can be seen as a 2D version of the polyacetylene chain obtained by replacing the C–H bonds by C–C bonds to another chain. Although the tight-binding models of graphene and polyacetylene look quite similar, there are also important differences. First, in graphene each $p_z$ orbital is coupled to 3 neighboring $p_z$ orbitals, instead of two neighbors for polyacetylene. Second, there is no Peierls instability in graphene, although dimerization effects have been studied: for a magnetic-field induced Peierls polyacetylene look quite similar, there are also important differences. First, in graphene each C–H bonds by C–C bonds to another chain. Although the tight-binding models of graphene and

$$c_\alpha (r) = \frac{1}{\sqrt{N}} \sum_k e^{i k \cdot r} c_\alpha (k), \quad (128)$$

where $l = A, B$ is the sublattice index and $N$ is the total number of unit cells. Note that the exponential phase factors contain the exact location of the atomic sites. After substitution of equation (128), the Hamiltonian equation (125) becomes diagonal in momentum and reads:

$$H_0 = \sum_k c_\dagger (k) [H_0 (k)]_{\alpha\beta} c_\alpha (k), \quad (129)$$
where $k$ is restricted to the first BZ. The (canonical) Bloch Hamiltonian $H_0(k)$, which acts on the sublattice isospin, is given by:

$$H_0(k) = d_x(k)\sigma_x + d_y(k)\sigma_y = |d(k)| (\cos \varphi_k \sigma_x + \sin \varphi_k \sigma_y),$$  \hspace{1cm} (130)

since only off-diagonal hopping amplitudes are included in the model defined by equation (125). The phase $\varphi_k$ is the azimuthal angle along the equator of the Bloch sphere. The real functions $d_x(k)$ and $d_y(k)$ are defined by:

$$d_x(k) = t \sum_{\alpha=1}^3 \cos(k \cdot \delta_\alpha), \quad \text{and} \quad d_y(k) = t \sum_{\alpha=1}^3 \sin(k \cdot \delta_\alpha),$$  \hspace{1cm} (131)

over the whole BZ. The functions $d_x(k)$ and $d_y(k)$ are respectively even and odd in momentum reversal $k \rightarrow -k$, which is related to time-reversal and inversion symmetries of pristine graphene. The electronic energy spectrum is given by the length of the vector $d = (d_x, d_y)$:

$$E(k) = \pm |d(k)| = \pm \sqrt{d_x^2(k) + d_y^2(k)},$$  \hspace{1cm} (132)

which describes a valence band (minus sign) and a conduction band (plus sign) that are symmetric with respect to $E = 0$. The zero energy corresponds to the common energy of the $p_z$ atomic orbitals on sublattices $A$ and $B$. The corresponding wave functions are the spinors:

$$|u_+(k)\rangle = \frac{1}{\sqrt{2}} \left| \begin{array}{c} 1 \\ e^{i\varphi_k} \end{array} \right|, \quad \text{and} \quad |u_-(k)\rangle = \frac{1}{\sqrt{2}} \left| \begin{array}{c} 1 \\ -e^{i\varphi_k} \end{array} \right|, $$   \hspace{1cm} (133)

in the upper and lower bands respectively. These are the same as for the SSH model albeit the angle $\varphi_k$ is now a function of a 2D quasimomentum given by $\varphi_k = \arg(k_x + ik_y)$. The Bloch Hamiltonian does not have the periodicity of the reciprocal lattice. The reason is that its definition involves the distance between sites within the unit cell. For the honeycomb lattice, the distance between nearest-neighbor $A$ and $B$ sites is actually one-third of a lattice vector. For example, $\delta_3 = (a_1 - 2a_2)/3$. This translates into the fact that the Bloch Hamiltonian has a triple periodicity, as can be seen by plotting the phase $\varphi_k$ (see figure 15(left)) [83]. This fact has measurable consequences which have been observed by quantum tomography using cold atoms in an optical lattice, see [159]. If instead of using the canonical Bloch Hamiltonian (130), one uses a periodic Bloch Hamiltonian:

$$\mathcal{H}_0(k) = |d(k)| (\cos \varphi_k \sigma_x + \sin \varphi_k \sigma_y),$$ \hspace{1cm} (134)

then the azimuthal angle is different and called $\phi_k$. The latter has the periodicity of the hexagonal BZ and is related to $\varphi_k$ by $\phi_k = \varphi_k - k \cdot \delta_3$ (see figure 15(right)), where $-\delta_3$ is the position of a $B$ site with respect to an $A$ site within the unit cell (see figure 14(left)).
The valence and conduction bands touch at isolated points of the BZ obtained by solving the equations $d_x(k) = d_y(k) = 0$. There are only two inequivalent solutions located at:

$$k = \pm \mathbf{K} = \pm \frac{4\pi}{3\sqrt{3}a} \mathbf{e}_x,$$

(135)

and called the Dirac points. Any other solutions of the equation $d(k) = 0$ can be linked by a reciprocal lattice vector to one of these two solutions, and therefore would describe the same physical state. In the SSH chain, the band touching conditions were also obtained by cancelling two functions $d_x$ and $d_y$, but there was only a single momentum component $k$, so it needed a fine tuning of hopping parameters to close the gap, namely $\delta = v - w = 0$. In graphene, the momentum $\mathbf{k}$ runs over a 2D manifold for graphene, so there is no need for fine tuning of the Hamiltonian parameters and one generically gets Dirac points. Even for anisotropic graphene, where the hoppings $t_{\alpha}$ along the bonding directions $\delta_{\alpha}$ might differ, there will be always some isolated points $\mathbf{k} = (k_x, k_y)$ satisfying the 2 equations: $d_x(k) = d_y(k) = 0$ [160].

6.1.2. Effective Hamiltonian near the Dirac points and symmetries

We consider now the low-energy theory for the single-particle states near the Dirac points [39], see equation (135). The momenta are written as $\mathbf{k} = \pm \mathbf{K} + \mathbf{q}$ where $\mathbf{q} = q_x \mathbf{e}_x + q_y \mathbf{e}_y$ is a small momentum deviation from the Dirac points, namely $|\mathbf{q}|a \ll 1$. The annihilation operators for these states are relabeled as $c_\alpha = c_\alpha(\mathbf{k} + \mathbf{q}) = c_\alpha(\pm \mathbf{K} + \mathbf{q})$ to indicate the valley they belong to. Expanding to first order in momenta, the Bloch Hamiltonian describing the low-energy excitations in the valley near $\xi \mathbf{K} (\xi = \pm 1)$ reads:

$$H_0(\xi \mathbf{K} + \mathbf{q}) = H_0^{(\xi)}(\mathbf{q}) = v_F \begin{pmatrix} 0 & \xi q_x - i q_y \\ \xi q_x + i q_y & 0 \end{pmatrix} = v_F(q_x \xi \sigma_x + q_y \sigma_y),$$

(136)

where $v_F = -3at/2 \approx 10^6$ m s$^{-1}$ is Fermi velocity. The Fermi velocity is roughly the bandwidth $t$ divided by the BZ size $1/a$ (when $h = 1$). The Hamiltonian can also be written $H_0^{(\xi)}(\mathbf{q}) = v_F q_\alpha \sigma_\alpha$ with $i, j = x, y = 1, 2$ here. Therefore, near each Dirac point, one obtains a 2D Weyl Hamiltonian describing massless fermions carrying a sublattice isospin coupled to their momentum. The two species of massless Dirac fermions are attached to a given valley, labelled by $\xi$. The chirality (or winding) of each Dirac point is given by $\chi_\xi = \text{sign } \text{det}(v_F) = \xi$. The low-energy dispersion of those fermions is valley independent and reads:

$$E_{\pm}(\mathbf{q}) = \pm v_F |\mathbf{q}|,$$

(137)

which is typical of a relativistic massless particle with velocity of light replaced by $v_F$. This linearized dispersion is reminiscent of the 1D fermions obtained in the SSH chain near $k = \pi$ at the gap closure point $m = v - w = 0$. It also has a 3D counterpart in Weyl and Dirac semimetals [161]. As a general reference on the distinction between Dirac (i.e. complex and having both chiralities), Weyl (i.e. complex and chiral) and Majorana (i.e. real and achiral) fermions, we recommend [162].

Nevertheless we would like to emphasize the differences between the Dirac (and Weyl) equation in the contexts of graphene and particle physics, respectively. In high-energy physics, the Dirac equation comes...
from Lorentz-invariance and very general considerations related to special relativity and quantum mechanics. Then, in 3+1 space-time dimensions, the minimal objects to satisfy Dirac equation are bispinors combining the spin and particle/antiparticle degrees of freedom. In graphene, the origin of the Dirac physics is totally different. As we have seen, the spinors originate from a $a, k, p$ expansion around special points of a particular band structure. Hence in graphene, there is no fundamental issue with the negative energy states that are just the valence band states (these states are in fact bounded from below by the bottom of the valence band).

Finally the emergent Lorentz invariance of equation (137) is only valid near the Dirac point, namely for wave vectors $q$ located in a disk whose radius is far smaller than the inverse lattice spacing $1/a$, whereas Lorentz invariance applies in the whole Minkowski space-time. Finally the 4 components of the spinors are associated to the sublattice isospin (instead of real spin), and to the valley index (instead of particle/antiparticle label).

### 6.1.3. Protection of the Dirac points by symmetry

Graphene is invariant under space inversions with respect to particular points of the lattice, which are the centers of the hexagons of carbon atoms and the centers of the carbon–carbon bonds. In the absence of any magnetic field or impurities, graphene is also time-reversal invariant and therefore all hopping parameters are real numbers.

It is rather difficult to gap out the Dirac points between two bands of spinless fermions. Indeed, there are only four possible types of perturbations which mathematically corresponds to the identity and the three Pauli matrices. A scalar perturbation (proportional to identity in sublattice isospin space) will just shift both bands in energy without separating them. Moreover perturbations in $\sigma_x$ or $\sigma_y$ would shift the position of the band touching to another location in the BZ without removing the degeneracy. Only a perturbation acting as $\sigma_z$ could gap out the Dirac points, but we are now going to explain that such a term is forbidden by the time-reversal and inversion symmetry of graphene, and more generally in any two-band model which is invariant under those two symmetries. Let us consider a generic two-band model in 2D, described by the Bloch Hamiltonian $H(k) = d(k) \cdot \sigma$, and with sublattice isospin ($T^z = +1$). The time-reversal $T$ symmetry condition reads:

$$H(k) = H^*(-k) \implies d_x(k) = d_x(-k), \quad d_y(k) = -d_y(-k) \quad \text{and} \quad d_z(k) = d_z(-k),$$

which is indeed satisfied by graphene and $h$-BN.

The inversion $I$ symmetry condition can be written:

$$\sigma_x H(k) \sigma_x = H(-k) \implies d_x(k) = d_x(-k), \quad d_y(k) = -d_y(-k) \quad \text{and} \quad d_z(k) = -d_z(-k),$$

which is satisfied by graphene but violated by $h$-BN.

Finally the combined $IT$ symmetry leads to:

$$\sigma_z H^*(k) \sigma_z = H(k) \implies d_x(k) = d_x(k), \quad d_y(k) = d_y(k) \quad \text{and} \quad d_z(k) = -d_z(k).$$

Therefore, the presence of both $T$ and $I$ enforces the function $d_x(k)$ to be odd and even in $k$, which means $d_x(k) = 0$ for all $k$. The third line shows that the symmetry $IT$ is enough to protect the existence of a robust contact, even if $T$ and $I$ were separately broken. Then a gap closing requires only the simultaneous cancellation of two functions $d_x(k) = d_y(k) = 0$ at some $k$ which can be varied in a 2D manifold. This is why isolated solutions are expected as it happens in graphene. The existence of such isolated solutions of $d(k) = 0$, preventing the system to become gapped, is robust even if some crystal symmetries are lost and more hopping amplitudes are added. For instance additional second-neighbor hopping will break the electron-hole symmetry discussed above, but will not affect the existence of Dirac points. Other perturbations, like an anisotropic deformations on one type of bond, only shift the Dirac points and modify the conical dispersion around them [160, 163, 164]. The touching points are protected by more fundamental symmetries, namely space inversion and time-reversal symmetries. Breaking at least one of these symmetries usually results in a gap opening at the Dirac points, as we will see below with noncentrosymmetric $h$-BN, and in the time-reversal breaking Haldane model.

### 6.2. Hexagonal boron nitride and massive Dirac fermions

Like graphene, hexagonal boron nitride ($h$-BN) crystallizes in the honeycomb structure, but two different elements (boron and nitrogen) occupy the $A$ and $B$ inequivalent sublattices respectively. The corresponding tight-binding model is given by $H_0$ in equation (125) plus a staggered potential $H_1$ that takes into account the difference in energy between the boron and nitrogen orbitals. The on-site staggered potential simply reads:

$$H_1 = M \sum_{\alpha} \left( c_A^\dagger(r_\alpha) c_A(r_\alpha) - c_B^\dagger(r_\alpha + \delta_\alpha) c_B(r_\alpha + \delta_\alpha) \right),$$

(141)
which is very similar to the asymmetry term in the RM model, albeit defined here on the honeycomb lattice. This perturbation $H_1$ breaks the inversion symmetry $I$ between $A$ and $B$ sublattices while still keeping time-reversal invariance. The effect of this staggered potential, first studied by Semenov [151], is to gap out the Dirac points, thereby turning graphene into an insulator, at half filling. If we take the limit $M \to \infty$, thereby making the nearest neighbor hopping amplitude negligible, the electrons are forced to sit on one of the triangular sublattice (the $B$ one) and one obtains an ionic insulator.

The corresponding Bloch Hamiltonian can be written:

$$H(k) = d_x(k) \sigma_x + d_y(k) \sigma_y + d_z(k) \sigma_z \quad \text{with} \quad d_z(k) = M, \quad (142)$$

the functions $d_x(k)$ and $d_y(k)$ being still given by equation (131). Hence at $k = \pm K$, the Bloch Hamiltonian reduces to $H(\pm K) = M \sigma_z$, and the Dirac points are gapped because $H(\pm q) = v_F(q) \xi_\sigma x + q_\sigma y + M \sigma_z$. The spectrum is:

$$E_{\pm}(k) = \pm |d(k)| \approx \pm \sqrt{v_F^2 q^2 + M^2}, \quad (143)$$

which corresponds to a massive Dirac fermion. The spinor eigenfunctions of equation (142), expressed in the ‘north gauge’, read:

$$|\psi_+ (k)\rangle = \left( \cos \frac{\theta_k}{2} \right), \quad \text{and} \quad |\psi_- (k)\rangle = \left( \sin \frac{\theta_k}{2} e^{-i \varphi_k} \right), \quad (144)$$

in the upper and lower bands respectively, and depends on the parameter $M$ via the angle $\theta_k$. The $k$-dependent quantities $\varphi_k$ and $\theta_k$ are the spherical coordinate angles of the unit vector:

$$\hat{d}(k) = \frac{d(k)}{|d(k)|} = \left( \frac{\cos \varphi_k \sin \theta_k}{\cos \theta_k} \right). \quad (145)$$

At each quasi-momentum $k$, these spinors are exactly the same as those of a spin one-half in a fictitious magnetic field $\hat{d}(k)$, given by equation (21). The mapping $k \to d(k) = d(k) / |d(k)|$ is essential and captures the global topological properties of the Hamiltonian $H(k) = d(k) \sigma$, as we will see in section 6.4. The image manifold of the 2-dimensional BZ $T^2$ is a subset of the unit sphere $S^2$, that is included in the north-hemisphere (resp. south hemisphere) for $M > 0$ (resp. $M < 0$).

Locally in $k$-space, h-BN has interesting and non trivial geometrical properties. This can be seen by computing the Berry curvature of boron nitride [85]:

$$F_{xy}^\pm(k) = \pm \sqrt{3} \frac{v_F M}{E(k)} \sin \left( k, \frac{\delta_2 - \delta_3}{2} \right) \sin \left( k, \frac{\delta_3 - \delta_1}{2} \right) \sin \left( k, \frac{\delta_1 - \delta_2}{2} \right), \quad (146)$$

a formula which is valid within the whole BZ (here $\pm$ refers to the band index). Near a Dirac point (in a given valley $\xi = \pm 1$), this formula simplifies as:

$$F_{xy}^\pm(\xi K + q) = \pm \frac{\xi v_F^2 M}{2(M^2 + v_F^2 q^2)^{3/2}}, \quad (147)$$

thereby recovering rotational invariance around the Dirac points. In the limit $M \to 0$, we recover that the Berry curvature is vanishing everywhere except at $q = 0$. In the limit first $q = 0$, and then $M \to 0$, the sign (even for a given valley) is not well defined. The sign depends on the valley (time-reversal symmetry). So the total flux is zero. The Berry curvature is plotted in figure 16(left). If, instead of using the (canonical) Bloch Hamiltonian, one wrongly uses a periodic Bloch Hamiltonian to compute the Berry curvature with the usual formula, it would give the result plotted figure 16(right). This obviously lacks the correct rotational symmetry [83, 85, 86].

Locally in reciprocal space, near each valley, boron nitride behaves as an interesting Dirac insulator. It therefore features the physics related to the Jackiw–Rebbi mechanism (see section 5.4). For example, if there is a domain wall (a grain boundary) between two different staggered on-site potentials $M > 0$ on one side and $M < 0$ on the other side, there should be a gapless chiral one-dimensional edge mode running along the boundary between the two domains. The Jackiw–Rebbi mechanism occurs in each valley. Therefore the corresponding gapless edge modes are actually valley-filtered: one valley has a given chirality and the other has the opposite chirality. These were studied in [137]. Such valley-filtered gapless edge modes are not very robust and may be gapped by any inter-valley scattering process. For example, they depend on whether the
boundary is of zig-zag or armchair type \[137\]. If inter-valley scattering is negligible then these modes give rise to a quantized valley Hall effect \[165\].

Even if the Berry curvature is finite and exhibits pronounced peaks near low gaps in the spectrum, its total flux must vanish in \(h\)-BN due to the time-reversal invariance of the model. This is because time-reversal invariance implies that the Berry curvature is an odd function of quasimomentum. Here, the Berry curvature contributions from the two valleys compensate each other. This cancellation of the Chern number also confirms the absence of quantized Hall effect in this model and boron nitride is considered as a trivial insulator. In the next section, we explain how Haldane devised a band insulator model on the honeycomb lattice in which the total flux of the Berry curvature does not vanish, and is even quantized when the Fermi level lies in the band gap. The expressions above suggest that the Berry flux cancellations between valleys is related to the fact that the sign of \(M\) is identical in both valleys. Therefore having a finite Chern number requires to have opposite signs of those masses through inverting the bands in a single valley (and not in the other). That’s indeed what the Haldane model does.

### 6.3. Haldane model of graphene and topology

In 1982, TKNN had considered the QHE in the presence of a periodic lattice \[12\], where the interplay of the magnetic field and lattice generates the famous Hofstadter spectrum \[13\]. In his seminal work \[16\], Haldane made a step forward by realizing that the mandatory condition for QHE was time-reversal symmetry breaking, while the homogeneous magnetic field is not necessary. To prove it, he devised a toy model based on a 2D single sheet of graphite (although at that time graphene was far from being an experimental reality) showing QHE without Landau levels. The recipe consists in breaking time-reversal symmetry while preserving a zero net magnetic flux per unit cell and therefore Bloch states. It is also important to have an insulator to get quantized Hall conductivity. Haldane realized that a specific pattern of complex second neighbor hoppings can bring all these ingredients together \[16\]. The Hamiltonian of the Haldane model reads:

\[
H = H_0 + H_1 + H_2,
\]

where the Hamiltonian of pristine graphene \(H_0\), equations (130) and (131), and the on-site staggered potential \(H_1\), equation (141), have already been encountered to describe \(h\)-BN \[151\]. In order to break time-reversal symmetry, Haldane introduced the term \(H_2\) which contains the complex NNN hoppings, equal either to \(t_2 e^{i\phi}\) or \(t_2 e^{-i\phi}\) depending on the NNN bond \[16\]. Those two values are not distributed randomly over the bonds but are organized in a way related to the orbital motion of electrons, see figure 17(a). The value \(t_2 e^{i\phi}\) is associated to any term \(\hat{c}_i^\dagger \hat{c}_j\) where \(i\) and \(j\) are second neighbors that have a common intermediate first neighbor \(l\) on the left of the vector joining site \(j\) to site \(i\), see figure 17(b). The Haldane term can be written in a compact form as:

\[
H_2 = t_2 \sum_{(i \leftarrow j)} \hat{c}_i^\dagger \hat{c}_j e^{i\nu_{ij}/\phi},
\]

Figure 16. (Left) Berry curvature for boron nitride as a function of \((k_x, k_y)\) (the hexagonal BZ is indicated by a black line). (Right) Same but computed from a periodic Bloch Hamiltonian instead of the canonical one. The latter result is therefore not correct. See [83, 85, 86].
where the sum is over all oriented second-neighbor pairs \( ((i \leftarrow j)) \). Each non-oriented pair \( ((i, j)) \) contributes to two conjugated terms \( ((i \leftarrow j)) \) and \( ((j \leftarrow i)) \), thereby making \( H_2 \) Hermitian. The index \( \nu_{ij} \) of the oriented pair \( ((i \leftarrow j)) \) is defined by:

\[
\nu_{ij} = e_z \cdot (\delta_{ii} \times \delta_{jj}) / || \delta_{ii} \times \delta_{jj} ||, \tag{150}
\]

where \( l \) is the common first-neighbor shared by the sites \( i \) and \( j \), see figure 17(b).

Going back to our notations with explicit mention to the \( A \) and \( B \) sublattices, this leads to:

\[
H_2 = t_2 \sum_{i=1}^{3} \left( \sum_{r_i} c^*_i(r_i) c_A(r_i + b_i) e^{i\phi} + \sum_{r_n} c^*_n(r_n) c_B(r_n + b_i) e^{-i\phi} \right) + \text{H.c.}, \tag{151}
\]

where \( b_1 = \delta_2 - \delta_3, b_2 = \delta_3 - \delta_1, \) and \( b_3 = \delta_1 - \delta_2, \) are the vectors connecting NNN sites (see figure 14). The Haldane term \( H_2 \) breaks time-reversal symmetry \( T \) because the hoppings \( t_2 e^{i\phi} \) are complex when \( \phi \neq 0 \) or \( \pi \), but respects inversion symmetry. Both the Semenov term \( H_1 \) and the Haldane term \( H_2 \) break the chiral symmetry because they do not change sign under the transformation: \( c_A \rightarrow c_A \) and \( c_B \rightarrow -c_B \). For the topological symmetry classes [41, 42], only \( T \) and \( S \) matter, so the Haldane model belongs to the \( A \)-class of the ten-fold periodic table in two dimension, just like the QHE.

Going to reciprocal space, the Bloch Hamiltonian for the Haldane model reads:

\[
H(k) = \epsilon_0(k) \sigma_0 + d_x(k) \sigma_x + d_y(k) \sigma_y + d_z(k) \sigma_z, \tag{152}
\]

and the corresponding band structure is given by:

\[
E_{\pm}(k) = \epsilon_0(k) \pm |d(k)|, \tag{153}
\]

where the functions \( d_x(k) \) and \( d_y(k) \) are still given by the same equations (131) as for pristine graphene. The specificity of the Haldane model lies in the following \( k \)-dependent terms:

\[
\epsilon_0(k) = 2t_2 \cos(\phi) \sum_{i=1}^{3} \cos(k.b_i) \quad \text{and} \quad d_z(k) = M + 2t_2 \sin(\phi) \sum_{i=1}^{3} \sin(k.b_i). \tag{154}
\]
18

Figure 18. (Left) Phase diagram of the Haldane model in the \((\phi, M/t_2)\) plane. The blue lines are the transition lines between the topological phase (colored in orange) and the trivial phase (white). (Right) A representation of the manifold \(M\) —i.e. the image of the whole BZ through the map \(k \rightarrow d(k)\) —in the \((d_\phi, d_\psi)\) plane, defined by \(d_2 = 0\), for various values of \(t_2\) (with \(M\) set to 1) and at \(\phi = \pi/2\). The blue lines do not enclose the origin and corresponds to values of \(M/t_2\) in the trivial phases, the flat horizontal segment being \(t_2 = 0\) (Semenov insulator). The red curve encloses the origin and is therefore in the topological phase. The green dashed line corresponds to \(M = 3\sqrt{3}t_2\).

The NNN perturbation exhibits spatial dispersion (\(k\)-dependence) because it is nonlocal in real space. The part of \(H(k)\) which is proportional to the identity just shifts the energies and breaks the electron–hole symmetry of the purely NN model. Nevertheless if \(M = 0\), the system remains gapless (if \(\phi = 0, \pi\)) under introduction of a real NNN hoppings \(t_2 \neq 0\), because both \(I\) and \(T\) are preserved for real NNN hoppings. In contrast, for complex hoppings (namely \(\phi \neq 0, \pi\)), the term proportional to \(\sigma_z\) opens gaps at the Dirac points. The mass term \(d_2(k)\) is the sum of a constant \(M\) and a contribution proportional to \(t_2\) which is odd in momentum and in particular changes sign in different valleys. Near the Dirac points, \(d_\phi\) and \(d_\psi\) vanish and one simply has to substitute \(k = K\) (or \(k = -K\)) in \(d_2\) as a zero order approximation:

\[
d_2(k = \xi K) = M - 3\sqrt{3}t_2\xi \sin(\phi),
\]

where \(\xi = \pm 1\) is the valley index. The term in \(t_2\) depends on the flux and vanishes for \(\phi = 0\) and \(\phi = \pi\) where the time-reversal symmetry is restored. Therefore there is a competition between \(M\) and \(t_2\sin(\phi)\), and three different scenarios are possible depending on parameters. First, when \(|M|\) is very large, the Semenov mass \(M\) dominates and the masses \(d_2(k = \xi K)\) have the same sign at both Dirac points. This situation can be connected continuously to an atomic insulator without closing any gap, so this corresponds to the trivial insulating phase. Second, when \(|M|\) is small (eventually zero), the \(\xi t_2\) term dominates and induces opposite masses at the Dirac points \(\xi = \pm 1\). This is also an insulating phase and we will see that it is a non-trivial phase in the sense that the valence band carries a non zero Chern number \(C_- = \pm 1\) (orange regions in figure 18). Finally, the transition between the trivial insulator and the non-trivial insulator occurs when one of the mass or the gap closes and changes sign at one of the Dirac point. This situation corresponds to the blue lines in figure 18 that are given by:

\[
\frac{M}{t_2} = 3\sqrt{3}\xi \sin(\phi),
\]

which are transition lines between the two classes of band structures.

This seminal paper by Haldane was not fully nor widely appreciated at that time, because graphene was not experimentally available and even considered as a toy model for theorists. Moreover the time-reversal breaking terms involved in the Haldane model seemed (and are still) very difficult to realize: it requires a pattern of alternating magnetic fluxes at the scale of the unit cell. The works of Kane and Mele around 2005 [28, 29] lead to a revival of interest in the Haldane model physics and attempts to realize it experimentally. This has been done in some condensed matter materials and in cold atom systems. That was the first example of a topological band insulator: a Chern or QAH insulator.

6.4. Chern number of two-band models in 2D

We consider the very important case of a crossing between two non-degenerate bands. This is the most elementary case in topological band theory where it plays a role similar to the TLS in atomic physics. In such
a case, the Berry curvature in one band is only originating by virtual transitions to the only other band. The general Bloch Hamiltonian that captures the two-band system is:

$$H(k) = d(k) \cdot \sigma.$$ (157)

We have seen examples with the models of spinless fermions on the honeycomb lattice described above.

6.4.1. Chern number as a wrapping number

To compute the Chern number, one can use equation (56) and inject:

$$\partial \sigma H(k) = \sigma_a \partial \sigma_a(k) \quad \text{and} \quad (E_n - E_n')^2 = 4d^2,$$ (158)

where summation over repeated index $a = x, y, z$ is implied. The Berry curvature is given by:

$$F_{\alpha\beta}^{\sigma}(k) = \frac{i}{4d^2} \langle u_n | \partial_\alpha H(k) \partial_\beta H(k) | u_n \rangle + \text{c.c.} = \frac{1}{2d^2} \epsilon_{\alpha\beta\gamma} \langle u_n | \sigma_\gamma | u_n \rangle \partial_x \sigma_a(k) \partial_y \sigma_b(k),$$ (159)

where the last line holds for the valence band because then $\langle u_n | \sigma_\gamma | u_n \rangle = -d_c/d$. The Berry curvature is opposite for the conduction band. Then we can express the Chern number of the valence band $n = -$ under the following form:

$$C_- = \frac{1}{2\pi} \int_{T^2} d^2k F_{\alpha\beta}^{\sigma}(k) = \frac{1}{4\pi} \int d^2k \left( \partial_x d / \partial k_x \times \partial_y d / \partial k_y \right) \cdot \vec{d} = \frac{1}{4\pi} \int d\Omega \in \mathbb{Z},$$ (161)

which is the Chern number of the fiber bundle. This formula is useful because it expresses the Chern number of a band as an integral over the parameter vector components with no explicit mention of spinors, unlike equation (49), nor velocity operator averages, unlike equation (56). It has a simple geometric interpretation, which explains why it has to be an integer. Let us consider the image manifold $M$ of the whole BZ through the map $k \rightarrow d(k)$ (figure 19). The number $C_n$ counts how many times the closed manifold $M$ wraps around the origin, because it is the global solid angle $\Omega$ from which this image manifold is seen from the origin divided by $4\pi$. If the origin is outside the image manifold $M$ then the integral vanishes and $C_n = 0$. If the origin is inside the image manifold, then the integrated is a multiple of $4\pi$, and therefore $C_n$ is a finite integer (see figure 18 right). To change $C_n$ it is necessary to change the parameter of the bulk Hamiltonian $d(k)$ in such a way that the bulk gap closes. Below, we will see that an alternative view on a Chern insulator is to consider it as a topological texture in reciprocal space. The corresponding invariant will be shown to be the skyrmion number (which is still another name for the Chern number, the Pontryagin index, the wrapping number, etc). An alternative formula using only the unit vector $\vec{d}$ is:

$$C_n = \frac{1}{4\pi} \int d^2k \left( \partial_x d / \partial k_x \times \partial_y d / \partial k_y \right) \cdot \vec{d}$$ (162)

which is the wrapping number of the map $k \rightarrow \vec{d}(k) = d(k)/|d(k)|$ between the BZ (torus $T^2$) and the unit sphere ($S^2$). In accordance with general classifications, $C_n$ is a single number that characterizes the general structure of wave functions globally in $k$-space. In section 4.4.3, we have seen that this topological number is observable and measures the charge Hall conductance of insulators in units of $e^2/\hbar$. In the following subsection, we evaluate $C_n$ for the Haldane model in its different insulating phases.

Finally, we can relate those Chern number of Bloch bands to expressions for Chern numbers in the TLS case. By integrating over the BZ and injecting the spinor components equation (144), one gets the Chern number:

$$C = \frac{1}{2\pi} \int_{\text{BZ}} dk \sin\theta_k \left( \frac{\partial \theta_k}{\partial k_x} \frac{\partial \phi_k}{\partial k_y} - \frac{\partial \phi_k}{\partial k_x} \frac{\partial \theta_k}{\partial k_y} \right) = \frac{1}{2\pi} \int_{\text{BZ}} d\theta d\phi \sin\theta \ \frac{1}{2},$$ (163)

where BZ′ is the image of the BZ on the Bloch sphere, or equivalently the central projection of the closed manifold $M$ onto the unit sphere. The last equality simply follows from the change of integration variables between the $k$ of the BZ and the polar angles on the Bloch sphere, the quantity between comes being the Jacobian of this transformation. This is the flux of a radial Berry curvature through the unit sphere.
6.4.2. Practical evaluation of the Chern number with a degree formula

Equation (162) expresses the Chern number as a winding number of the map $k \to \hat{d}(k) = d(k)/|d(k)|$, which is the composition of the map $k \to d(k)$ followed by the central projection $\pi$ to the unit sphere. It is possible to transform this integral expression as a discrete sum by using the Brouwer degree of the map. The Brouwer degree reads [166]:

$$C = \sum_{y \in Y_z} \sum_{k \in \partial^{-1}(y)} \text{sign} \left[ (\partial_k d \times \partial_y d) \cdot n \right], \quad (164)$$

where $z$ is an arbitrary point on $S^2$, $n$ the unit vector towards $z$, and $Y_z = M \cap \pi^{-1}(z)$. The procedure for evaluating $C$ consists in choosing a point $z$ on $S^2$ and then determining the intersections of $M$ with the ray originating from $(0,0,0)$ to $z$. These intersections typically form a discrete set $Y_z$ of isolated points of $\mathbb{R}^3$. For our purpose, it is very useful to choose $z$ to be on one of the $\sigma$ axis, let us say $z$ is the north pole of the Bloch sphere. Then we have to sum in equation (164) over points that satisfy $d_z(k) = d_z(k) = 0$.

Let us treat the case of the Haldane model to be more specific. Then $d_z(k) = d_z(k)$ leads to the Dirac points of pristine graphene $k = \pm K$, but note that we are considering the gapped Haldane model, so that importantly $d_z(\pm K)$ is non zero. Near those points the Bloch Hamiltonians read:

$$H(\xi K + q) = H^{(\xi)}(q) = \nu_F (\xi q_x \sigma_x + q_y \sigma_y) + d_z(\xi K) \sigma_z, \quad (165)$$

$d_z = (\xi q_x, \nu y q_y, d_z(\xi K))$. At this point, one needs to consider different cases depending on the signs of $d_z(\pm K)$. Let us first assume that $d_z$ is positive at both Dirac points $\pm K$. Then the set $Y_z$ consists in two points and the Chern number is:

$$C = \sum_{\xi = \pm 1} \text{sign} \left[ (\partial_\theta d_\xi \times \partial_y d_\xi) \cdot e_z \right] = \sum_{\xi = \pm 1} \text{sign}(\nu y^2) = 0, \quad (166)$$

reflecting that the local orientations of the two sides of the surface $M$ are opposite. If $d_z$ is negative at both Dirac points $\pm K$, then set $Y_z = -$ and immediately $C = 0$. Therefore one finds that the valence band is trivial in the $h$-BN model and also in regions of the Haldane phase diagram where the masses have the same sign at the Dirac point.

We now consider the case with opposite signs of the masses, for instance $d_z(K) > 0$ and $d_z(-K) < 0$. In such a case, the ray $z$ intersects the surface $M$ only once and the Chern number is simply:

$$C = \sum_{\xi = \pm 1} \text{sign} \left[ (\partial_\theta d_\xi \times \partial_y d_\xi) \cdot e_z \right] = \text{sign}(\nu y^2) = 1. \quad (167)$$

For the opposite choice of signs, $d_z(K) < 0$ and $d_z(-K) > 0$, the intersection $Y_z$ also reduces to a single point but which maps back to valley $\xi = -1$ in the BZ. Hence $C = -1$. The above formulas can be symmetrized by considering the whole axis $z$ instead of the ray $z$, leading to:

$$C = -\frac{1}{2} \sum_{\xi = \pm 1} \text{sign}(M_\xi) = \frac{1}{2} \left[ \text{sign}(M_-) - \text{sign}(M_+) \right], \quad (168)$$

where $M_\xi = d_z(\xi K)$.

Finally, the above formula has been obtained in a more general way in [166] which is valid for any two-band insulator and reads:

$$C = \frac{1}{2} \sum_{k_0} \chi(k_0) \text{sign} \left[ d_z(k_0) \right], \quad (169)$$

Figure 19. Mapping $k \to \hat{d}(k) = d(k)/|d(k)|$ between the BZ (torus) and the Bloch sphere. Each point on the unit sphere represents a spinor parametrized by the angles $\theta_k$ and $\phi_k$. 

49
where \( \chi(k_0) = \pm 1 \) denotes the chirality or winding number of the Dirac point located at \( k_0 \) (for example, in graphene, this chirality is \( \chi = \text{sign det}(\eta_{ij}) = \xi \) and sign \([d_i(k_0)] = \pm 1 \) is the sign of the Dirac fermion mass. Lattice Dirac fermions always occur in pairs (fermion doubling theorem), which makes the Chern number an integer. Each massive Dirac fermion contributes \( \pm 1/2 \) to the Chern number. However, it is not always easy to identify all massive Dirac fermions from the energy spectrum alone. Indeed, there are often cases with spectator Dirac fermions of very large mass that are totally undetectable in the energy spectrum. Let us illustrate that in the case of two bands with \( H(k) = d(k) \cdot \sigma \). The way to identify Dirac fermions correctly is to find points in reciprocal space at which two (out of three) components of \( d(k) \) vanish (e.g. \( d_i(k_0) = d_i(k_0) = 0 \)). The third component \( d_i(k_0) \) giving the mass of the Dirac fermion (it can not be zero otherwise the system would be gapless at \( k_0 \)). The choice of components is arbitrary: it does not matter which two components vanish and which plays the role of the finite mass. Spectator fermions are discussed in [16] and also in section 8.2 of [18].

Let us now consider the example of graphene in presence of some inversion breaking and time-reversal breaking terms. So the mass matrix is \( (M_3 - 3\sqrt{3}t_2 \sin(\phi)\xi\sigma_z) \) implying that the gap can close for \( M_3 = 3\sqrt{3}t_2 \sin(\phi) \) in one valley (labelled by \( \xi = \pm 1 \)). This equality signals a one-electron topological quantum transition separating a QAH insulator and a trivial atomic insulator. Finally we would like to make a comment on the terminology. This type of phase transition is purely a change between two one-electron Hamiltonians. It has in particular nothing to do with TO defined by Wen [167, 168]. In particular the transition discussed here is not a transition between two TOs. It is rather a transition between two band-insulators having distinct topological invariants (which characterize the winding of one-electron wave functions). More on this distinction in the conclusion, see section 9.

When proposing his model, Haldane had a two-fold goal [16]. One was to find a QHE without Landau levels, i.e. as a band effect, breaking time-reversal symmetry but preserving translation symmetry. This goal was achieved. His other goal was to have a condensed matter realization of the parity anomaly known to occur in the 2D massless Dirac equation. One manifestation of this anomaly (i.e. a symmetry which exists classically but does not survive quantization) is the fact that, in a perpendicular magnetic field, there should be a half-quantized QHE. However, in the Haldane model on the honeycomb lattice, Dirac fermions always occur in pairs even if one member of the pair has a zero mass and the other has a very large mass (of the order of the bandwidth). Therefore, no half-quantized Hall effect occurs in the Haldane model and the parity anomaly of one Dirac fermion (say the massless one in one valley) is compensated by the other Dirac fermion (say the spectator one in the other valley). Hence, from this perspective the second goal was not achieved. Achieving this parity anomaly requires a further trick, namely to produce a single 2D massless Dirac equation at the surface of a 3D topological insulator [32–34].

### 7. Topological insulators: a bigger picture

We have already discussed in detail a few examples of topological insulators, including the Haldane model of spinless fermions on the honeycomb lattice (an intrinsic topological insulator), and the Rice–Mele chain with inversion symmetry (a topological crystalline insulator). We have also alluded to the Kane–Mele model as an example of time-reversal invariant topological insulator (i.e. a topological insulator protected by an extraordinary symmetry such as time-reversal, charge conjugation or chiral). In this section, we build upon those representative examples and enlarge our point of view to describe the current ‘standard model’ of insulating topological phases. Below, we give two definitions of a topological insulator. The first (see section 7.1) mainly involves cell-periodic Bloch states in reciprocal space and focuses on bulk invariants and quantized electromagnetic responses. The second (see section 7.2) mainly involves Wannier functions in real space and focuses on their localization. Topological metals will be the subject of the next section 8.

#### 7.1. Bulk quantized electromagnetic response and gapless boundary states

In this section, we use a characterization of topological insulators in reciprocal space focusing on two ingredients: a bulk topological invariant (usually associated to a quantized electromagnetic response) and the related existence of gapless surface states. We first discuss intrinsic topological insulators and then SPT insulators. In the latter case, we further distinguish between extraordinary symmetries (time-reversal, charge conjugation or chiral) and crystalline symmetries (such as rotations, mirrors, inversion, etc). These three symmetries are called extraordinary because either they are anti-unitary (time-reversal and charge conjugation) or/and they anti-commute with the Bloch Hamiltonian (charge conjugation and chiral symmetry). We refer to the first case simply as SPT insulators and to the second case as topological crystalline insulators.
7.1.1. Intrinsic topological insulators

A well-known topological insulator, the Haldane model [16], has been discussed in detail in this review (see section 6.3). This is the telltale example and the representative of a larger class of topological insulators, called Chern insulators or QAH insulators. Its phase diagram shows several gapped phases separated by lines where the gap closes and identified by a specific value of the Chern number $C_-$ of the occupied valence band (figure 18, left). The Chern number of a band is a topological invariant which measures a global property of the Bloch wave functions of this band over the whole BZ. It can only be defined in the presence of a finite bulk gap. Each gapped phase is robust in the sense that modifying continuously the parameters of the model will not change the Chern number. The only way to jump from an integer value of the Chern number to another one is to cross a gap closing. At the gap closure, $C_-$ is ill-defined as can be seen from the formula (161) where the denominator $d$ vanishes, or in more geometric terms, the manifold $\mathcal{M}$ crosses the origin which is the source of the Berry curvature flux (figure 18, right). Note that the total Chern number, including the two bands of the Haldane model (the occupied valence and the empty conduction band) is always zero: $C = C_- + C_+ = 0$. Each band can separately carries both zero Chern number (trivial phase), or opposite $C_\pm = \pm 1$ (topological phase). More generally, a Chern insulator is a two-dimensional band insulator that breaks time-reversal symmetry and that has a non-vanishing total Chern number of the occupied bands at zero temperature (it may have much more bands than the Haldane model). The QH insulator is a close relative where the translational invariance is broken by the magnetic field and replaced by the Landau level structure in the continuum, and eventually magnetic subbands and Hofstadter spectrum in presence of a periodic potential.

We now turn to the characterization of topological insulating phases in terms of macroscopic observables, like the transverse Hall conductance which is a bulk response function odd with respect to time-reversal. Via the TKNN formula (83), the Hall conductance is directly proportional to the TKNN number with a prefactor $e^2/h$ depending only on fundamental constants. There is a subtle distinction here between the TKNN number and Chern numbers. A Chern number characterizes a band (or a group of bands) whereas a TKNN number characterizes a gap. The TKNN number is simply the sum of the Chern numbers of the occupied bands below the gap in which the Fermi level is assumed to be. This explains why each insulating phase is distinguished by a robust and quantized Hall conductance numbers of the occupied bands below the gap in which the Fermi level is assumed to be. This explains why bands) whereas a TKNN number characterizes a gap. The TKNN number is simply the sum of the Chern numbers of the occupied bands below the gap in which the Fermi level is assumed to be. The gap closing makes possible the presence of gapless conducting states at the interface between distinct topological phases. At such an interface, the parameters of the model have to be tuned in such a way that the Chern number jumps by an integer number, which implies a region where the gap closes. The gap closing makes possible the presence of gapless conducting states at the interface between topologically-distinct phases. In addition, the existence of chiral gapless modes is guaranteed by the bulk-edge correspondence and can also be understood as following from the Jackiw–Rebbi mechanism (see section 5.4). Bulk-edge correspondence states that the number of gapless chiral modes per edge is equal to the jump of the bulk topological invariant [178]. These boundary modes have unique properties, meaning that they typically cannot be realized in a lattice system. For Chern insulators, the 1D edge states are chiral, i.e. they circulate only in one direction around the insulating bulk (see figure 5). The chiral nature of motion indicates the time-reversal breaking. Such chiral fermions can only emerge at the boundary at a non trivial insulator: there are no 1D chain or wire with unidirectional electrons. In the QHE, the direction of motion is determined by the orientation of the magnetic field, while in the Haldane model it is determined by the internal inhomogeneous magnetic field. These modes are protected from backscattering by impurities or disorder because there is no available state for the carrier to back scatter on the same edge and tunneling to the opposite edge is exponentially forbidden.

From the above discussion, we can formulate a first definition of a topological insulator. It is a bulk band insulator, that is characterized by a non-zero integer, called a (bulk) topological invariant, and a robust quantized bulk response. To change this bulk topological invariant it is necessary to close the bulk gap. The edges of a topological insulator necessarily host gapless modes that are robust because they are chiral and...
therefore immune to disorder. In addition there is a relation between the number of these edge modes and the bulk topological invariant: this is the so-called bulk-edge correspondence. The case of the Chern insulator is particular in that it requires no symmetry for its protection and is truly robust to any perturbation, as long as it does not close the bulk gap. It is intrinsically topological.

### 7.1.2. Symmetry-protected topological insulators

It was long believed that broken time-reversal symmetry and space dimension two were a mandatory setting for the appearance of topological phases. In 2005, Kane and Mele [28, 29] uncovered a new type of topological insulator where time-reversal plays the role of a protective symmetry for metallic edges, and allows one to draw a topological distinction between two insulating states. They introduced a model of spinful fermions in graphene consisting in two independent copies of the Haldane model respectively attached to each spin projection, each copy being the time-reversed of the other, e.g. with Chern number $C_\uparrow = +1$ for spin-up fermions and $C_\downarrow = -1$ for spin down fermions. As a result the overall Kane–Mele model is time-reversal invariant and describes graphene with a very particular intrinsic SO coupling conserving the spin projection $S_z$. The total Chern number of the occupied bands is $C = C_\uparrow + C_\downarrow = 0$ (no QHE), but there is a $\mathbb{Z}_2$ topological invariant $\nu = (C_\uparrow - C_\downarrow)/2$ [2], which distinguishes only two classes: trivial ($\nu = 0$) and non-trivial ($\nu = 1$).

Because the $S_z$-conserving Kane–Mele model consists in two independent copies of a Haldane model with opposite Chern numbers, it has counterpropagating edge states with a full spin-momentum locking property: spin up propagates in only one direction, and spin-down propagates in the opposite direction. This spin-momentum locking insures that backscattering by a non-magnetic impurity (i.e. an impurity that respects time-reversal symmetry) is not possible. One could expect that the gapless helical edge states are an artefact of this $S_z$ conserving juxtaposition of two chiral edge states, and that any spin mixing perturbation would mix the two spins and gap the edge states. In fact the gapless character of the helical edge states is robust again both staggered potential perturbations and Rashba SO perturbations (mixing spin up and down projections) provided those perturbations do not close completely the bulk gap. With those time-reversal symmetric perturbations, the only way to get rid of the helical edge state is to close the bulk gap. In contrast, a time-reversal breaking perturbation (like a Zeeman effect due to transverse magnetic field in $x$ or $y$ directions) would immediately spoil the metallic character of the helical edge. The crossing of the left and right movers branches is protected by Kramers’ theorem as $T^2 = -1$. In principle, it is possible to construct models with a larger number of pairs of counter-propagating modes. If the number of Kramers’ pairs is even, then disorder may eventually gap out all the edge states. In contrast, for an odd number of Kramers’ pairs there will always be a gapless pair left as long a the bulk gap is finite and time-reversal symmetry is obeyed. This expresses that the bulk $\mathbb{Z}_2$ index is the parity of the number of Kramers’ pairs of edge states. Intrinsic SOC, at the heart of the Kane and Mele proposal, is too small in graphene for the effect to be measurable. However, QSHIs have been realized experimentally in HgTe/CdTe quantum wells following the theoretical proposal by Bernevig, Hughes and Zhang, see [18] for review.

In contrast to the Chern insulator (QHE), that only exists in two space dimension [179], this type of time-reversal symmetric insulator has a three-dimensional counterpart known as a strong topological insulator [32–34]. The strong topological insulator is also classified by a $\mathbb{Z}_2$ index and also has a topological metal on its surface. It is a single massless two-dimensional Dirac fermions that is spin-momentum locked. It appears similar to graphene, except that it has a single (instead of four) Dirac cone. Indeed, the real spin is actually the one appearing in the Dirac equation (and not a sublattice pseudo-spin as in graphene) and there is also no valley degeneracy. Such a strange metal can only exist as a surface mode of a higher dimensional system and has a true parity anomaly and a half-quantized QHE.

In conclusion, the $\mathbb{Z}_2$ topological index and the helical edge states are both protected by time-reversal symmetry. There is no smooth path (in the space of Hamiltonians) connecting the $\nu = 1$ and $\nu = 0$ phase while simultaneously keeping the bulk gap open and respecting the time-reversal invariance. To change the $\mathbb{Z}_2$ topological index, it is necessary either to use time-reversal breaking perturbations or to close the bulk gap. In 2D, the odd and even phases of the Kane–Mele model can be characterized by a bulk quantized response which is a spin Hall conductance. Nevertheless, this only holds for the $S_z$ conserving models, and it is difficult to build a macroscopic observable that could play the role of the Hall conductance for Chern insulators. The bulk characterizations of the 2D or 3D time-reversal invariant phases relies on $\mathbb{Z}_2$ pumps [180, 181]. The 3D strong topological insulators have a specific orbital magnetolectric polarizability described by axion electrodynamics (with a topological $\theta$-term similar to the one discussed in the 1D context, see equation (123)) that can play such a role [182].

The ideas of Kane and Mele have been extended to extraordinary symmetries such as time-reversal, particle-hole (charge conjugation) and chiral and to the concept of SPT insulators. There is a general classification of topological insulators (and superconductors) for non-interacting fermions that is known as...
the ten-fold way periodic table [41, 42, 128, 129]. Depending on space dimension and on three extraordinary symmetries resulting in ten Altland-Zirnbauer classes, it indicates whether there is a single type of band insulator (only trivial), two types of band insulators ($\mathbb{Z}_2$ index separating trivial from non-trivial band insulators) or as many band insulators as relative integers $\mathbb{Z}$. Kitaev has realized that, as a function of space dimension $D$, there is a form of regularity or pattern in this table related to real and to complex K-theory and known as Bott periodicity [42]. Such topological insulators that are protected by a extraordinary symmetry are also called strong topological insulators [183] as they resist the absence of crystalline symmetries that we study next.

7.1.3. Topological crystalline insulators

So far we have discussed topological band insulators protected by time-reversal symmetry, which is a fundamental and generic symmetry of space-time. As initiated by Liang Fu [136], it is also possible to use spatial symmetries (e.g. rotations or reflections) of the crystal itself to protect and maintain topological distinctions between various insulating phases. Such phases have been coined topological crystalline insulators (TCI) and have also been classified (see e.g. [184, 185]).

We have already discussed a very simple example, namely an inversion-symmetric insulator in 1D. In section 5.5, we have seen that provided inversion symmetry is maintained, it is impossible to connect the SSH insulating chain (at $\delta \neq 0$ and $\Delta = 0$) to the CDW insulator (at $\delta = 0$ and $\Delta \neq 0$). Both phases are characterized by a distinct robust observable, which is the electric polarization. It is vanishing in the SSH phase and quantized in the CDW phase.

This simple example of a quantized electric polarization in 1D has been generalized to quantized electric multipole insulators [120] leading to the concept of higher order topological insulators (HOTI) [186]. These latter systems are insulators that do not possess gapless surface states but rather gapless hinge states, which are gapless states living on the boundary of a boundary. For example, a 2D second order topological insulator known as the quadrupole insulator does not host 1D gapless edge states but possesses 0D gapless corner states for a square patch protected by mirror symmetries. In other words, the surface of this second order topological insulator is not a metal but is itself a 1D topological insulator. HOTI escape the simple bulk-boundary correspondence between a $D$-dimensional insulator and $D−1$ metal but rather feature a correspondence to $D−2$ or $D−3$ gapless states. They only exist under the protection of point-group symmetries.

7.1.4. Conclusion: first definition of a topological insulator

A topological insulator is a band insulator (it has a bulk gap) characterized by a bulk topological invariant computed from its cell-periodic Bloch states (e.g. a Chern number or a $\mathbb{Z}_2$ invariant). This invariant is typically related to a quantized electromagnetic response (e.g. a quantized Hall conductivity or a quantized electric polarization). In order to change this bulk topological invariant it is necessary to close the bulk gap or to break the protecting symmetry. In a finite sample with a surface that does not break the protecting symmetry, a topological insulator host gapless surface (or hinge) states that have some form of robustness towards perturbations that respect the corresponding symmetry.

7.2. Obstruction to exponentially-localized and symmetric Wannier functions

The picture of the previous section focuses first on the topology of Bloch states in the bulk and second on the eventual gapless boundary states (and their dispersion in momentum space) as a hallmark signature of a topological insulator. Now we present an alternative point of view based on the real-space localization of the Wannier functions in the bulk (see section 4.3.3).

According to Kohn [187], a (trivial) insulator is exponentially insensitive to boundary conditions. This is similar to an atomic insulator, which can be seen as a periodic solid made from bringing closer initially isolated atoms. As in isolated atoms, valence electrons remain close to their nuclei (or ions). In the language of solid-state physics, the Wannier functions are exponentially-localized on the ions. In the case of an atomic insulator, the Wannier functions are simply the atomic orbitals. As a consequence, the many-body ground-state (for non-interacting electrons) is barely sensitive to what happens near the boundary of a macroscopic sample. In contrast, a topological insulator is a band insulator that is not smoothly or adiabatically connected to an atomic insulator. Smoothly means without closing the gap for intrinsic topological insulators, and without closing the gap while keeping some protective symmetry for SPT insulators.

Such a viewpoint of a topological insulator featuring an obstruction to exponentially-localized and symmetric Wannier functions has been put forward recently [140, 188] and is sometimes called topological quantum chemistry (see [183] for review).
7.2.1. Intrinsic topological insulators

Bloch functions and Wannier functions are two dual possible representations of the electronic states of a crystal. Intrinsic topological insulators (Chern insulator and QH states) are characterized by obstructions in the realization of either of these representations. First, we explain the obstruction for the Bloch states in reciprocal space, and second the corresponding veto for exponentially-localized Wannier functions in real space.

A band with a finite Chern number implies an obstruction in finding a unique gauge for the cell-periodic Bloch states which would be smooth over the whole BZ. Let us consider the explicit case of the Haldane model and use the Bloch sphere representation to visualize the spinors (cell-periodic Bloch states) of the valence band. In the trivial phase, $d_z(k)$ being always positive it is possible to pick the $(n)$-gauge which has no singularity except at the south pole. In the topological phase, $d_z(k)$ changes sign somewhere in the BZ, and the spinor explores both north and south poles: it is not possible to use a single gauge and one has to use the Wu-Yang construction with overlapping $(n)$ and $(s)$ regions. From this perspective, going from a trivial to a topological phase requires a band inversion (in one of the valley for the Haldane model) which necessarily occurs via a gap closing. In other words, there is an obstruction to smoothly or adiabatically connect a topological insulator to an atomic insulator. Note that the vacuum is considered to be a trivial (atomic) band insulator. This obstruction is also a well-known property of the QH states [189]. Reciprocally, without this obstruction, it would be possible to apply Stokes’ theorem to the whole BZ and to obtain that the flux of the Berry curvature vanishes in contradiction with the hypothesis of finite Chern number.

Let us now turn to the Wannier function representation. Wannier functions are the Fourier transforms of Bloch states that, in conventional insulators, are usually exponentially-localized around specific locations of the solid, the Wannier centers. The Wannier functions are not unique, because there is a gauge freedom to define the Bloch functions. Changing, locally in $k$, the $U(1)$ phases of the Bloch functions results in changing the spread of each Wannier function, while keeping the Wannier center of the band (or isolated group of bands) invariant. Marzari and Vanderbilt [190] devised a procedure to compute the maximally localized Wannier functions of a given lattice model, see [80] for a review. Going back to the Haldane model, Thonhauser and Vanderbilt have shown that this procedure fails at the topological transition, where the spread of the Wannier functions diverges [97]. Indeed, at the transition, Wannier functions are completely delocalized, as in any metallic gapless state. For a band with finite Chern number, it is not possible to build a basis of exponentially-localized Wannier functions, meaning that they are only algebraically localized [96, 97]. This is in fact a general property shared by all Chern insulators [95] and by QH insulators [96].

Finally, this impossibility to represent a topological phase in terms of localized basis functions shows that the quantum degrees of freedom are not stored locally in such phases. As a plausible consequence, the non-triviality of such phases should be evidenced by their sensitivity to twists in the boundary conditions. For the QHE, Niu et al [175] gave a real-space expression for the Chern number. This requires putting the system on a real space torus and twisting the boundary conditions by inserting fluxes in the inequivalent non-contractible loops of the torus. The space of boundary conditions replaces the BZ. The beauty of this formulation is that translation invariance is no longer needed. The topological invariant can be defined directly for the many-body ground-state (instead of ‘band by band’) and is valid also for a disordered (and in some cases interacting) system.

7.2.2. Symmetry-protected topological insulators

The idea of obstruction is also relevant to investigate SPT phases, but here the obstruction is based on the fact that one enforces a protective symmetry. Without such a constraint, the obstruction would generally be lifted.

For concreteness, we consider the case where the symmetry is time-reversal and first discuss real space. According to [95], it is always possible to construct a basis of exponentially-localized Wannier functions for time-reversal invariant insulators both in the even (trivial) and odd (topological) phases. This is true but it comes always at a price, which is that one cannot maintain the time-reversal invariance of the exponentially-localized Wannier functions in the topological phase [191, 192]. In a topological insulator, there is an obstruction to finding symmetric and exponentially-localized Wannier functions [140, 188].

In reciprocal space, the obstruction is manifest in the fact that it becomes impossible to find a smooth gauge for the whole BZ that respects time-reversal symmetry. Although it is possible to find a smooth gauge over the whole BZ (because the Chern number vanishes), it necessarily breaks time-reversal symmetry. This is nicely reviewed in [57].

7.2.3. Topological crystalline insulators

We start by considering the simplest possible 1D monoatomic chain and search for topological insulator phases protected by inversion symmetry (the simplest point-group symmetry). The two possible configurations that respects inversion symmetry correspond to either having electrons wave functions
localized at every site, or alternatively having them localized at the center of each bond. Those two situations have respectively electric polarization $0$ and $P_g/2$ modulo the quantum of polarization $P_q$. Of course it seems very simple to go continuously from one to the other situation by simply shifting all electrons by half the lattice constant. But this way breaks inversion symmetry at every steps between the two configurations. Alternatively one may keep inversion symmetry at all stage if one first spreads the wave functions in a symmetric way before reconcentrating them around the new localization centers. Nevertheless this latter protocol goes through a fully delocalized state that is gapless. In conclusion, it seems impossible to go continuously from the configuration $P_{\text{tot}} = 0$ to the configuration $P_{\text{tot}} = P_g/2$ while simultaneously keeping the whole system gapped and inversion symmetric at each intermediate steps. Such a topological transition can be implemented in a 1D model with coupled $s$ and $p$ bands [21, 141]. The trivial phase is the atomic insulator (a filled $s$ band) with vanishing polarization and the topological phase is a covalent insulator (a filled $sp$ hybridized bonding band) with quantized polarization.

 Actually, these two 1D insulating phases have exponentially-localized and inversion-symmetric Wannier functions. They are therefore both eligible for being atomic limits [188]. One is a naive atomic limit (the atomic insulator with Wannier center on the ions), while the other has been called an obstructed atomic limit (the covalent insulator with Wannier center on the bonds) [140]. What distinguish the two is whether the Wannier center is located on the ions or mid-bonds. The conclusion is that in this example, there is no obstruction in finding symmetric and localized Wannier functions. Still, in a finite chain, as shown by Shockley, the covalent insulator hosts edge states, but not the atomic insulator [141].

A general theory based on the obstruction in finding symmetric localized Wannier functions has recently been developed [140, 188] and is reviewed in [183] (it is sometimes called topological quantum chemistry). It is based on the key idea of a mismatch or a compatibility between the symmetry of atomic orbitals in real-space and the topology of bands in the BZ. First, starting from well-localized atomic orbitals, the notion of elementary band representations is used to build all possible atomic limits. These are essentially the cases where symmetric localized Wannier functions exist. Second, starting from the opposite view of delocalized electrons in Bloch states, one considers all possible band structures, which defines quasi-band representations. Third, one analyses the compatibility of those two views of band theory. Topological semimetals correspond to the case where compatibility relations cannot be satisfied. When they can be satisfied, as there are more quasi-band representations than elementary band representations, one defines atomic limits (i.e. trivial insulators) as being the quasi-band representations that are elementary band representations. There exists several such limits and hence the notion of obstructed atomic limits. Quasi-band representations that are not elementary band representations are called topological insulators. In order to detect such topological insulators it is possible to use symmetry-based indicators [188] following the initial idea of Fu and Kane for inversion-symmetric topological insulators [32]. Another interesting notion that emerges from this framework is that of fragile topological bands [193]. Using these ideas, several catalogues of topological materials based on the 230 space groups appeared recently [194–196].

7.2.4. Conclusion: second definition of a topological insulator
An atomic limit corresponds to exponentially-localized and symmetric Wannier functions. A trivial insulator is such that it can be smoothly continued to an atomic limit. Due to the different possible protecting symmetries, there may be several distinct atomic limits and therefore several types of trivial insulators (see the notion of an obstructed atomic limit). A topological insulator is such that it cannot be continued to an atomic limit without closing a gap or breaking a protecting symmetry. In a SPT insulator there is an obstruction to finding symmetric and exponentially-localized Wannier functions.

This second definition of a topological insulator is actually more stringent than the first one. For example, according to the first definition, a 1D inversion-symmetric band insulator with quantized electric polarization (e.g. the CDW chain realizing an ionic insulator, or the Shockley model of coupled $s$ and $p$ bands in the covalent insulator phase) is a topological insulator. But according to the second definition, a 1D inversion-symmetric band insulator is always trivial: it is either a naive atomic limit if the Wannier centers are on the ions (e.g. the CDW chain, or the Shockley model in the atomic insulator limit) or an obstructed atomic limit if the Wannier center are on the bonds (e.g. the SSH chain realizing a molecular insulator, or the Shockley model in the covalent insulator limit).

8. Topological metals: Fermi surface as a topological defect

Up to now, we have mainly dealt with band insulators having non-trivial topology, i.e. topological insulators. Here, we want to expand the picture and consider also metals, i.e. fermionic systems having a Fermi surface, and their topological characterization. For example, graphene is a peculiar 2D semi-metal, in which the Fermi surface is limited to two points corresponding to the contacts between two bands. The idea of
classifying Fermi surfaces as topological defects goes back to Volovik (see [45] for review, the original idea dates from the 1980s), following the pioneering work of Lifshitz [197] in the 1960s. The latter understood that there could be a phase transition in a metal that is not related to a change in symmetry but rather to a change in the topology of the Fermi surface. Is this surface made of a single piece or of two disconnected pieces, for example? Volovik later took the step of considering Fermi surfaces as topological defects in reciprocal space and of treating either the inverse Green function (or single-electron propagator) or the Bloch Hamiltonian as a kind of order parameter. These systems are sometimes called topological (semi-)metals. An extension of the idea of a topological defect (namely a topological texture) will eventually lead us to an alternative point of view on topological insulators.

8.1. Introduction to homotopy groups for topological defects and textures

These ideas followed from earlier developments in the classification of topological defects in real space using homotopy groups, that we now review. This subject flourished in the 1970s mainly under the impulsion of Mineev, Volovik, Toulouse, Kléman and Mermin (see [198] for review, see also the very accessible lectures notes by Sethna [199]). These authors have shown that, in the context of phase transitions resulting from spontaneous symmetry breaking and characterized by an order parameter, defects (i.e. zeroes) in the order parameter could be systematically classified using homotopy theory. We will use similar ideas to discuss Fermi surfaces as if they were topological defects.

Consider the ordered phase of a 3D ferromagnet. The order parameter is the magnetization \( \mathbf{M} \), its direction \( \mathbf{n} \) is uniform at equilibrium and its norm \( |\mathbf{M}| \approx M_0 \) is almost constant at low temperature. Spatially varying the direction of the magnetization corresponds to excitations in the system. If these variations are small perturbations around equilibrium, these are known as spin waves (or generically as soft or Goldstone modes). But there are other type of excitations of the order parameter, which are not smooth and small perturbations. These corresponds to situations in which the order parameter vanishes on points or lines or planes and are known as defects. Some of these defects are remarkably stable due to a topological protection. A topological defect is a singularity of the order parameter that can not be patched or repaired by any local rearrangement of the microscopic spins. In order to characterize a singularity as a topological object, one proceeds in four steps:

First, one needs to trap the defect inside a cage \( C \). What is the dimensionality of this cage? If the defect is point-like and space is 3D, then the cage is a sphere \( S^2 \) of dimensionality 2. If the defect is line-like, then the cage is a circle \( S^1 \) of dimensionality 1. The general formula for the corresponding dimensions is known as the hunter’s rule and reads [200]:

\[
\text{space} = \text{defect} + \text{cage} + 1.
\] (170)

It is often written \( d = d' + r + 1 \), where \( d \) is the space dimension, \( d' \) the defect dimension (0 for a point, 1 for a line, 2 for a plane, etc), \( r \) the cage dimension (2 for a sphere \( S^2 \), 1 for a circle \( S^1 \), 0 for two points \( S^0 \)).

Second, one identifies the order parameter space \( V \), which is the remaining freedom for the order parameter once deep in the ordered phase. For the ferromagnet, the magnetization is a 3-dimensional vector. In the ordered phase at low temperature, its norm is essentially fixed, only its direction has some freedom and therefore \( V = S^2 \).

Third, in order to represent the configuration of the magnetization around the defect, one needs to consider maps from the cage space \( C \) to the target space \( V \). For example, for a point-like defect in a 3D ferromagnet where the magnetization \( \mathbf{M} \approx M_0 \mathbf{n} \):

\[
\mathbf{n} : C = S^2 \rightarrow V = S^2
\]

\[
r \rightarrow \mathbf{n}(r).
\] (171)

Fourth, we need to use a mathematical tool from topology known as homotopy groups. Homotopy groups answer the following question: how many inequivalent classes of maps exist from \( C \) (where \( C \) is a sphere \( S^2 \)) to \( V \)? By inequivalent, we mean in a topological sense: we are allowed to smoothly deform the cage space that we are applying onto the target space but we are not allowed to cut it or paste it. Homotopy groups are noted \( \Pi_r(V) \). In the present case, we are interested in the second homotopy group of the sphere \( S^2 \). This is \( \Pi_2(S^2) = \mathbb{Z} \) as can be found in tables. This means that there are as many inequivalent classes as there are integers. And therefore that point-like defects in a 3D ferromagnet can be attributed an integer topological charge. In the present context, it is known as the wrapping number.

In order to better understand the notion of a homotopy group, we give several simple examples before going back to the wrapping number. Consider maps going from a cage \( C = S^1 \) to a target space \( V = S^1 \). We want to know how many inequivalent ways there are to map a circle onto a circle or, to put it differently, to
draw a closed path on a circle. This is given by the first homotopy group (a.k.a. the fundamental group) of the circle $\Pi_1(S^1) = \mathbb{Z}$ and is known as the winding number $W$ (see figure 20). Another example concerns maps from the circle $C = S^1$ to the sphere $V = S^2$. This is actually the trivial group $\Pi_1(S^2) = 0$, which means that there is a single class of closed paths on a sphere. All paths can be smoothly deformed to a null path, or to paraphrase Sethna [199]: 'one cannot lasso a basketball'. What about closed path on the torus? We are interested in maps from the circle $C = S^1$ to the torus $V = T^2$. As we have seen, the torus is the direct product of two circles $T^2 = S^1 \times S^1$. There are therefore two fundamentally distinct non-contractible loops on a torus. The first homotopy group is therefore $\Pi_1(T^2) = \mathbb{Z} \times \mathbb{Z}$. And classes of closed paths on the torus are characterized by a pair of integers.

As we have seen, homotopy groups can be generalized from the first homotopy group by considering maps that start from other spheres then $S^1$. For example starting from $S^2$ or from $S^0$. In the latter case, the zeroth homotopy group, by convention, gives the number of connected components of the target space. For example, $\Pi_0(S^0) = \mathbb{Z}_2$ because $S^0$ consists of two points, whereas $\Pi_0(S^r) = 0$ when $r \geq 1$.

The wrapping number corresponding to the above $\Pi_2(S^2) = \mathbb{Z}$ answers to the question: how many inequivalent ways of wrapping a sphere with a sphere? A pictorial way is to imagine that the target space is a basketball and the cage space is a specially designed spherical bag made of rubber and with a zipper that is used to carry the ball. A dummy would leave the ball outside the closed bag, corresponding to a wrapping of $0$. When the ball is inside the bag it corresponds to a wrapping of $+1$. Wrapping the ball twice with the bag (harder to imagine!), would correspond to $+2$. One could also turn the bag inside-out before enclosing the ball, which would correspond to a wrapping of $−1$.

Going back to our initial question, we have seen that point-like defects in a 3D ferromagnet carry an integer topological charge known as the wrapping number. The elementary topological defect with wrapping $+1$ is known as the hedgehog (or Bloch point) in the context of ferromagnets. One can picture it as a sphere, the surface of which carries outward pointing arrows (see figure 21).

The stereographic projection of the hedgehog on the 2D plane, gives rise to another interesting topological object known as a texture (see figure 21). In the present case the texture is called a (baby or 2D) skyrmion. A texture is not a defect because there is no singularity (no zero) in the order parameter on the 2D plane. In addition, it is related to special boundary conditions. Here, because of the stereographic projection sending the north pole of the sphere to the points at infinity on the plane, we see that on the boundary on the plane at infinity, all vectors $n$ point up. Note that the hedgehog is a topological point-like defect for a 3D ferromagnet, whereas the skyrmion is a topological texture (not a defect) for a 2D ferromagnet. Topological textures (also known as topological solitons or topological configurations) are discussed in [202] and are classified by relative homotopy groups. Topological textures are interesting in that they extend the topological ideas of homotopy groups to cases that have no defects. The price to pay is that the boundary conditions should be fixed.

### 8.2. 3D Weyl semi-metal as topological defect

We are now back to band theory. Consider a contact point between two bands in 3 space dimensions. This can be modeled using a $2 \times 2$ Bloch Hamiltonian:

$$H(k) = d(k) \cdot \mathbf{\sigma},$$  

(172)

where $\mathbf{k} = (k_x, k_y, k_z)$. The contact point correspond to $\mathbf{k}$, such that $d(\mathbf{k}_c) = 0$. When considered as a topological defect, it should be characterized as follows. We first need to put this defect into a cage $C$. According to the hunter’s rule (170), the cage has dimension 2 as the defect is 0-dimensional. Here the cage can be taken as a 2-dimensional sphere $C = S^2$ surrounding the contact point. The target space $V$, which in the context of phase transitions is known as the order parameter space in the ordered phase, is here the Bloch sphere $S^2$ spanned by $d = d/|d|$. The defects are characterized by the inequivalent classes of maps going from $C = S^2 \to V = S^2$, i.e. by the second homotopy group $\Pi_2(S^2) = \mathbb{Z}$. The corresponding topological invariant

![Figure 20. First homotopy group of the circle and winding number W. Examples from W = +3 to −2.](image-url)
Figure 21. Hedgehog (topological defect of a 3D ferromagnet) above and its stereographic projection into a skyrmion (topological texture in a 2D ferromagnet) below (adapted from [201]). In reciprocal space, the same figure would be an illustration of the relation between a 3D Weyl semi-metal (above) and a 2D Chern insulator (below).

is the wrapping number $N_3$ (3 refers to the co-dimension of the defect, which is $\text{codim} = \text{space–defect} = 3–0$). Here we are using the notation of Volovik $N_{\text{codim}}$ for the topological indices [45]. The wrapping number is actually also a Chern number in the present context and reads:

$$N_3 = \frac{1}{2\pi} \int_{S^2} ds \cdot F_n(k).$$  \hfill (173)

A Weyl point corresponds to an elementary such topological defect also known as a hedgehog and having a wrapping number or chirality $N_3 = \pm 1$. In the vicinity of the contact point, the Bloch Hamiltonian is:

$$H(k_c + q) = H_W(q) = \sum_{ij} v_{ij} q_i \sigma_j,$$  \hfill (174)

where $q = k - k_c$ and $i, j = x, y, z$. In that case, the wrapping number (or chirality) is given by $N_3 = \text{sign} \det(v_{ij})$. Such a Hamiltonian was proposed by Weyl shortly after the discovery of the Dirac equation in order to describe hypothetical chiral fermions [203]. A Weyl fermion can be roughly thought as being half a Dirac fermion in 3D, in the sense that its Hamiltonian is $2 \times 2$ rather than $4 \times 4$. In addition chirality implies masslessness. For a nice discussion of Dirac versus Weyl versus Majorana fermions, see [162]. In the original version, the Weyl Hamiltonian reads:

$$H_W = \pm c \mathbf{p} \cdot \sigma,$$  \hfill (175)

where $c$ is the velocity of light and the $\pm$ indicates the chirality (i.e. left or right Weyl fermion). The dispersion relation is $E = \pm |\mathbf{p}|$.

An alternative way to understand that a contact point between two bands in 3D is a stable (topological) defect is to realize that the Bloch Hamiltonian being a $2 \times 2$ matrix can be decomposed onto the three Pauli matrices (the identity matrix plays no role) and that the three corresponding coefficients $d = (d_x, d_y, d_z)$ are each analytical functions of three variables $k = (k_x, k_y, k_z)$. A contact point is a triplet $k_c$ such that $d(k_c) = 0$. As there are three linear equations ($d_j = 0$) and three unknowns ($k_j$), there generically exists a solution in reciprocal space. This type of reasoning belongs to von Neumann and Wigner [204]. It was applied to band theory by Herring [205]. Such an accidental degeneracy is not required by symmetry but is topologically robust. To get rid of Weyl points, one needs to merge two of them with opposite chirality [206]. A Weyl point
corresponds to a singular source (or sink) of Berry flux. It is the reciprocal-space analog of a Dirac magnetic monopole, which Berry calls a diabolical point and which we call a Berry monopole. Weyl points can also be seen as the 3D analogs of the Dirac cones of graphene.

Nielsen and Ninomiya have shown that in a lattice realization, chiral (Weyl) fermions occur in pairs [207]. This is known as the fermion doubling theorem. It is related to the fact that a Weyl point carries a chirality of $N_3 = \pm 1$, which is equivalent to an elementary Berry monopole, and that the total Berry flux across the BZ should vanish. In order to have a Weyl point, one needs a non-vanishing Berry curvature, and therefore either inversion or time-reversal symmetry must be broken, which leads to two types of Weyl semi-metals: inversion symmetric or time-reversal symmetric.

An inversion-symmetric Weyl semi-metal necessarily breaks time-reversal symmetry: when there is a Weyl point at $k$ with chirality $N_3$, inversion symmetry implies that there is another one at $-k$ with chirality $-N_3$. Therefore the minimal number of Weyl points is two in that case as $N_3 + (-N_3) = 0$.

A time-reversal-symmetric Weyl semimetal necessarily breaks inversion symmetry and when there is a Weyl point at $k$ with chirality $N_3$, time-reversal symmetry guarantees that there is another one at $-k$ with the same chirality. In order for the total chirality in the BZ to vanish, one therefore needs a minimum of four Weyl points in that case.

Early on, Nielsen and Ninomiya have proposed a lattice realization of a Weyl semi-metal in order to study the associated chiral anomaly [207]. A nice and simple tight-binding model is discussed in [208]. These authors study a model for 3D spinless electrons on a lattice with two sites per unit cell, that breaks time-reversal symmetry and preserves inversion. In that way, they are able to find a situation in which there are only two Weyl points, which they study in detail including the resulting surface behavior. Indeed, a remarkable feature of Weyl semi-metals is that they host open Fermi arcs on their surface [209]. Open Fermi arcs are two-dimensional Fermi surfaces (i.e. lines) that are opened instead of closed. Actually, a Fermi arc connects to the projection of the Weyl points on the surface. The Fermi arcs are the 3D equivalent of the zero-energy edge mode present in 2D graphene nanoribbons [142]. To put it differently, some of the surfaces of a 3D Weyl semi-metal host a peculiar two-dimensional metal, which is chiral and has an opened Fermi line.

For a general review of theoretical and experimental aspects of Weyl semimetals, see [161].

8.3. 2D Chern insulator as topological texture

From the perspective of topological defects, there is also an interesting relation between the Weyl point in 3D and the Chern insulator in 2D. A band insulator has no Fermi surface, it is gapped at every point in the BZ, and therefore it can not be characterized as a topological defect. However, it may still contain some kind of twist. This is related to the notion of a topological texture [202, 210]. A topological texture is not a defect, as there is no vanishing of the ‘order parameter’, no defect, but it can nevertheless be characterized by a topological invariant. Topological textures are classified by relative homotopy groups [210]. They are best explained on the example of the 2D skyrmion, which is a simple topological texture. Let us consider the 2D reciprocal infinite plane $k = (k_x, k_y) \in \mathbb{R}^2$ (for the moment, we forget about the BZ torus $T^2$ relevant to the 2D Chern insulator and replace it by $\mathbb{R}^2$), in which at each point $k$ the vector field $d = (d_x, d_y, d_z)$ is defined. Imagine that at infinity $k \to \infty$, the vector $d$ points in the $z$ direction and that it smoothly evolves towards the center $k \to 0$, where it points in the $-z$ direction (see figure 21). Note that, in particular the vector $d$ never vanishes (one could consider that its norm is constant), so that there is no defect. The 2D plane with such a boundary condition at infinity $\mathbb{R}^2 + \{\infty\}$ can be compactified into a sphere $S^2$, that will play a role similar to that of the cage in the case of a defect (remember that a point-like defect in 3D is trapped in an $S^2$ cage). The relevant mapping that we now consider are from this compactified complete space $\mathbb{R}^2 + \{\infty\} \sim S^2$ to the Bloch sphere $S^2$. We already know that the relevant homotopy group is $\Pi_2(S^2) = \mathbb{Z}$. This means that topological textures in 2D are characterized by a wrapping number, just like topological defects in 3D. The elementary non-trivial topological texture is known as the 2D skyrmion and has a wrapping number $N_3 = \pm 1$. The notation here is again that of Volovik [45]: the tilda means that it is a topological index related to a texture and not to a defect. In other words, there is a deep relation between the 2D skyrmion $N_3 = \pm 1$ and the 3D hedgehog $N_3 = \pm 1$. The former can be seen as the stereographic projection of the latter (see figure 21 bottom). In the above discussion, the parameter space, which is the BZ torus $T^2$, was replaced by a sphere $S^2$ (compactified plane). As explained in [211] e.g. maps from $T^2$ to $S^2$ are actually topologically equivalent to maps from $S^2$ to $S^2$ (despite the fact that the sphere and the torus do not have the same genus). This means that topological textures also exist when the parameter space is a torus and are classified by a wrapping number.

In the context of band structures, the relation between the skyrmion and the hedgehog means that there is a relation between the 2D Chern insulator and the 3D Weyl semi-metal. In order to clearly expose this relation, we consider the phase transition between two phases of a 2D Chern insulators driven by a parameter

\[ W_{\text{Chern}} = \sum_{k} n_{\text{Chern}}(k) \cdot \chi(k) \]

where $n_{\text{Chern}}(k)$ is the Chern number at wavevector $k$, and $\chi(k)$ is the chirality of the Weyl point at $k$. The Chern number is a topological invariant that counts the number of crossings of the Fermi energy with the Fermi surface. The chirality of a Weyl point is defined as the sign of the chirality of the fermion field at the Weyl point.

\[ n_{\text{Chern}}(k) = \frac{1}{2\pi} \int d^2k \cdot \epsilon_{jkkl} \hat{v}_j(k) \cdot \hat{v}_l(k) \]

where $\epsilon_{jkkl}$ is the Levi-Civita symbol and $\hat{v}_j(k)$ is the gradient of the Fermi momentum $V_j(k)$. The chirality of a Weyl point is defined as the sign of the chirality of the fermion field at the Weyl point.

\[ \chi(k) = \frac{1}{2\pi} \int d^2k \cdot \epsilon_{jkkl} \hat{v}_j(k) \cdot \hat{v}_l(k) \]

The Chern number is a topological invariant that counts the number of crossings of the Fermi energy with the Fermi surface. The chirality of a Weyl point is defined as the sign of the chirality of the fermion field at the Weyl point. The Chern number is related to the topology of the Fermi surface, while the chirality is related to the chirality of the fermionic field at the Weyl point.
λ. One may think of the Haldane model and of tuning the inversion-breaking mass \( M \) (i.e. \( \lambda = M \) for example) so as to go from a phase with Chern number \( \tilde{N}_3^{(1)} \) to a phase with Chern number \( \tilde{N}_3^{(2)} \neq \tilde{N}_3^{(1)} \) (here we use the notation of Volovik for a topological texture, but the Chern number \( \tilde{N}_3 \) is really the same thing as what we called \( C \) before). The corresponding phase transition can be thought of as being a Weyl point in the space \((k_x, k_y, \lambda)\). The wrapping number \( N_3 \) of the corresponding Weyl point is related to the change in the Chern number as follows:

\[
\tilde{N}_3^{(2)} - \tilde{N}_3^{(1)} = N_3. \tag{176}
\]

For a detailed proof see [212]. Here, we recall our discussion in section 4.3.4: a Chern insulator is a 2D band insulator that has a non-zero number of Berry monopoles (or Weyl points) inside (i.e. enclosed by) its BZ.

In conclusion, gapped non-interacting fermionic systems can therefore be classified using ideas from topology. Two alternative viewpoints on topological insulators (or superconductors/superfluids) are:

1. twisted fiber bundles and
2. topological textures. The first viewpoint goes back to Thouless et al. [12], the second to Volovik [17].

### 8.4. 2D Dirac point as symmetry-protected topological defect

In contrast to the 3D Weyl semi-metal, in 2D, a contact point between two bands is unlikely (unstable) because it requires the vanishing of three functions \( d = (d_x, d_y, d_z) \) that only depend on two variables \( k = (k_x, k_y) \). The corresponding homotopy analysis involves a cage \( C = S^1 \), a target space \( V = S^2 \) and a homotopy group \( \Pi_1(S^2) = 0 \), which means that no topological invariant exists that protects a 2D contact point between two bands. Then, one may wonder about the case of graphene: why would Dirac points be stable? The contact points in graphene are actually not topologically stable but protected only as long as a certain symmetry is preserved. This symmetry is actually \( \mathcal{I} \mathcal{T} \), i.e. the product of inversion and time-reversal transformations, and ensures that the target space is restricted to a great circle \( V = S^1 \) of the Bloch sphere (instead of the whole Bloch sphere \( S^2 \)). This is known as a SPT defect. Indeed, in that case the relevant homotopy group is \( \Pi_1(S^1) = Z \) and the corresponding topological invariant is known as the winding number \( N_2(\mathcal{I} \mathcal{T}) \) using Volovik’s notation (the subscript 2 indicates the codimension and \( \mathcal{I} \mathcal{T} \) indicates that this topological invariant requires the existence of a symmetry for its existence). In graphene, the Dirac points have winding number \( N_2(\mathcal{I} \mathcal{T}) = \pm 1 \). Breaking the \( \mathcal{I} \mathcal{T} \) symmetry, the Dirac points can be immediately gapped, which shows that they are not stable against any perturbation (see e.g. boron nitride that breaks inversion symmetry or the Haldane model that breaks time-reversal symmetry). However, if this symmetry is maintained, the Dirac points can only disappear via a merging (Lifshitz) transition [213]. This topological invariant \( N_2(\mathcal{I} \mathcal{T}) \) is known as the winding number or the chirality \( \chi = \text{sign} \text{det}(\nu_3) \). Symmetries are actually able to protect degeneracies between more than two bands (i.e. pseudo-spin 1/2). For a discussion on how site-permutation symmetries can be used to obtain the equivalent of pseudo-spin 1 or 2 fermionic quasiparticles, see e.g. [214].

### 8.5. Classification of topological metals

The general idea of Volovik is to classify Fermi surfaces as topological defects [45, 206]. Using elementary ideas from homotopy groups, he realized that in 3D, the Fermi surface \( d' = 2 \) is stable (and characterized by a winding number), the Fermi line \( d' = 1 \) is unstable and the Fermi point \( d' = 0 \) is stable (also with an integer invariant). The latter is now known as a 3D Weyl semi-metal. In 2D, he found that the Fermi line is stable (with integer invariant) and the Fermi point is unstable (unless protected by a symmetry as in graphene). And in 1D, the Fermi point being the natural ’Fermi surface’ is stable (also with an integer invariant).

As in the case of gapped systems (topological insulators or superconductors), one may generalize the above ideas and classify all types of Fermi surfaces as topological defects. Important concepts are space dimension and the nature of the Fermi surface (does it involve complex (Dirac) fermions or real (Majorana) fermions?). As in the ten-fold periodic table, the classes are of three types (0, \( Z_2 \) or \( Z \)) and there is a form of regularity known as Bott periodicity and inherent to K-theory. This issue is beyond the scope of the present review. The interested reader will find more information in [43, 44].

### 9. Conclusion

To conclude this review, we would like to highlight some take-home messages:

- In band theory, the physical properties are not only determined by the energy level properties, but also by the Bloch wave functions. For instance the electrical polarization of crystals and the QHE cannot be understood solely from the energy bands. In the central section of this review (section 4), we explain how
the geometrical concepts (Berry connection and curvature, quantum metrics) and the topological invariants (Chern number) are useful to build a complete description of crystal band structures, going beyond the energy level spectrum. Then these latter concepts are applied to several systems: one-dimensional lattices (section 5), electrons on honeycomb lattices (section 6), and finally 3D Weyl/Dirac semimetals (section 8).

• Interesting physics emerges when bands are not independent but are rather coupled by virtual transitions due to external fields (e.g. electric and magnetic fields). These virtual transitions lead to geometrical effects locally in reciprocal $k$-space and to topological effects globally in reciprocal space (section 4). Topologically trivial insulators (e.g. boron nitride) may still contain interesting geometrical effects. For example, Dirac insulators host zero-modes trapped on topological defects of the Dirac mass (see the Jackiw–Rebbi mechanism, section 5.4).

• Berry phase concepts (including Berry connection and curvature) are not mandatory to describe the inter-band effects. In principle, if one could solve completely the energy spectrum in the presence of external fields all the physical properties could be calculated without resorting to Berry phase concepts. The Berry connection appears when one projects onto a single isolated band, or an isolated subset of bands. This is already clear in the context of the 0D TLS (section 3) and pertains for $D$-dimensional lattice systems (section 4). The alternative when studying a crystal in an external field is therefore: either use the zero-field energy bands and Bloch states (and therefore Berry phase effects); or compute the energy spectrum in the presence of external fields.

• In practice, it is very important to distinguish a periodic Bloch Hamiltonian and the canonical Bloch Hamiltonian in the case of a crystal structure with several atoms per unit cell (i.e. lattice with a basis). A periodic Hamiltonian contains less information than the canonical Hamiltonian. The usual formula for the Berry connection or curvature, Zak phases are typically written in terms of the canonical Bloch Hamiltonian and the corresponding cell-periodic Bloch states.

• A definition for a topological insulator consists in a bulk insulator with a quantized bulk response, and a topologically protected metal at its boundary. Within this picture, there is only one way to be a trivial insulator (section 7.1).

• An alternative paradigm consists in defining as topological an insulator that cannot be continuously deformed (while keeping a protective symmetry) into an atomic insulator characterized by localized symmetric Wannier functions. In this framework, there can be several trivial atomic limits (section 7.2).

• It is good to keep in mind a distinction between topological insulators (discussed in the present review) and the notion of TO (not discussed). See appendix for a short comparison between the two notions.

Data availability statement

No new data were created or analyzed in this study.

Acknowledgments

We acknowledge useful discussions with János Asbóth, Jens Bardarson, Hélène Bouchiat, Alexandre Buzdin, David Carpentier, Frédéric Combes, Pierre Delplace, Balasz Dora, Benoît Douçot, Clément Dutreix, Pierre Gosselin, Adolfo Grushin, Rony Ilan, Pavel Kalugin, Lih-King Lim, Andrej Mesaros, Roderich Moessner, Hervé Mohrbach, Joel Moore, Rémy Mosseri, Louis Nouri, Hugo Perrin, Arnaud Raoux, Pascal Simon, Doru Sticlet, André Thiaville, Max Trescher, Pierre Vallet, Julien Vidal and Ashvin Vishwanath. Special thanks to the gang of four in Orsay—Mark Goerbig, Gilles Montambaux and Frédéric Piéchon (like the musketeers)—for the enlightening discussions over the years since the first experimental papers on graphene that sparked our interest in these matters. We also thank Christophe Brun for organizing lectures at INSP in Jussieu that were a strong motivation in putting all this material together. Jérôme Cayssol acknowledges support from the Quantum Matter Network at Bordeaux University under project TaQuaMaUC.

Appendix. Topological insulator versus topological order

In this appendix and in order to recap what we learned on topological insulators, we wish to compare and distinguish them as clearly as possible from another type of gapped systems featuring some topological property and called topological order. The following discussion is essentially summarized in table 2. A good reference for this distinction is the commentary written by Fisher [215]. Topological insulators are discussed in the books by Bernevig and Hugues [18], Asbóth et al [60] and Vanderbilt [54], whereas topological order (TO) is discussed in detail in the book by Wen [167] and in the review [168].
Table 2. Symmetry-protected topological phases (topological insulators) versus intrinsic topological order.

|                      | Topological insulators (SPT)                        | Topological order (TO)                         |
|----------------------|------------------------------------------------------|------------------------------------------------|
| Full name            | Symmetry-protected topological phases                | Intrinsic topological order                    |
| Space dimension      | Any D                                                | Mainly 2D                                       |
| Gapped bulk          | Yes                                                  | Yes                                             |
| Any small perturbation | Not robust                                           | Robust                                          |
| Effective description | Topological band theory                             | Topological quantum field theory               |
| Interactions         | Not needed                                           | Required                                        |
| Ground-state         | Non-degenerate                                       | Degenerate on manifolds with non-trivial topology |
| Bulk excitations     | Non-exotic                                           | Fractionalized, anyons, topological, created in pairs related by a string |
| Robust gapless edge states | Always                                              | Not always (chiral versus achiral TO)           |
| Ground-state entanglement | Short-range                                      | Long-range                                      |
| Related to           | Topological textures                                 | Lattice gauge theories, tensor categories, CFT |
| Historical example   | Integer quantum Hall effect (IQHE)                  | Fractional quantum Hall effect                  |
| Further examples     | Chern insulators, $\mathbb{Z}_2$ TI in 2D and 3D    | Quantum spin liquids                            |
| Models               | Haldane, Kane–Mele, Bernevig-Hughes-Zhang, Volovik $p_x + ip_y$ superfluid, Majorana chain | Toric code (Kitaev), string nets (Levin-Wen), Kitaev honeycomb |
| Historical roots     | Thouless–Kohmoto–Nightingale–den Nijs 1982          | Wen's topological order 1990                   |

We restrict the discussion to two-dimensional gapped systems. On the one hand, topological insulators are essentially non-interacting fermionic band insulators. Their topological nature comes from the fact that they can not be adiabatically—i.e. without closing the bulk gap—deformed into an atomic insulator. This obstruction is typically characterized by a topological invariant such as a Chern number. Usually, the obstruction in doing so only occurs provided a symmetry condition is imposed. These are therefore also known as symmetry-protected topological (SPT) insulators. A good example is the Kane–Mele model for a quantum spin Hall insulator (QSHI), which is protected by time-reversal symmetry. As a consequence of this obstruction, a boundary with the vacuum (that behaves as a trivial insulator) necessarily hosts gapless edge modes, that have some form of robustness. In the case of the QSHI, the edge modes are spin-momentum locked and known as helical edge states, their direction of motion being tied to their spin projection. Therefore an impurity or a defect that does not act on spin can not backscatter. The groundstate of a SPT insulator is unique and is a Slater determinant. It has only short-range quantum entanglement. Bulk excitations are similar to that in a trivial band insulator: they are electronic quasiparticles. Topological insulators in this sense are a refinement of band insulators.

On the other hand, systems featuring TO are typically strongly interacting. Examples are fractional quantum Hall (FQH) states and quantum spin liquids (QSL). In the FQH effect, microscopics is defined in terms of interacting electrons in a perpendicular magnetic field. In the QSL, the microscopics typically involves localized spins with exchange interaction and some form of frustration that prevents spontaneous symmetry breaking and long-range order. The topological nature of TO phases is revealed by their ground-state degeneracy depending on the genus of the space manifold: it is different for the system placed on a sphere or on a torus, for example. This property is intrinsic (the topological degeneracy is robust to any perturbation which is sufficiently small) and does not rely on the protection by a specific symmetry. The ground-state features long-range quantum entanglement despite the fact that correlation functions decay exponentially on a length scale set by the bulk gap. Furthermore, some of the excitations are exotic: they are topological excitations, created in pairs, related by an unobservable string and having fractionalized quantum numbers and exchange statistics. The latter properties mean that they are anyons. A defining property of TO is the presence of such fractionalized excitations.

In this dichotomy between SPT phases and TO phases, there are some cases which are not so clear cut. For example, the IQHE seems to be the historical example of a topological insulator but it exists in the absence of any protecting symmetry (it is intrinsically topological). However its excitations are not fractionalized. Also, superconductors are usually described as resulting from the spontaneous symmetry breaking of a gauge symmetry, despite the fact that a gauge symmetry can not be broken. Actually, they have been shown to be topologically ordered [216]. Furthermore, topological superconductors such as chiral $p_x + ip_y$ superconductors, when treated at the mean-field level so that they effectively appear as being non-interacting, belong to the $D$-class of the ten-fold periodic table and are characterized by a Chern
invariant [41, 42]. As such they are a form of SPT. But they are also known to host fractionalized excitations trapped in vortex cores and known as Majorana zero modes (or Majorana bound states). The latter behave as non-Abelian (Ising) anyons.

**ORCID iDs**

J Cayssol @ https://orcid.org/0000-0002-2021-9080  
J N Fuchs @ https://orcid.org/0000-0003-1950-8261

**References**

[1] Hoddeson L, Baym G and Eckert M 1987 *Rev. Mod. Phys.* 59 287  
[2] Wilson A 1931 *Proc. R. Soc. A* 133 458  
[3] Bloch F 1928 *Z. Phys.* 52 555  
[4] Peierls R E 1929 *Z. Phys.* 53 255  
[5] Jones H and Zener C 1934 *Proc. R. Soc. A* 144 101  
[6] Ashcroft N W and Mermin N D 1976 *Solid State Physics* (Philadelphia, PA: Saunders College Publishing)  
[7] Blount E 1962 *Solid State Phys.* 13 305  
[8] Karplus R and Luttinger J M 1954 *Phys. Rev.* 95 1145  
[9] Kohn W 1959 *Phys. Rev.* 115 1460  
[10] von Klitzing K, Dorda G and Pepper M 1980 *Phys. Rev. Lett.* 45 494  
[11] Laughlin R B 1981 *Phys. Rev. B* 23 5632  
[12] Thouless D J, Kohmoto M, Nightingale M P and den Nijs M 1982 *Phys. Rev. Lett.* 49 405  
[13] Hofstadter D R 1976 *Phys. Rev. B* 14 2239  
[14] Berry M 1984 *Proc. R. Soc. A* 392 45  
[15] Simon B 1983 *Phys. Rev. Lett.* 51 2167  
[16] Haldane F D M 1988 *Phys. Rev. Lett.* 61 2015  
[17] Volodin G E 1988 *Sov. Phys. JETP* 67 1804  
[18] Bernevig B A 2013 *Topological Insulators and Topological Superconductors* ed T L Hughes (Cambridge: Cambridge University Press) ([https://doi.org/10.1515/9781400846733](https://doi.org/10.1515/9781400846733))  
[19] Zak J 1989 *Phys. Rev. Lett.* 62 2747  
[20] King-Smith R D and Vanderbilt D 1993 *Phys. Rev. B* 47 1651(R)  
[21] Vanderbilt D and King-Smith R D 1993 *Phys. Rev. B* 48 4442  
[22] Resta R 1994 *Rev. Mod. Phys.* 66 899  
[23] Resta R and Vanderbilt D 2007 *Physics of Ferroelectrics* (Berlin: Springer)  
[24] Xiao D, Chang M-C and Niu Q 2010 *Rev. Mod. Phys.* 82 1959  
[25] Thonhauser T 2011 *Int. J. Mod. Phys.* B 25 1429  
[26] Chang M-C and Niu Q 1996 *Phys. Rev. B* 53 7010  
[27] Sundaram G and Niu Q 1999 *Phys. Rev. B* 59 14915  
[28] Kane C L and Mele E J 2005 *Phys. Rev. Lett.* 95 226801  
[29] Kane C L and Mele E J 2005 *Phys. Rev. Lett.* 95 146802  
[30] Bernevig B, Hughes T L and Zhang S-C 2006 *Science* 314 1757  
[31] König M, Wiedmann S, Brüne C, Roth A, Buhmann H, Molenkamp L W, Qi X L and Zhang S-C 2007 *Science* 318 766  
[32] Fu L, Kane C L and Mele E J 2007 *Phys. Rev. Lett.* 98 106801  
[33] Moore J and Balents L 2007 *Phys. Rev. B* 75 121306(R)  
[34] Roy R 2009 *Phys. Rev. B* 79 195322  
[35] Hasan M Z and Kane C L 2010 *Rev. Mod. Phys.* 82 3045  
[36] Novoselov K, Geim A, Morozov S V, Jiang D, Katsnelson M, Grigorieva S, Dubonos I V and Firsov A 2005 *Nature* 438 197  
[37] Zhang Y, Tan Y-W, Stormer H L and Kim P 2005 *Nature* 438 201  
[38] Wallace P R 1947 *Phys. Rev.* 71 622  
[39] DiVincenzo D P and Mele E J 1984 *Phys. Rev. B* 29 1685  
[40] Berry M 2010 *Nat. Phys.* 6 148  
[41] Schynder A P, Ryu S, Furusaki A and Ludwig A W W 2008 *Phys. Rev. B* 78 195125  
[42] Kitaev A 2009 *AIP Conf. Proc.* 1134 22  
[43] Horava P 2005 *Phys. Rev. Lett.* 95 016405  
[44] Zhao Y X and Wang Z D 2013 *Phys. Rev. Lett.* 110 240404  
[45] Volodin G E 2003 *The Universe in a Helium Droplet* (Oxford : Oxford University Press)  
[46] Dirac P A M 1929 *Proc. R. Soc. A* 117 610  
[47] Cayssol J 2013 *C. R. Phys.* 14 760  
[48] Shen S-Q 2017 *Topological Insulators: Dirac Equation in Condensed Matter* (Berlin: Springer) ([https://doi.org/10.1007/978-3-662-32858-9](https://doi.org/10.1007/978-3-662-32858-9))  
[49] Duplanter B, Rivasseau V and Fuchs J-N (eds) 2017 *Dirac Matter (Dirac Matter vol 71)* (Basel: Birkhäuser)  
[50] Dirac P A M 1930 *Proc. R. Soc. A* 133 60  
[51] Wu T and Yang C N 1975 *Phys. Rev. D.* 12 3845  
[52] Resta R 2000 *J. Phys.: Condens. Matter* 12 R107  
[53] Resta R 2011 *Eur. Phys. J. B* 79 121  
[54] Vanderbilt D 2018 *Berry Phases in Electronic Structure Theory* (Cambridge: Cambridge University Press) ([https://doi.org/10.1017/9781107166620](https://doi.org/10.1017/9781107166620))  
[55] Qi X-L and Zhang S-C 2011 *Rev. Mod. Phys.* 83 1057  
[56] König M, Buhmann H, Molenkamp L W, Hughes T, Liu C-X, Qi X-L and Zhang S-C 2008 *J. Phys. Soc. Japan* 77 031007  
[57] Fruchart M and Carpentier D 2013 *C. R. Phys.* 14 779
If the crystal has inversion symmetry, the position origin is often chosen on an inversion center, in which case the Zak phase is quantized: it is either 0 or $\pi$. Actually in that case the wave function is no longer single-valued, which is a problem.

One may think that it would be a good idea to have a more symmetric phase by having $e^{\pm \phi/2}$. In that case the Zak phase is quantized: it is either 0 or $\pi$. Actually in that case the wave function is no longer single-valued, which is a problem.

References:

[58] Witten E 2016 Riv. Nuovo Cimento 7 313
[59] Shankar R 2018 (arXiv:1804.06471)
[60] Asboth J K, Oroszlány I and Pályi A 2016 A Short Course on Topological Insulators (A Short Course on Topological Insulators) (Berlin: Springer) (https://doi.org/10.1007/978-3-319-25607-8)
[61] Cooper N R, Dalibard J and Spielman I B 2019 Rev. Mod. Phys. 91 015005
[62] Ozawa T et al 2019 Rev. Mod. Phys. 91 015006
[63] Ryder L H 1985 Quantum Field Theory (Cambridge; Cambridge University Press)
[64] Mathematically this is expressed in the fact that the second cohomology group of the sphere is non trivial: $H^2(S^3) = \mathbb{Z}$
[65] Thouless D 1989 Topological Quantum Numbers in Nonrelativistic (Singapore: World Scientific)
[66] Ryder L H 1980 J. Phys. A: Math. Gen. 13 437
[67] Minami M 1979 Prog. Theor. Phys. 62 1128
[68] Wilczek F and Shapere A 1989 Geometric Phases in Physics (Singapore: World Scientific)
[69] One may think that it would be a good idea to have a more symmetric phase by having $e^{\pm \phi/2}$. Actually in that case the wave function is no longer single-valued, which is a problem.
[70] Schroer M D, Kolodrubetz M H, Kindel W F, Sandberg M, Gao J, Vissers M R, Pappas D P, Polkovnikov A and Lehnert K W 2014 Phys. Rev. Lett. 113 056402
[71] Roushan P et al 2014 Nature 515 241
[72] Urbanikie H K 2003 J. Geom. Phys. 46 125
[73] Nakahara M 2003 Geometry, Topology and Physics 2nd edn (Bristol: IOP Publishing)
[74] Berry M 1989 Geometric Phases in Physics ed F Wilczek and A Shapere (Singapore: World Scientific)
[75] Thouless D J 1983 Phys. Rev. B 27 6083
[76] Avron J E, Seiler R and Simon B 1983 Phys. Rev. Lett. 51 51
[77] Provost J P and Vaille G 1980 Commun. Math. Phys. 76 289
[78] Piéchon F, Raoux A, Fuchs J-N and Montambaux G 2016 Phys. Rev. B 94 134423
[79] Liang L, Vanhala T I, Pootta S, Tero T, Harju A and Törmä P 2017 Phys. Rev. B 95 024515
[80] Marzari N, Mostofi A A, Yates J R, Souza I and Vanderbilt D 2012 Rev. Mod. Phys. 84 1419
[81] Fläschner N, Rem B S, Tarnowski M, Vogel D, Lüthmann D-S, Sengstock K and Weitenberg C 2016 Science 352 1091
[82] Gianfrate A et al 2020 Nature 578 381
[83] Lim L-K, Fuchs J N and Montambaux G 2015 Phys. Rev. A 92 063627
[84] Bena C and Montambaux G 2009 New J. Phys. 11 095003
[85] Fuchs J-N, Piéchon F, Goerbig M O and Montambaux G 2011 Eur. Phys. J. B 77 351
[86] Fruchart D C M and Gawedzki K 2014 Europhys. Lett. 106 60002
[87] Raoux A, Morigi M, Fuchs J-N, Piéchon F and Montambaux G 2014 Phys. Rev. Lett. 112 026402
[88] Combes F, Treuher T, Piéchon F and Fuchs J-N 2016 Phys. Rev. B 94 135109
[89] Aharony Y and Bohn D 1959 Phys. Rev. 115 485
[90] If the crystal has inversion symmetry, the position origin is often chosen on an inversion center, in which case the Zak phase is quantized: it is either 0 or $\pi$.
[91] Wannier G H 1937 Phys. Rev. 52 191
[92] Kivelson S 1982 Phys. Rev. B 26 4269
[93] Kohn W 1959a Phys. Rev. 115 809
[94] Stiarnati G 1978 Phys. Rev. B 18 4104
[95] Brouder C, Panati G, Calandra M, Mourouguane C and Marzari N 2007 Phys. Rev. Lett. 98 046402
[96] Thouless D J 1984 J. Phys. C: Solid State Phys. 17 L325
[97] Thouless D and Vanderbilt D 2006 Phys. Rev. B 74 235111
[98] Vinyuela O, Rivas A and Martin-Delgado M A 2014 Phys. Rev. Lett. 113 076408
[99] Peierls R 1955 Quantum Theory of Solids (Oxford: Oxford University Press)
[100] Peierls R E 1933 Z. Phys. 80 763
[101] Gosselin P, Ménas F, Bérard A and Mohrbach H 2006 Europhys. Lett. 76 651
[102] Gao Y, Yang S A and Niu Q 2014 Phys. Rev. Lett. 112 116601
[103] Xiao D, Shi J and Niu Q 2005 Phys. Rev. Lett. 95 137204
[104] Halperin B I 1982 Phys. Rev. B 25 2185
[105] Oh S 2013 Science 340 153
[106] Nagaosa N, Sinova J, Onoda S, MacDonald A H and Ong N P 2010 Rev. Mod. Phys. 82 1539
[107] Sinova J, Valenzuela S O, Wunderlich J, Back C H and Jungwirth T 2015 Rev. Mod. Phys. 87 1213
[108] Moore J E and Orenstein J 2010 Phys. Rev. Lett. 105 026805
[109] Sodemann I and Fu L 2015 Phys. Rev. Lett. 115 216806
[110] Morimoto T, Zhong S, Orenstein J and Moore J E 2016 Phys. Rev. B 94 245121
[111] Parker D E, Morimoto T, Orenstein J and Moore J E 2019 Phys. Rev. B 99 045121
[112] Nandy S and Sodemann I 2019 Phys. Rev. B 100 195117
[113] Matsuyishin O and Sodemann I 2019 Phys. Rev. Lett. 123 246602
[114] Ma Q et al 2019 Nature 565 337
[115] Kang K, Li T, Soh E, Shan J and Mak K F 2019 Nat. Mater. 18 324
[116] Shvetsov O O, Eskin V D, Timonina A V, Kolesnikov N N and Deviatov E V 2019 IETT Lett. 109 715
[117] de Juan F, Grushin A G, Morimoto T and Moore J E 2017 Nat. Commun. 8 15995
[118] Wilczek F and Zee A 1984 Phys. Rev. Lett. 52 2111
[119] Alexandradinata A, Dai X and Bernevig B A 2014 Phys. Rev. B 89 155114
[120] Banerjee W A, Bernevig B A and Hughes T L 2017 Phys. Rev. B 96 245115
[121] Su W P, Schrieffer J R and Heeger A J 1979 Phys. Rev. Lett. 42 1698
[122] Heeger A J, Kivelson S, Schrieffer J R and Su W P 1988 Rev. Mod. Phys. 60 781
[123] Rice M J and Mele E J 1982 Phys. Rev. Lett. 49 1455
[124] Onoda S, Murakami S and Nagaosa N 2004 Phys. Rev. Lett. 93 167602
[125] Jackiw R and Rebbi C 1976 Phys. Rev. D 13 3398
[126] Goldstone J and Wilczek F 1981 Phys. Rev. Lett. 47 986
[127] Kitaev A Y 2001 Phys.-Usp. 44 131
[128] Qi X-L, Hughes T L and Zhang S-C 2008 Phys. Rev. B 78 195424
[129] Chiu C-K, Teo J C Y, Schnyder A P and Ryu S 2016 Rev. Mod. Phys. 88 035005
[130] Ryu S and Hatsugai Y 2002 Phys. Rev. Lett. 89 077002
[131] Delplace P, Ullmo D and Montambaux G 2011 Phys. Rev. B 84 195452
[132] Atala M, Aidsedlerburger M, Barreiro J T, Abanin D, Kitagawa T, Demler E and Bloch I 2013 Nat. Phys. 9 795
[133] The polarization ‘quantum’ has nothing to do with quantum mechanics
[134] Co W P, Schrieffer J R and Heeger A J 1980 Phys. Rev. B 22 2099
[135] Jackiw R 2012 Phys. Scr. 2012 014005
[136] Fu L I 2011 Phys. Rev. Lett. 106 106802
[137] Semenoff G W, Semenoff V and Zhou F 2008 Phys. Rev. Lett. 101 087204
[138] Spaldin N A 2012 J. Solid State Chem. 195 2
[139] Chen K T and Lee P A 2011 Phys. Rev. B 84 125119
[140] Bradlyn B, Elcoro L, Cano J, Vergniory M G, Wang Z, Parkin S S P, Bernevig B A and Neupert T 2018 Phys. Rev. B 97 085125
[141] Qi X-L, Hughes T L and Zhang S-C 2008 Phys. Rev. B 78 195424
[142] Goerbig M O 2011 Rev. Mod. Phys. 83 1193
[143] Dirac P A M 1930 Proc. R. Soc. A 126 360
[144] Novoselov K, Geim A, Morozov S V, Jiang D, Zhang Y, Dubonos S, Grigorieva I and Firsov A 2004 Science 306 666
[145] Wang G, Chernikov A, Glazov M M, Heinz T F, Marie X, Amand T and Urbaszek B 2018 Rev. Mod. Phys. 90 021001
[146] Qian X, Liu J, Fu L and Kane C L 2006 Phys. Rev. Lett. 97 085402
[147] Fu L and Kane C L 2008 Phys. Rev. B 78 195424
[148] The polarization ‘quantum’ has nothing to do with quantum mechanics
[149] Fei Z
[150] Zhang T, Jiang Y, Song Z, Huang H, He Y, Fang Z, Weng H and Fang C 2019 Nature 566 475
[151] Vergniory M G, Elcoro L, Felser C, Regnault N, Bernevig B A and Wang Z 2019 Nature 566 480
[152] Tang F, Po H C, Vishwanath A and Wan X 2019 Nature 566 486
[153] Dirac P A M 1930 Proc. R. Soc. A 126 360
[154] Nagaosa N, Rice J R and Fujita T 1994 Phys. Rev. B 49 13161
[155] The polarization ‘quantum’ has nothing to do with quantum mechanics
[156] Zhang T, Jiang Y, Song Z, Huang H, He Y, Fang Z, Weng H and Fang C 2019 Nature 566 475
[157] The polarization ‘quantum’ has nothing to do with quantum mechanics
[158] Tang F, Po H C, Vishwanath A and Wan X 2019 Nature 566 486
[159] Dirac P A M 1930 Proc. R. Soc. A 126 360
[151] Dirac P A M 1930 Proc. R. Soc. A 126 360
[197] Lifshitz I 1960 Sov. Phys. JETP 11 1130
[198] Mermin N D 1979 Rev. Mod. Phys. 51 591
[199] Sethna J P 1992 1991 Lectures in Complex Systems (Santa Fe Institute Studies in the Sciences of Complexity, Proc. vol XV) ed I. Nagel and D Stein (Reading, MA: Addison-Wesley)
[200] Toulouse G and Kléman M 1976 J. Phys. Lett. 37 149
[201] Hoffmann M, Zimmermann B, Müller G P, Schürhoff D, Kiselev N S, Melcher C and Blügel S 2017 Nat. Commun. 8 308
[202] Chaikin P M and Lubensky T C 1995 Principles of Condensed Matter Physics (Cambridge: Cambridge University Press)
[203] Weyl H 1929 Z. Phys. 56 330
[204] von Neumann J and Wigner E 1929 Phys. Z. 30 467
[205] Herring C 1937 Phys. Rev. 52 365
[206] Volovik G E 2007 Quantum Phase Transitions From Topology in Momentum Space (Lecture Notes in Physics vol 718) (Berlin: Springer) p 31
[207] Nielsen H B and Ninomiya M 1983 Phys. Lett. B 130 389
[208] Delplace P, Li J and Carpentier D 2012 Europhys. Lett. 97 67004
[209] Wan X, Turner A M, Vishwanath A and Savrasov S Y 2011 Phys. Rev. B 83 205101
[210] Michel L 1980 Rev. Mod. Phys. 52 617
[211] Moore J, Ran Y and Wen X-G 2008 Phys. Rev. Lett. 101 186805
[212] Belissard J 1995 (arXiv:cond-mat/9504030)
[213] Montambaux G, Piéchon F, Fuchs J-N and Goerbig M O 2009 Phys. Rev. B 80 153412
[214] Crasto de Lima F and Ferreira G J 2020 Phys. Rev. B 101 041107(R)
[215] Fisher M P A 2013 J. Club Condens. Matter Phys. (https://doi.org/https://doi.org/10.36471/JCCM_February_2013_01)
[216] Hansson T H, Oganesyan V and Sondhi S L 2004 Ann. Phys. 313 497
[217] Kohmoto M, Halperin B I and Wu Y-S 1992 Phys. Rev. B 45 15488