

# Topological electronic bands in crystalline solids

A. T. Boothroyd<sup>a,b</sup>

<sup>a</sup>Department of Physics, Oxford University, Clarendon Laboratory, Oxford, UK

<sup>b</sup>LINXS Institute of Advanced Neutron and X-ray Science, Lund, Sweden

## ARTICLE HISTORY

Compiled November 5, 2023

## ABSTRACT

Topology is now securely established as a means to explore and classify electronic states in crystalline solids. This review provides a gentle but firm introduction to topological electronic band structure suitable for new researchers in the field. I begin by outlining the relevant concepts from topology, then give a summary of the theory of non-interacting electrons in periodic potentials. Next, I explain the concepts of the Berry phase and Berry curvature, and derive key formulae. The remainder of the article deals with how these ideas are applied to classify crystalline solids according to the topology of the electronic states, and the implications for observable properties. Among the topics covered are the role of symmetry in determining band degeneracies in momentum space, the Chern number and  $\mathbb{Z}_2$  topological invariants, surface electronic states, two- and three-dimensional topological insulators, and Weyl and Dirac semimetals

## KEYWORDS

Topological band structure; Berry phase; Berry curvature; topological insulator; Weyl semimetal; Dirac semimetal; Fermi arcs.

## 1. Introduction

Over the past few decades, topology has become an increasingly valuable tool in condensed matter physics, providing a radically different way of thinking about materials and enabling the prediction and discovery of exotic states with intriguing physical properties. The importance of topology in condensed matter was recognised by the award of the 2016 Nobel Prize in Physics to Thouless, Haldane and Kosterlitz.

Amongst topological themes of current interest in the area of quantum materials are, in no particular order, topological defects and phase transitions [1,2], topological insulators [3], graphene [4], magnetic skyrmions [5], and Majorana fermions for quantum computation [6]. The use of topology is enriching other fields too, including photonics [7], polymer statistics [8], and biological matter [9]. Topological systems are also appealing from the point of view of practical applications, due in large part to a special property known as *topological protection*, which can enhance the stability of a state.

In short, concepts from topology have become widespread in condensed matter and are making a significant impact. My focus here is on one particular area within this new field, namely the topology of electronic bands in crystalline solids and associated

observable phenomena. The topics I shall review came to prominence in the early years of the twenty-first century, first with the demonstration of the extraordinary electronic properties of graphene [10], and very shortly after with the prediction [11,12] and subsequent experimental verification [13] of the quantum spin Hall insulator (a two-dimensional version of a topological insulator). These ideas were soon extended into three dimensions, igniting the field of topological electronic bands in solids and leading to a range of predictions and discoveries of novel effects in topological insulators, semimetals, metals and superconductors.

Several important themes cut across this new realm of electronic phenomena:

- (1) The *Berry phase*. This is a quantum-mechanical phase picked up by systems when they are slowly deformed. Berry phases cause interference effects due to the wave-like nature of particles.
- (2) The *Berry curvature*. This is an emergent gauge field which acts like a magnetic field in momentum space and gives rise to unconventional transport phenomena, such as the anomalous Hall and Nernst effects.
- (3) *Band touching points*. These are points in momentum space where two or more electronic bands become degenerate. They are special because of the associated Berry phase effects, and also because if the chemical potential is located near such a point the electronic excitations can be exotic.
- (4) *Electronic topology*. The materials we are concerned with here contain topologically non-trivial electronic states. In other words, the geometry of electrons in the solid is distinct from the geometry of a free electron. Topologically inequivalent states cannot be smoothly deformed into one another, and are often separated by an energy barrier, making them stable against perturbations.

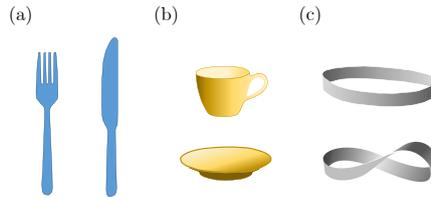
Accounts of the above-mentioned concepts are plentiful but can be somewhat technical. My aim here is to provide a mathematically rigorous but pedagogical introduction, and to show as simply as possible how these ideas translate into observable phenomena. Some proficiency in quantum mechanics and solid-state physics is required, but not beyond what is taught in a standard undergraduate course. For readers who aspire to know more, there are some excellent books and technical review articles which go into much greater detail, of which Refs. [14–27] are a good selection.

## 2. Basic ideas of topology

I begin with a brief introduction to a few important concepts in topology that are relevant to the theme of this article.

### 2.1. Topological invariants

Topology is first and foremost a branch of mathematics used to describe geometrical properties that remain invariant when an object is continuously deformed. Two objects are considered to be topologically equivalent, or *homeomorphic*, if one object can be transformed into the other by bending, stretching or twisting, but not cutting or sticking together. For example, a knife and fork are homeomorphic, Figure 1(a), whereas a cup and saucer are not, Figure 1(b). Similarly, to convert a simple cylindrical band into a Möbius strip, Figure 1(c), you need to cut the band, apply a half-twist, and then reattach the ends, so these shapes are not topologically equivalent.



**Figure 1.** Examples of pairs of objects which are (a) topologically equivalent, and (b)–(c) not topologically equivalent.

In topology, objects are classified by means of *topological invariants*, which quantify some generic property which they have in common, e.g. the number of holes or twists. Topological invariants have discrete values, usually normalised to be integers. A well-known example is Euler’s characteristic for geometric solids, defined as  $\chi = F + V - E$ , where  $F$ ,  $V$  and  $E$  are the number of faces, vertices and edges. All convex polyhedra have  $\chi = 2$ , and are topologically equivalent to (and hence continuously deformable into) a sphere. Other examples of topological invariants are, winding numbers (of closed curves) and the Kauffman bracket polynomial (for knots).

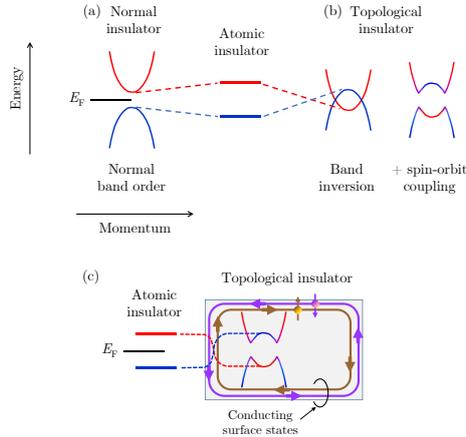
Topological equivalence in quantum systems concerns whether the Hamiltonian of one system can be continuously deformed into that of another while preserving a topological invariant (also known as a topological quantum number). Here, ‘deforming’ means slowly changing one or more of the parameters of the Hamiltonian. The topology of electronic band states in crystals is characterised by a topological invariant called the *Chern number*, which I shall define in Section 4.9.

## 2.2. Topological protection

We know from experience that it is often difficult to convert topologically inequivalent states into one another. A knot, for example, cannot easily be removed from a piece of string simply by shaking it around. It requires a series of coordinated steps that are highly unlikely to occur at random. It is easy to appreciate, therefore, that a state which is topologically distinct from its surroundings can be highly immune to perturbations, making it especially robust. States which are topologically protected in this way can be very useful in devices. For example, topologically protected spin-polarised surface conducting states have potential applications in high-speed electronics and spintronics, and an entirely new field of topological quantum computation based on the manipulation of Majorana edge modes is envisaged.

## 2.3. Topological order

Topologically inequivalent states are regarded as being differently ordered, but not in a conventional sense. Topological electronic order is characterised by a Chern number, rather than by a broken symmetry such as loss of rotation, inversion or time-reversal symmetry. When topological transitions take place between states with different Chern numbers something dramatic must happen — something like a twisting of the electronic wave function. One consequence is the existence of edge modes at the boundary



**Figure 2.** Illustration of band inversion and edge currents in a topological insulator. For simplicity, two atomic levels are shown, the lower level being occupied and the upper level empty, so the Fermi level ( $E_F$ ) is in the gap. (a) The band gap of a normal insulator is smoothly connected to the gap between the atomic levels. (b) In a topological insulator there is a gap between the valence and conduction bands, and for some range of momenta the bands are inverted relative to the atomic levels. (c) Gap closure and level crossing occurs at the interface between the topological insulator and vacuum, resulting in a conducting surface. The surface currents are spin-polarised. The up-spin and down-spin currents flow in opposite directions around the boundary.

between different topological phases. An example is the conducting states that exist on the surface of a topological insulator.

#### 2.4. What is a topological insulator?

The classification of electronic states in crystalline solids according to their topology was pioneered by Thouless and coworkers, who in 1982 identified the topological invariant for the integer quantum Hall effect [28]. It would take more than twenty years for this way of thinking to become mainstream. Theoretical investigations on quantum spin Hall insulators [11,12,29], now more commonly known as two-dimensional topological insulators, led to the first experimental observations, reported in 2007, on HgTe/CdTe quantum wells [13]. The concepts were soon extended to describe three-dimensional topological insulators [30–32], which were identified in experiments shortly afterwards [33,34].

The concept of the topological insulator is actually very simple. Figure 2 illustrates the basic idea. The energy levels of isolated atoms are quantised. When the atoms come together to form a solid the levels spread out to form a band of states whose energy varies (disperses) with momentum. An insulator is a material with an energy gap that separates the (filled) valence bands from the (empty) conduction bands. If, as the atoms come together and bands form, the gap between the valence and conduction bands never closes, then we have a trivial (i.e. non-topological) insulator or semiconductor. This situation is illustrated in Figure 2(a) for a single pair of levels.

Alternatively, if the levels cross as the solid forms, then for at least some states the energies will be inverted compared with the isolated atom. If, in addition, a gap

forms in the solid where the lower and upper atomic levels cross, then we still have an insulator in the bulk, but now it is a topological insulator, Figure 2(b). The reason why this insulator is topological is that when the levels cross, the valence and conduction bands twist, metaphorically speaking, a bit like going from the simple cylindrical band to the Möbius strip illustrated in Figure 1(c). Put another way, the insulating states in the solid are not smoothly connected to the atomic states.

One interaction which can open a gap where the levels cross in the bulk is spin-orbit coupling (SOC). Accordingly, topological insulators are often made from elements with high atomic numbers (hence large SOC). Despite the bulk gap, the surface (or edge in two dimensions) of a topological insulator must necessarily be conducting in the absence of a magnetic interaction. In simple terms, the states which are inverted in the bulk must cross back to their usual band ordering on moving from inside to outside the solid. When they cross, the energy gap closes and charge can flow. This band crossing happens just beneath the surface, where the bulk lattice periodicity of the material is terminated, Figure 2(c). Gapless surface or edge states are then inevitable. Their existence depends only on there being a transition between two topologically inequivalent states below and above the surface. This aspect of topological materials is referred to as the *bulk-boundary correspondence*.

More detailed analysis shows that the boundary states come in Kramers degenerate pairs (see Section 5.3) which are spin-polarised, meaning that states with opposite spin move in opposite directions along a given boundary, see Figure 2(c). In consequence, these states cannot be backscattered by non-magnetic impurities, because no conduction channel exists for electrons whose direction of motion is reversed but whose spin is unchanged (scattering by non-magnetic impurities preserves the electron spin direction). The flow of spin current, therefore, is dissipationless, which is highly advantageous for low-power spintronics applications. Further exotic properties emerge when TIs are coupled to electric and magnetic fields, or to superconductors [35–37].

In summary, topological insulators are insulating in the bulk but metallic on the surface, carry dissipationless surface spin currents, and are topologically distinct from a normal insulator.

### 3. Electronic band structure in crystalline solids

In this section I review some of the features of electronic states in crystalline solids, as first considered almost a century ago by Felix Bloch in his PhD thesis [38]. We are going to be concerned primarily with systems in which the electrons can be described adequately as non-interacting particles, and for the time being I shall neglect spin.

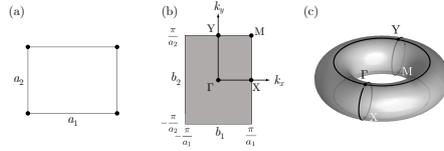
#### 3.1. Lattice Hamiltonian

The single-particle states are solutions of the time-independent Schrödinger equation

$$\mathcal{H}\psi(\mathbf{r}) = E\psi(\mathbf{r}). \quad (1)$$

The Hamiltonian  $\mathcal{H}$  contains the crystal potential  $V(\mathbf{r})$  which satisfies the periodicity condition

$$V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r}) \quad (2)$$



**Figure 3.** (a) Rectangular lattice with primitive lattice vectors  $a_1$  and  $a_2$ . (b) Wigner–Seitz cell (Brillouin zone) centred on  $k_x = k_y = 0$  for the rectangular lattice in (a), with primitive reciprocal lattice vectors  $b_1 = 2\pi/a_1$  and  $b_2 = 2\pi/a_2$ . (c) Toroidal geometry of the BZ when the Bloch functions are defined on the periodic gauge. The rectangular BZ in (b) is first rolled around the  $k_y$  axis, and then the edges at  $k_y = \pm\frac{\pi}{a_2}$  are joined together to form a torus.

for any lattice vector

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \quad (n_1, n_2, n_3 \text{ integers}), \quad (3)$$

where  $\mathbf{a}_1$ ,  $\mathbf{a}_2$  and  $\mathbf{a}_3$  are the primitive lattice vectors.

The Fourier transform of the real-space lattice defines the *reciprocal lattice*, which is an essential tool for describing the properties of waves propagating in a periodic potential. The reciprocal lattice is the set of discrete vectors

$$\mathbf{G} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3 \quad (m_1, m_2, m_3 \text{ integers}), \quad (4)$$

where  $\mathbf{b}_i = (2\pi/v_0)\mathbf{a}_j \times \mathbf{a}_k$  are primitive reciprocal lattice vectors with the property  $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$ . Here,  $v_0$  is the volume of the primitive unit cell of the direct lattice and  $\delta_{ij}$  is the Kronecker delta. Panels (a) and (b) of Figure 3 illustrate the unit cell of a primitive rectangular lattice and the corresponding reciprocal lattice.

### 3.2. Bloch functions

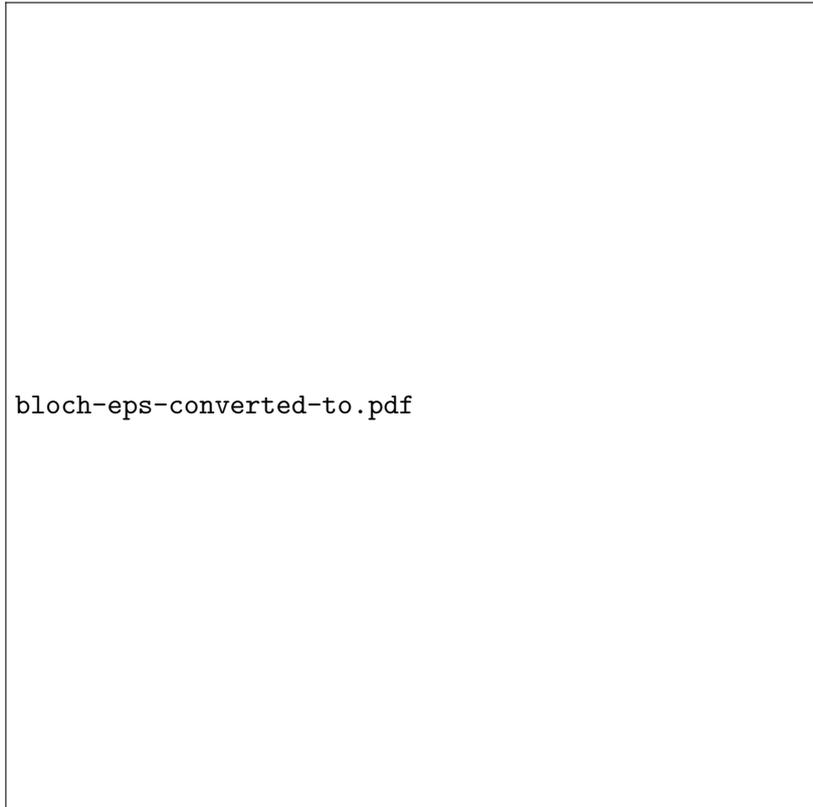
Bloch showed that the solutions of Equations (1)–(3) have the generic form (see Figure 4)

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r}), \quad (5)$$

which comprises a plane wave modulated by a function  $u_{\mathbf{k}}(\mathbf{r})$  that is periodic in the lattice,

$$u_{\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{\mathbf{k}}(\mathbf{r}). \quad (6)$$

The *Bloch wavevector*  $\mathbf{k}$  resembles the wavevector of a free particle, but differs in two important respects. First, it is quantised due to the confinement of the particle in the solid. Second, translational symmetry makes  $\mathbf{k}$  states that differ by a reciprocal lattice vector equivalent, in the sense that  $\psi_{\mathbf{k}}(\mathbf{r})$  and  $\psi_{\mathbf{k}+\mathbf{G}}(\mathbf{r})$  correspond to the same energy eigenstate for any reciprocal lattice vector  $\mathbf{G}$ .



**Figure 4.** Representation of a one-dimensional Bloch function. The real part of the Bloch function is plotted on the left, and the lattice-periodic function is shown on the right. A value  $k = 0.5/a$  has been used, where  $a$  is the spacing between the lattice points (represented by dots).

These two properties mean that the  $\mathbf{k}$  vectors which label unique solutions of the Schrödinger equation are finite in number and can be contained in a bounded volume of  $\mathbf{k}$  space. This volume is known as the Brillouin zone (BZ), and is usually chosen to be the Wigner–Seitz unit cell<sup>1</sup> centred on  $\mathbf{k} = 0$ . The BZ of a primitive rectangular lattice is illustrated in Figure 3.

The fact that  $\mathbf{k}$  and  $\mathbf{k} + \mathbf{G}$  correspond to the same energy eigenstate means that the Bloch functions  $\psi_{\mathbf{k}}(\mathbf{r})$  and  $\psi_{\mathbf{k}+\mathbf{G}}(\mathbf{r})$  can differ at most by a trivial phase factor  $\exp(-i\theta)$ . In what follows we shall choose  $\theta = 0$ , a condition known as the *periodic gauge*. With this choice of phase the BZ has periodic boundary conditions on all its faces, giving it the topology of a  $d$ -torus in  $d$  dimensions, i.e. a circle in one dimension, the surface of a doughnut in two dimensions, and a 3-torus in three dimensions. Figure 3(c) illustrates the 2-torus geometry of the BZ in two dimensions.

The eigenvalues  $E_{\mathbf{k}}$  associated with the Bloch functions form a band of states in  $\mathbf{k}$  space which originates from the overlap of atomic orbitals. The degree of orbital overlap in the solid determines the electronic band width. Atoms have multiple orbitals, so for a given  $\mathbf{k}$  there will be multiple energy levels  $E_{n\mathbf{k}}$ . The additional quantum number  $n$ , called the band index, is used to label the different bands. Each energy band has its own characteristic variation with  $\mathbf{k}$ , and different bands can cross over one another. Band crossing points play an important role in the phenomena we shall discuss here.

## 4. The Berry phase

### 4.1. Quantum mechanical phase factors

*Phase factors* are an important feature of quantum mechanical functions. They take the form  $e^{-i\theta}$ , where  $\theta$  is a (real-valued) phase angle. Some phase factors are trivial. For example, suppose that  $\psi$  is a solution of the time-independent Schrödinger equation, Equation (1). The gauge transformation  $\tilde{\psi} = e^{-i\beta}\psi$ , where  $\beta$  commutes with  $\mathcal{H}$ , simply generates an equivalent solution  $\tilde{\psi}$  with the same eigenvalue  $E$ . There are no observable consequences of such a gauge transformation.

Another example is the dynamical phase of the eigenfunction, which comes from the time-dependent Schrödinger equation,

$$\mathcal{H}\psi = i\hbar\frac{\partial\psi}{\partial t}. \quad (7)$$

Substituting (1) in (7) and integrating for a time-independent  $\mathcal{H}$ , we obtain

$$\psi(t) = e^{-iEt/\hbar}\psi(0), \quad (8)$$

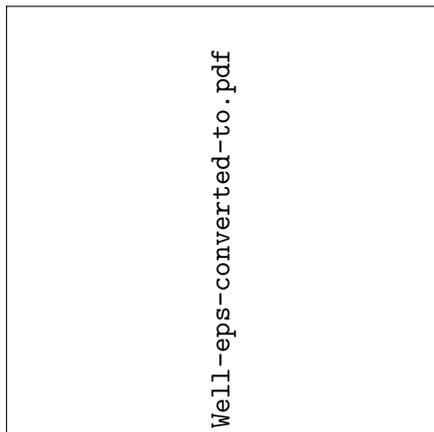
or, more generally,

$$\psi(t) = e^{-i\int E(t)dt/\hbar}\psi(0). \quad (9)$$

We see that the time dependence of the eigenfunction is contained in a time-dependent phase factor.

---

<sup>1</sup>The Wigner–Seitz unit cell is defined as the region of  $\mathbf{k}$  space centred on a reciprocal lattice point  $\Gamma$  such that it contains all  $\mathbf{k}$  states which are closer to  $\Gamma$  than to any other reciprocal lattice point.



**Figure 5.** Illustration of the adiabatic theorem in quantum mechanics. The plot shows the probability density  $|\psi^2|$  for the ground state wave function of a particle in a one-dimensional potential well, as the width  $W$  and depth  $V_0$  of the well are varied around a closed path. If the parameters are changed sufficiently slowly then the wave function adapts so that the particle remains in the instantaneous ground state throughout the cycle.

#### 4.2. Adiabatic changes and the Berry phase

The Berry phase is a *non-trivial* phase that can have observable consequences, and is the central idea of topological band theory. The Berry phase arises when a quantum state is varied adiabatically, through slow changes in the parameters on which the state depends. As we shall show, variations in the Berry phase are particularly strong in the vicinity of points where electronic levels become degenerate.

In his original work [39], Berry considered a general system whose Hamiltonian depends on a set of parameters that can be varied as a function of time. He showed that if the system starts in an eigenstate  $\psi(0)$  of the Hamiltonian, and the system is taken slowly around a closed path in parameter space, then at the end of the cycle the system returns to  $\psi(0)$  but acquires a phase factor  $e^{-i\gamma}$  *in addition to* the usual dynamical phase factor given in Equation (9). The phase angle  $\gamma$ , now known as the *Berry phase*, depends upon how the eigenstates of the Hamiltonian change around the path. In other words, the Berry phase is a function of the geometry of the Hamiltonian in parameter space.

The type of process we are considering is described by the *adiabatic theorem* in quantum mechanics, and is illustrated in Figure 5. The picture shows how the ground-state of a particle in a one-dimensional square potential well evolves as the width and depth of the well are varied around a closed path. If the changes occur sufficiently slowly, then the wave function has time to adapt and the particle remains in the instantaneous ground state throughout the process. By ‘sufficiently slowly’, we mean that the characteristic time-scale for the process (e.g. the time to complete the cycle) must be large compared with  $\hbar/\Delta E$ , where  $\Delta E$  is the separation between the ground state and the rest of the spectrum of eigenstates.

#### 4.3. Cell-periodic Bloch functions and the Bloch Hamiltonian

We are interested in results that apply to systems of non-interacting electronic quasi-particles in periodic potentials. The parameter space in which adiabatic changes take place is in this case the space spanned by  $\mathbf{k}$ , which for present purposes can be re-

garded as a quasi-continuous variable rather than a discrete quantum number.<sup>2</sup> To emphasise the role of  $\mathbf{k}$  as a (quasi-continuous) parameter we shall from now on write the  $\mathbf{k}$ -dependence as an argument rather than as a subscript, e.g.  $E_n(\mathbf{k})$  instead of  $E_{n\mathbf{k}}$ .

In what follows it will be necessary to work with the lattice-periodic part of the Bloch function (see Section 3), which is known as the *cell-periodic* Bloch function,

$$u_n(\mathbf{k}) = e^{-i\mathbf{k}\cdot\mathbf{r}}\psi_n(\mathbf{k}). \quad (10)$$

The energy eigenvalue equation (1) may then be written

$$\mathcal{H}(\mathbf{k})u_n(\mathbf{k}) = E_n(\mathbf{k})u_n(\mathbf{k}), \quad (11)$$

where we have defined the *Bloch Hamiltonian*

$$\mathcal{H}(\mathbf{k}) = e^{-i\mathbf{k}\cdot\mathbf{r}}\mathcal{H}e^{i\mathbf{k}\cdot\mathbf{r}}. \quad (12)$$

The cell-periodic Bloch functions are seen to be eigenstates of  $\mathcal{H}(\mathbf{k})$ , and form a complete set of basis states defined on the primitive unit cell. Hereafter, we shall adopt Dirac's notation and represent the basis state  $u_n(\mathbf{k})$  by the ket  $|u_n(\mathbf{k})\rangle$ .

#### 4.4. Derivation of the Berry phase and Berry connection

Having dispensed with the preliminaries, we now come to the essence of the problem which is to obtain expressions for the Berry phase and related quantities when electronic quasiparticles move slowly around a closed path  $\mathcal{C}$  in  $\mathbf{k}$  space. The physical significance of the calculations is that they apply to the motion of quasiparticles in electric and/or magnetic fields under many conditions of interest, and therefore describe a variety of magnetotransport phenomena.

To proceed, we write the state of the system at time  $t$ , when it has reached a point  $\mathbf{k}$  on the path, as

$$|\psi(t)\rangle = e^{-i\theta(t)}|u_n(\mathbf{k})\rangle, \quad (13)$$

where  $|u_n(\mathbf{k})\rangle$  is one of the normalised eigenstates of  $\mathcal{H}(\mathbf{k})$ , see Equation (11). Consistent with the adiabatic theorem, we assume that the system remains in the same quantum level  $n$  (in other words, in the same band) throughout the process. Substituting  $|\psi(t)\rangle$  into the Schrödinger Equations (1) and (7), and taking the inner product<sup>3</sup> with  $\langle u_n(\mathbf{k})|$ , we obtain

$$\hbar\frac{d\theta}{dt} = E_n(t) - i\hbar\langle u_n(\mathbf{k})|\frac{d}{dt}|u_n(\mathbf{k})\rangle. \quad (14)$$

Hence, if it takes time  $T$  to complete the full closed path, then

$$\theta(T) = \frac{1}{\hbar}\int_0^T E_n(t)dt - i\int_0^T \langle u_n(\mathbf{k})|\frac{d}{dt}|u_n(\mathbf{k})\rangle dt. \quad (15)$$

---

<sup>2</sup>The number of allowed  $\mathbf{k}$  states in the first BZ is equal to the number of primitive unit cells in the crystal, so for a macroscopic system the allowed  $\mathbf{k}$  are very close together relative to the size of the BZ.

<sup>3</sup>The inner product is an integral over the unit cell in real space.

The first term is the usual dynamical phase — see Equation (9). The second term defines the *Berry phase*,

$$\gamma_n = i \int_0^T \langle u_n(\mathbf{k}) | \frac{d}{dt} | u_n(\mathbf{k}) \rangle dt. \quad (16)$$

The Berry phase is always contained in a phase factor  $e^{-i\gamma_n}$  in expressions that relate to observable quantities, and so values of  $\gamma_n$  that differ by an integer multiple of  $2\pi$  are the same for all practical purposes. In other words, the Berry phase is defined modulo  $2\pi$ . This property has important implications for band topology, as will be discussed in Section 4.9.

Time can be eliminated from (16) via the chain rule,

$$\frac{d}{dt} = \nabla_{\mathbf{k}} \cdot \frac{d\mathbf{k}}{dt}, \quad (17)$$

where  $\nabla_{\mathbf{k}}$  stands for  $(\frac{\partial}{\partial k_x}, \frac{\partial}{\partial k_y}, \frac{\partial}{\partial k_z})$ , whereupon

$$\gamma_n = i \oint_{\mathcal{C}} \langle u_n(\mathbf{k}) | \nabla_{\mathbf{k}} | u_n(\mathbf{k}) \rangle \cdot d\mathbf{k}. \quad (18)$$

The matrix element  $\langle u_n | \nabla | u_n \rangle$  is an imaginary number, so the Berry phase is always real.<sup>4</sup> The integrand in Equation (18) defines the *Berry connection* (or *Berry potential*)  $\mathcal{A}_n(\mathbf{k})$ ,

$$\mathcal{A}_n(\mathbf{k}) = i \langle u_n(\mathbf{k}) | \nabla_{\mathbf{k}} | u_n(\mathbf{k}) \rangle, \quad (19)$$

so that

$$\gamma_n = \oint_{\mathcal{C}} \mathcal{A}_n(\mathbf{k}) \cdot d\mathbf{k}. \quad (20)$$

As we shall see in Equation (24),  $\mathcal{A}_n(\mathbf{k})$  is analogous to the magnetic vector potential in electromagnetism. One property they have in common is that both potentials are gauge dependent. To see this, we apply a gauge transformation to  $|u_n(\mathbf{k})\rangle$ :

$$\begin{aligned} |\tilde{u}_n(\mathbf{k})\rangle &= e^{-i\beta(\mathbf{k})} |u_n(\mathbf{k})\rangle \\ \rightarrow \tilde{\mathcal{A}}_n &= \mathcal{A}_n + \nabla_{\mathbf{k}}\beta(\mathbf{k}). \end{aligned} \quad (21)$$

An observable quantity cannot depend on an arbitrary choice of phase  $\beta(\mathbf{k})$ , so  $\mathcal{A}_n$  is not an observable quantity.

#### 4.5. Parallel transport

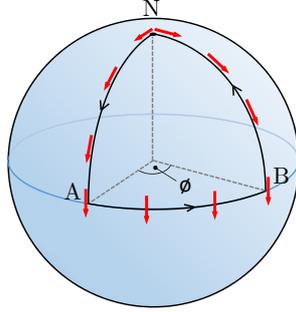
Let us pause for a moment. The Berry phase is an example of a phenomenon called *anholonomy*, in which a system variable fails to return to its original value when the

---

<sup>4</sup>This can be seen by applying  $\nabla_{\mathbf{k}}$  to the normalization condition  $\langle u_n | u_n \rangle = 1$ :

$$\nabla \langle u_n | u_n \rangle = \langle \nabla u_n | u_n \rangle + \langle u_n | \nabla u_n \rangle = \langle u_n | \nabla u_n \rangle^* + \langle u_n | \nabla u_n \rangle = 0,$$

hence  $\langle u_n | \nabla u_n \rangle = -\langle u_n | \nabla u_n \rangle^*$  is purely imaginary and therefore  $\gamma_n$  is purely real.



**Figure 6.** Parallel transport of a vector on a sphere. The vector travels around the closed path N–A–B–N while remaining parallel to the surface of the sphere. The angle between the initial and final directions of the vector is  $\phi$ .

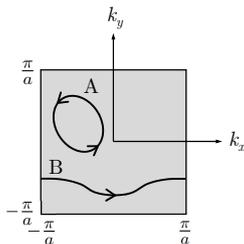
system is driven by other variables around a cycle. A simple classical example is the parallel transport of a vector (e.g. an arrow) on a sphere, Figure 6. Starting at the north pole (N), an arrow is transported down a meridian until it reaches the equator (A), at which point it follows the equator to point B, then returns back to N along the meridian through B. The arrow remains parallel to the local surface at all points on the path by the minimum of adjustments to its orientation. At the end of the journey, the arrow is seen to point in a different direction than at the start of the journey.

In this example, the parameters of the path are the coordinates on the surface of the sphere, and the anholonomy occurs because of the curved geometry of the parameter space. If the closed path had been on a flat surface there would be no anholonomy. The Berry phase is a quantum version of parallel transport, in which the Bloch Hamiltonian plays the role of the curved surface and the wavevectors  $\mathbf{k}$  act as the parameters that define the closed path. For obvious reasons, the Berry phase is often called a geometrical phase.

#### 4.6. Berry curvature

At this point we need to make a clarification about closed paths in  $\mathbf{k}$  space. Consider the two paths drawn on the 2D BZ in Figure 7. The simplest type of closed path is a *contractible path* (path A), i.e. one that can be continuously shrunk to a point. However, because Bloch functions are periodic in the Brillouin zone a *non-contractible* closed path is also possible (path B), where the path crosses the BZ boundary then wraps around to the equivalent point on the opposite face of the BZ. The Berry phase associated with the latter type of path is often called a Zak phase [40].

In the case of a contractible path  $\mathcal{C}$ , one can use Stokes' theorem to convert the



**Figure 7.** Closed paths in two-dimensional  $\mathbf{k}$  space. The shaded square is the BZ of a square lattice (lattice parameter  $a$ ). Path A is contractible and path B is non-contractible. Path B is still a closed path because points on opposite sides of the BZ related by a reciprocal lattice vector are equivalent.

contour integral in Equation (20) into a surface integral:

$$\oint_{\mathcal{C}} \mathcal{A}_n \cdot d\mathbf{k} = \int_{\mathcal{S}} (\nabla_{\mathbf{k}} \times \mathcal{A}_n) \cdot d\mathbf{S}, \quad (22)$$

where  $\mathcal{S}$  is a surface in  $\mathbf{k}$  space bounded by  $\mathcal{C}$ . For Equation (22) to apply,  $\mathcal{A}_n(\mathbf{k})$  must be well defined over  $\mathcal{S}$ . Assuming that to be the case, the phase change around  $\mathcal{C}$  may be written

$$\gamma_n = \int_{\mathcal{S}} \mathcal{B}_n \cdot d\mathbf{S}, \quad (23)$$

where

$$\mathcal{B}_n = \nabla_{\mathbf{k}} \times \mathcal{A}_n \quad (24)$$

is an emergent gauge field called the *Berry curvature*, and is a  $\mathbf{k}$ -space analogue of the magnetic induction.<sup>5</sup> We see from (23) that the Berry phase around  $\mathcal{C}$  is the flux of Berry curvature (*Berry flux*) through the surface  $\mathcal{S}$ .

From Equations (21) and (24), together with the identity  $\nabla \times (\nabla \phi) = 0$ , one can see that  $\mathcal{B}_n$  is gauge-invariant and therefore an observable quantity. Since  $u_n$  and  $\psi_n$  are related by a phase factor, Equation (5), and  $\psi_n$  is periodic in the reciprocal lattice, it follows that  $\mathcal{B}_n$  is also periodic in the reciprocal lattice.

From (23) it looks like  $\gamma$  should also be gauge-invariant. As we shall show in Section 4.9,  $\gamma$  is actually gauge invariant to within an integer multiple of  $2\pi$ . Hence, the Berry phase factor is observable too. Note that the Zak phase is defined from a non-contractible path, and so it cannot be expressed in term of a Berry curvature.

#### 4.7. Berry curvature in terms of virtual transitions

Although the results derived above for  $\gamma_n$  and  $\mathcal{B}_n$  are expressed purely in terms of Bloch functions for a single band  $n$ , a non-vanishing Berry phase/curvature actually

<sup>5</sup>In the literature, the Berry curvature is often denoted by  $\mathbf{\Omega}$  or  $\mathbf{F}$ . Here, I use  $\mathcal{A}$  and  $\mathcal{B}$  for the Berry potential and Berry curvature to reinforce the analogy with magnetic fields.

depends upon the existence of other bands  $m \neq n$ . To see this, combine Equations (19), (23) and (24), and write (omitting the  $\mathbf{k}$  dependence for simplicity)

$$\begin{aligned}\gamma_n &= -\text{Im} \int_{\mathcal{S}} (\nabla \times \langle u_n | \nabla | u_n \rangle) \cdot d\mathbf{S} \\ &= -\text{Im} \int_{\mathcal{S}} \langle \nabla u_n | \times | \nabla u_n \rangle \cdot d\mathbf{S} \\ &= -\text{Im} \int_{\mathcal{S}} \sum_{m \neq n} \langle \nabla u_n | u_m \rangle \times \langle u_m | \nabla u_n \rangle \cdot d\mathbf{S}.\end{aligned}\tag{25}$$

The last line is obtained by the quantum mechanical closure relation, and the  $n = m$  term is omitted from the summation because this term is real and we are only concerned with the imaginary part. An alternative form of Equation (25) is obtained from the following identity,<sup>6</sup>

$$\langle u_m | \nabla | u_n \rangle = \frac{\langle u_m | (\nabla \mathcal{H}) | u_n \rangle}{E_n - E_m}, \quad n \neq m.\tag{26}$$

Putting (26) in (25), and comparing with (23), one obtains

$$\mathcal{B}_n = -\text{Im} \sum_{m \neq n} \frac{\langle u_n | (\nabla \mathcal{H}) | u_m \rangle \times \langle u_m | (\nabla \mathcal{H}) | u_n \rangle}{(E_n - E_m)^2}.\tag{27}$$

Formula (27) shows that the Berry curvature in the  $n^{\text{th}}$  band is determined by virtual transitions to other bands, and that bands which are closest in energy to band  $n$  will tend to contribute the most. In other words, the Berry curvature is an effective (emergent) field which arises when we project the effects of inter-band transitions onto the motion of an electronic quasiparticle in a single band.

Finally, note that  $\sum_n \mathcal{B}_n = 0$ , and hence  $\sum_n \gamma_n = 0$ . This follows because the numerator in (27) has the opposite sign for the terms  $(n, m)$  and  $(m, n)$  in the double summation.

#### 4.8. Linear band crossings in 3D

Suppose that the contour  $\mathcal{C}$  lies close to a point  $P$  where two or more bands cross.<sup>7</sup> The denominator of Equation (27) implies that the Berry curvature for band  $n$  is dominated by the contributions from the other level(s)  $m$  involved in the degeneracy. For simplicity, we shall consider a two-level system, and without loss of generality we

<sup>6</sup>One can obtain Equation (26) by applying the  $\nabla$  operator to the Schrödinger equation (11),

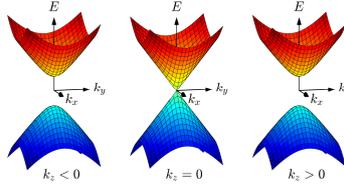
$$\nabla(\mathcal{H}|u_n\rangle) = E_n \nabla|u_n\rangle + (\nabla E_n)|u_n\rangle,$$

and then using the fact that  $\langle u_m | u_n \rangle = \delta_{mn}$  to show that for  $m \neq n$ ,

$$\begin{aligned}E_n \langle u_m | \nabla | u_n \rangle &= \langle u_m | \nabla \mathcal{H} | u_n \rangle \\ &= \langle u_m | \mathcal{H} | \nabla u_n \rangle + \langle u_m | (\nabla \mathcal{H}) | u_n \rangle \\ &= E_m \langle u_m | \nabla | u_n \rangle + \langle u_m | (\nabla \mathcal{H}) | u_n \rangle,\end{aligned}$$

which leads directly to Equation (26).

<sup>7</sup>The degeneracy point  $P$  must not lie on  $\mathcal{C}$ , otherwise the adiabatic condition is not satisfied, see Section 4.2.



**Figure 8.** Dispersion surfaces in the vicinity of a linear band crossing point at  $\mathbf{k} = 0$ . The upper and lower bands  $E^\pm(\mathbf{k})$ , calculated from Equation (29), are plotted as a function of  $(k_x, k_y)$ , for three  $k_z$  values that span the crossing point.

take P to be at  $(E, \mathbf{k}) = (0, 0)$ . We shall also assume for now that the dispersion  $E_n(\mathbf{k})$  is linear and isotropic (the case of an anisotropic dispersion is discussed below).

As there are two Bloch functions for each  $\mathbf{k}$  the Hamiltonian can be represented by a  $2 \times 2$  Hermitian matrix, which in the vicinity of P takes the form

$$\mathcal{H}(\mathbf{k}) = \hbar v \begin{pmatrix} k_z & k_x - ik_y \\ k_x + ik_y & -k_z \end{pmatrix}. \quad (28)$$

We see that the parameters of  $\mathcal{H}$  are  $k_x, k_y$  and  $k_z$ . The eigenvalues of (28) are

$$\begin{aligned} E^\pm(\mathbf{k}) &= \pm \hbar v \sqrt{k_x^2 + k_y^2 + k_z^2}, \\ &= \pm \hbar v k, \end{aligned} \quad (29)$$

which is the assumed linear dispersion with an isotropic speed  $v$ , see Figure 8. The normalised eigenvectors written in spherical coordinates in  $\mathbf{k}$  space<sup>8</sup> are

$$|+\rangle = \begin{pmatrix} \cos \frac{\theta_{\mathbf{k}}}{2} e^{-i\phi_{\mathbf{k}}} \\ \sin \frac{\theta_{\mathbf{k}}}{2} \end{pmatrix}, \quad |-\rangle = \begin{pmatrix} \sin \frac{\theta_{\mathbf{k}}}{2} e^{-i\phi_{\mathbf{k}}} \\ -\cos \frac{\theta_{\mathbf{k}}}{2} \end{pmatrix}. \quad (30)$$

The two components of the eigenvectors represent the amplitudes of the two basis vectors of the two-level system, e.g. two different orbitals.

We can now calculate the Berry curvature. In spherical coordinates  $\nabla_{\mathbf{k}} = \hat{\mathbf{k}}(\frac{\partial}{\partial k}) + \hat{\boldsymbol{\theta}}_{\mathbf{k}}(\frac{1}{k} \frac{\partial}{\partial \theta_{\mathbf{k}}}) + \hat{\boldsymbol{\phi}}_{\mathbf{k}}(\frac{1}{k \sin \theta_{\mathbf{k}}} \frac{\partial}{\partial \phi_{\mathbf{k}}})$ , and hence from Equation (19) we find

$$\mathcal{A}^+(\mathbf{k}) = \frac{1}{2k} \cot \frac{\theta_{\mathbf{k}}}{2} \hat{\boldsymbol{\phi}}_{\mathbf{k}}, \quad \mathcal{A}^-(\mathbf{k}) = \frac{1}{2k} \tan \frac{\theta_{\mathbf{k}}}{2} \hat{\boldsymbol{\phi}}_{\mathbf{k}}, \quad (31)$$

---

<sup>8</sup> $(k_x, k_y, k_z) = k(\sin \theta_{\mathbf{k}} \cos \phi_{\mathbf{k}}, \sin \theta_{\mathbf{k}} \sin \phi_{\mathbf{k}}, \cos \theta_{\mathbf{k}})$ .

for the  $|+\rangle$  and  $|-\rangle$  eigenstates. By writing the curl of a vector in spherical coordinates<sup>9</sup> we can obtain from Equations (24) and (31)

$$\mathcal{B}^\pm(\mathbf{k}) = \mp \frac{1}{2k^2} \hat{\mathbf{k}}. \quad (32)$$

We see that  $\mathcal{B}^-(\mathbf{k})$  and  $\mathcal{B}^+(\mathbf{k})$  resemble the fields from point charges (monopoles) in  $\mathbf{k}$  space of strength  $+\frac{1}{2}$  and  $-\frac{1}{2}$ , respectively, see Figure 9. The positive monopole charge is associated with the negative energy solution, and *vice versa*. The monopole and anti-monopole coincide at the nodes, but the associated Berry curvature fields do not cancel one another because they belong to different bands. Degeneracy points, therefore, are sources and sinks of Berry curvature.

The Berry phase for adiabatic transport around  $\mathcal{C}$  can be obtained from either the Berry connection, Equations (20) and (31), or the Berry curvature, Equations (23) and (32). The result will be the same, modulo  $2\pi$ . In the latter case, we have

$$\begin{aligned} \gamma_n^\pm &= \mp \frac{1}{2} \int_{\mathcal{S}} \frac{\hat{\mathbf{k}} \cdot d\mathbf{S}}{k^2} \\ &= \mp \frac{\Omega}{2}, \end{aligned} \quad (33)$$

where  $\Omega$  is the solid angle subtended by  $\mathcal{S}$  at  $P$ . A special case is when  $\mathcal{C}$  is restricted to a plane which contains  $P$ . Then,  $\gamma_n^\pm = \mp\pi$  if  $\mathcal{C}$  encloses  $P$ , or  $\gamma_n^\pm = 0$  otherwise.

The crossing point of two singly degenerate bands in 3D  $\mathbf{k}$  space is called a *Weyl point* or *Weyl node*, for the reason given in Section 8. It is conventional to assign a chirality  $\chi$  to the nodes, such that  $\chi = +1$  if the lower-energy band acts as a source of Berry curvature (positive monopole charge) and  $\chi = -1$  if the lower-energy band acts as a sink of Berry curvature (negative monopole charge). The Weyl node described above has  $\chi = +1$ . A change of sign in front of the Hamiltonian (28) generates the opposite chirality. Electronic quasiparticles in the vicinity of Weyl nodes have special properties which differ from those of conventional electrons. We will elaborate more in Section 8.

#### 4.9. Chern theorem and Chern number

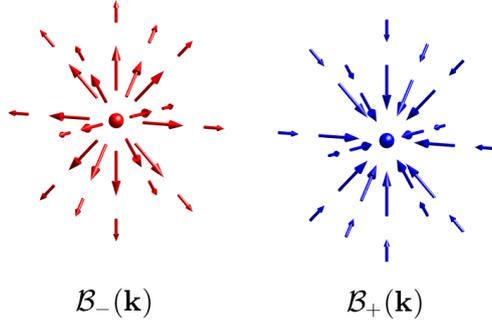
From Equation (32) and Gauss' theorem, it follows that the integral of  $\mathcal{B}^\pm(\mathbf{k})$  over a surface surrounding  $P$  in  $\mathbf{k}$  space is  $\mp 2\pi$ . This is a particular case of the *Chern theorem*, which states that the Berry flux calculated for any closed 2D surface  $\mathcal{S}$  in  $\mathbf{k}$  space is quantised in multiples of  $2\pi$ , i.e.

$$\oint_{\mathcal{S}} \mathcal{B}_n \cdot d\mathbf{S} = 2\pi C_n. \quad (34)$$

---

9

$$\begin{aligned} \nabla_{\mathbf{k}} \times \mathbf{F} &= \frac{1}{k \sin \theta_{\mathbf{k}}} \left\{ \frac{\partial}{\partial \theta_{\mathbf{k}}} (F_\phi \sin \theta_{\mathbf{k}}) - \frac{\partial F_\theta}{\partial \phi_{\mathbf{k}}} \right\} \hat{\mathbf{k}} \\ &+ \frac{1}{k} \left\{ \frac{1}{\sin \theta_{\mathbf{k}}} \frac{\partial F_k}{\partial \phi_{\mathbf{k}}} - \frac{\partial}{\partial k} (k F_\phi) \right\} \hat{\boldsymbol{\theta}}_{\mathbf{k}} \\ &+ \frac{1}{k} \left\{ \frac{\partial}{\partial k} (k F_\theta) - \frac{\partial F_k}{\partial \theta} \right\} \hat{\boldsymbol{\phi}}_{\mathbf{k}}, \end{aligned}$$



**Figure 9.** Representation of the Berry curvature associated with the lower (left) and upper (right) bands that form a Weyl point, Equation (32). Both vector fields,  $\mathcal{B}^-(\mathbf{k})$  and  $\mathcal{B}^+(\mathbf{k})$ , are centred on  $\mathbf{k} = 0$ .

The integer  $C_n$  is a topological invariant known as the *Chern number* of the surface. The Fermi surface can be such a closed surface, in which case the Fermi surface will have a well-defined Chern number. For the isotropic band crossing analysed in Section 4.8 the Fermi surface is spherical, and depending whether the Fermi level is higher or lower in energy than the Weyl node (i.e. whether the Fermi surface is electron-like or hole-like), the Fermi surface has a Chern number  $C_n = -1$  or  $+1$ , respectively.

### *Proof of the Chern theorem*

To understand the Chern theorem (34) better we need to appreciate the effect of a gauge transformation on the different expressions for the Berry phase, specifically (see Sections 4.4 and 4.6)

$$\gamma = \oint_{\mathcal{C}} \mathcal{A}(\mathbf{k}) \cdot d\mathbf{k} = \int_{\mathcal{S}} \mathcal{B} \cdot d\mathbf{S}. \quad (35)$$

The band index  $n$  is suppressed to simplify the notation. Now consider the gauge transformation

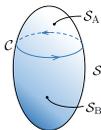
$$|\tilde{u}(\mathbf{k})\rangle = e^{-i\beta(\mathbf{k})}|u(\mathbf{k})\rangle. \quad (36)$$

If the functions  $u(\mathbf{k})$  and  $\tilde{u}(\mathbf{k})$  are to be well-behaved they must be single-valued functions of  $\mathbf{k}$ , and hence  $\exp(-i\beta)$  must also be single-valued. Therefore, around any closed path  $\mathcal{C}$  in  $\mathbf{k}$  space the initial and final values of  $\beta$  must satisfy

$$\beta_f - \beta_i = 2\pi m, \quad (m = \text{integer}). \quad (37)$$

The effect of a gauge transformation on  $\mathcal{A}$  was given in Equation (21),

$$\tilde{\mathcal{A}} = \mathcal{A} + \nabla_{\mathbf{k}}\beta(\mathbf{k}). \quad (38)$$



**Figure 10.** Surfaces for proof of the Chern theorem. The closed surface  $\mathcal{S}$  is divided into two surfaces  $\mathcal{S}_A$  and  $\mathcal{S}_B$  having a common boundary  $\mathcal{C}$ .

Integrating this equation around  $\mathcal{C}$  and using (20) and (37), we obtain

$$\tilde{\gamma} = \gamma + 2\pi m. \quad (39)$$

At first glance this result seems at odds with Equation (35). Previously we found that the Berry curvature  $\mathcal{B}$  is gauge-invariant, which according to (35) implies that  $\gamma$  is gauge-invariant, whereas Equation (39) shows that under a gauge transformation  $\gamma$  can change by an integer multiple of  $2\pi$ . The reason for this apparent paradox is that the second equality in (35) came from application of Stokes' theorem, which requires  $\mathcal{A}$  to be well-behaved over the entire surface  $\mathcal{S}$ . If we choose a gauge in which this requirement is satisfied, then Equation (35) will hold as an equality. If not, (35) will only hold modulo  $2\pi$ .

In order to prove the Chern theorem, Equation (34), we shall evaluate the integral of  $\mathcal{B}$  over a closed 2D surface  $\mathcal{S}$ , using what we have just learnt about the effects of a gauge transformation. To proceed, we divide  $\mathcal{S}$  into two regions A and B having a common boundary  $\mathcal{C}$ , so that  $\mathcal{S} = \mathcal{S}_A + \mathcal{S}_B$ , see Figure 10. We choose a gauge such that Stokes' law applies in region A, and calculate the Berry phase  $\gamma_A$  from the first equality in (35). The line integral around  $\mathcal{C}$  is taken in the sense related by the right-hand rule to the outward normal of  $\mathcal{S}_A$ . We then repeat for region B, but choosing a gauge in which Stokes' law applies in B. The gauges used for B and A need not be the same. The direction of circulation around  $\mathcal{C}$  is reversed because the outward normal of B is opposite to that of A, so the line integral gives  $-\gamma_B$ . Combining the two terms, we obtain

$$\begin{aligned} \oint_{\mathcal{S}} \mathcal{B} \cdot d\mathbf{S} &= \int_{\mathcal{S}_A} \mathcal{B} \cdot d\mathbf{S} + \int_{\mathcal{S}_B} \mathcal{B} \cdot d\mathbf{S} \\ &= \gamma_A - \gamma_B. \end{aligned} \quad (40)$$

If the gauges used for regions A and B are the same then  $\gamma_A = \gamma_B$  and the total Berry flux is zero. However, if the gauges are different, Equation (39) tells us that  $\gamma_A$  and  $\gamma_B$  can differ by an integer multiple of  $2\pi$ , and when put in Equation (40) this proves the Chern theorem.

The existence of a non-zero Chern number signals that  $\mathcal{A}$  is not well-behaved over the entirety of  $\mathcal{S}$ . In other words, there is some kind of 'twist' in the Bloch function manifold. For example, in the band-crossing model presented in Section 4.8 one can

see from Equation (31) that there are singularities in  $\mathcal{A}^\pm$  at the poles,  $\theta = 0$  or  $\pi$ . Therefore, a non-zero Chern number tells us that the Bloch function manifold has a non-trivial topology on  $\mathcal{S}$ .

#### 4.10. Fermion doubling theorem

The Chern theorem leads us to another fundamental result known as the Nielsen–Ninomiya *fermion doubling* theorem [41], which requires that the sum of the chiralities of the Weyl nodes in a 3D BZ must vanish,

$$\sum_{j \in \text{BZ}} \chi_j = 0. \quad (41)$$

It follows that Weyl nodes always come in pairs with opposite chirality (hence, *fermion doubling*).

To prove this result let us choose  $\mathcal{S}$  to enclose the entire BZ, i.e.  $\mathcal{S}$  is formed by the BZ boundary. Recall that the Berry curvature is periodic in the reciprocal lattice. On opposite faces of the BZ, therefore, the Berry curvature is the same but the outward normals of the faces point in opposite directions. Hence, the contribution to the integral in Equation (34) from such a pair of BZ faces cancels. For any band, therefore, the net Chern number defined over the entire BZ surface must vanish. Equation (41) then follows immediately because the Chern number of a closed surface is simply the sum of the chiralities of the Weyl nodes inside the surface.

### 5. Symmetry considerations

Symmetry plays a key rôle in determining the existence and distribution of band degeneracies in  $\mathbf{k}$  space.

#### 5.1. Symmetry and degeneracy

If a system possesses a symmetry  $S$  it means that the system is invariant under  $S$ , i.e. it looks the same before and after the  $S$  transformation. Another way to say this is that the energy of the system remains unchanged under  $S$ .

Let  $U$  be the quantum mechanical operator that represents  $S$ , and let  $|\psi\rangle$  be an eigenfunction of  $\mathcal{H}$  with eigenvalue  $E$ ,

$$\mathcal{H}|\psi\rangle = E|\psi\rangle. \quad (42)$$

If the state  $U|\psi\rangle$  is to have the same energy as  $|\psi\rangle$ , then

$$\begin{aligned} \mathcal{H}U|\psi\rangle &= EU|\psi\rangle \\ &= U\mathcal{H}|\psi\rangle. \end{aligned} \quad (43)$$

Therefore,  $U$  commutes with  $\mathcal{H}$ . Now suppose that  $U|\psi\rangle = |\phi\rangle$ . The first line of (43) may then be written

$$\mathcal{H}|\phi\rangle = E|\phi\rangle. \quad (44)$$

If  $|\phi\rangle = c|\psi\rangle$ , where  $c$  is a (complex) constant, then (44) is equivalent to (42) and there is nothing interesting to say. But if  $|\psi\rangle$  and  $|\phi\rangle$  are not simply proportional to one another, then  $|\psi\rangle$  and  $U|\psi\rangle$  are distinct eigenfunctions of  $\mathcal{H}$  belonging to the same energy level  $E$ . In other words, the  $S$  symmetry of  $\mathcal{H}$  enforces the degeneracy of  $|\psi\rangle$  and  $|\phi\rangle$ .

This analysis illustrates how symmetry can cause degeneracy. Conversely, when one observes degenerate eigenstates there is often (but not always) an underlying symmetry that is responsible for the degeneracy. When degeneracy is associated with symmetry it is called *symmetry-protected degeneracy*. When there is no symmetry involved, it is called *accidental degeneracy*.

## 5.2. Inversion and time-reversal symmetry

The inversion (or parity) transformation  $\mathcal{P}$  maps  $\mathbf{r}$  onto  $-\mathbf{r}$ , relative to a suitably defined origin. If we apply  $\mathcal{P}$  to a Bloch function  $\psi_{\mathbf{k}}(\mathbf{r})$  (see Section 3.2) we obtain another Bloch function  $\psi_{-\mathbf{k}}(\mathbf{r})$ :

$$\begin{aligned}\mathcal{P}\psi_{\mathbf{k}}(\mathbf{r}) &= e^{-i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(-\mathbf{r}) \\ &= e^{i(-\mathbf{k})\cdot\mathbf{r}}u_{-\mathbf{k}}(\mathbf{r}) \\ &= \psi_{-\mathbf{k}}(\mathbf{r})\end{aligned}\tag{45}$$

Here we have simply changed the name of  $u_{\mathbf{k}}(-\mathbf{r})$  to  $u_{-\mathbf{k}}(\mathbf{r})$ . The function  $\psi_{-\mathbf{k}}(\mathbf{r})$  is a Bloch function because  $u_{-\mathbf{k}}(\mathbf{r})$  is cell-periodic:  $u_{-\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{\mathbf{k}}(-\mathbf{r} - \mathbf{R}) = u_{\mathbf{k}}(-\mathbf{r}) = u_{-\mathbf{k}}(\mathbf{r})$ . Note that all lattices have inversion symmetry, so if  $\mathbf{R}$  is a lattice vector then  $-\mathbf{R}$  is one too. We shall sometimes write that  $\mathcal{P}$  maps  $\mathbf{k}$  to  $-\mathbf{k}$ , by which we mean is that  $\mathcal{P}$  transforms the Bloch function associated with  $\mathbf{k}$  into the Bloch function associated with  $-\mathbf{k}$ .

The treatment of time-reversal symmetry is more subtle. In classical mechanics, the equation of motion for a conservative force is invariant under  $t \rightarrow -t$ , and we expect a similar relation to hold in quantum mechanics. However, a simple reversal of time in Equation (7) does not produce the desired effect. Instead, one needs to define an operator  $\mathcal{T}$  which performs complex conjugation in addition to reversing time. Such an operator is termed *antiunitary*, and has the antilinear property,

$$\mathcal{T}c|\psi\rangle = c^*\mathcal{T}|\psi\rangle,\tag{46}$$

for a complex number  $c$ . In contrast,  $\mathcal{P}$  is a *unitary* operator,  $\mathcal{P}c|\psi\rangle = c\mathcal{P}|\psi\rangle$ .

As a simple example, consider the effect of  $\mathcal{T}$  on the function  $\psi = e^{i(kx - \omega t)}$ , which describes a plane wave travelling in the  $+x$  direction. Reversing the sign of  $t$  and taking the complex conjugate we obtain  $\mathcal{T}\psi = e^{i(-kx - \omega t)}$ , which is a plane wave travelling in the  $-x$  direction, as expected. We also see that in this example,  $\mathcal{T}$  is equivalent to changing  $k$  into  $-k$ . Next, apply  $\mathcal{T}$  to a Bloch function:

$$\begin{aligned}\mathcal{T}\psi_{\mathbf{k}}(\mathbf{r}) &= e^{-i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}^*(\mathbf{r}) \\ &= e^{i(-\mathbf{k})\cdot\mathbf{r}}\tilde{u}_{-\mathbf{k}}(\mathbf{r}) \\ &= \tilde{\psi}_{-\mathbf{k}}(\mathbf{r}).\end{aligned}\tag{47}$$

Very much as in Equation (45), we have renamed  $u_{\mathbf{k}}^*(\mathbf{r})$  as  $\tilde{u}_{-\mathbf{k}}(\mathbf{r})$ . The tilde is to

indicate that  $\tilde{u}_{-\mathbf{k}}(\mathbf{r})$  and  $\tilde{\psi}_{-\mathbf{k}}(\mathbf{r})$  in Equation (47) are not necessarily the same as  $u_{-\mathbf{k}}(\mathbf{r})$  and  $\psi_{-\mathbf{k}}(\mathbf{r})$  in Equation (45), for example  $\mathcal{T}$  also reverses the direction of spin. The function  $\psi_{-\mathbf{k}}(\mathbf{r})$  is a Bloch function because  $u_{\mathbf{k}}^*(\mathbf{r})$  is lattice periodic. So the effect of  $\mathcal{T}$  is equivalent to changing a Bloch function represented by  $\mathbf{k}$  into one represented by  $-\mathbf{k}$ .

In practice, time-reversal symmetry breaking in crystalline solids is usually associated with magnetic order or the presence of magnetic fields, because magnetic moments and fields are generated by circulating currents, and currents reverse direction when time is reversed. The presence or absence of inversion symmetry can be identified from the crystallographic space group or, if the material is magnetically ordered, from the magnetic space group.

### 5.3. Kramers' theorem

Further analysis of the time-reversal symmetry operation in quantum mechanics [42] reveals that for a state  $|\psi\rangle$  with total angular momentum quantum number  $J$ ,

$$\mathcal{T}^2|\psi\rangle = (-1)^{2J}|\psi\rangle. \quad (48)$$

We can use this result to prove a general property of half-integer spin systems with time-reversal symmetry, which is that every energy level is  $n$ -fold degenerate, where  $n$  is an even integer (not necessarily the same for each level). This is known as Kramers' theorem.

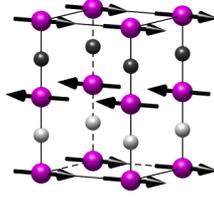
We infer from the arguments in Section 5.1 that if  $|\psi\rangle$  is an eigenfunction of a  $\mathcal{T}$ -invariant Hamiltonian, then  $|\psi\rangle$  and  $\mathcal{T}|\psi\rangle$  are degenerate eigenfunctions providing they are not linearly related. Let us assume the opposite, that  $\mathcal{T}|\psi\rangle = c|\psi\rangle$ . In that case, the antilinear property (46) means that  $\mathcal{T}^2|\psi\rangle = \mathcal{T}c|\psi\rangle = c^*\mathcal{T}|\psi\rangle = |c|^2|\psi\rangle$ . Since  $|c|^2 > 0$ , our assumption contradicts Equation (48), which says that for a half-integer spin system  $\mathcal{T}^2|\psi\rangle = -|\psi\rangle$ . We conclude, therefore, that  $|\psi\rangle$  and  $\mathcal{T}|\psi\rangle$  must be distinct eigenfunctions. It follows that the eigenfunctions belonging to each energy level come in degenerate, time-reversal related, pairs, which establishes the theorem.

### 5.4. Effect of $\mathcal{P}$ and $\mathcal{T}$ on electronic bands

If a crystal potential has inversion symmetry, then from Equation (45) the energy of the Bloch states  $\mathbf{k}$  and  $-\mathbf{k}$  are the same:  $E_n(\mathbf{k}) = E_n(-\mathbf{k})$ . The relation  $E_n(\mathbf{k}) = E_n(-\mathbf{k})$  also holds if the crystal has time-reversal symmetry, even if it does not have inversion symmetry. In this case Equation (47) applies, and the degeneracy is enforced by Kramers' theorem.

If both  $\mathcal{P}$  and  $\mathcal{T}$  are present, then  $\mathcal{P}$  combined with  $\mathcal{T}$  (written  $\mathcal{P} \times \mathcal{T}$ ) maps  $\mathbf{k} \rightarrow \mathbf{k}$ , and so Kramers' theorem requires that each band is at least doubly degenerate throughout the BZ. The same applies for systems which are invariant under  $\mathcal{P} \times \mathcal{T}$  but not under  $\mathcal{P}$  and  $\mathcal{T}$  separately. An example of the latter type of system is the antiferromagnetic structure illustrated in Figure 11.

If only  $\mathcal{T}$  or only  $\mathcal{P}$  is present, then bands are singly degenerate except at points where bands cross accidentally and at certain special points in the BZ where other spatial or magnetic symmetries leave  $\mathbf{k}$  invariant enforcing degeneracies at those positions. An exception is the so-called *time-reversal invariant momenta* (TRIM) for the case when only  $\mathcal{T}$  is present. The TRIM are points where  $-\mathbf{k} = \mathbf{k} + \mathbf{G}$ . At these points,



**Figure 11.** An A-type antiferromagnet that has neither  $\mathcal{P}$  nor  $\mathcal{T}$  symmetry, but has  $\mathcal{P} \times \mathcal{T}$  symmetry. In the paramagnetic phase there are centres of inversion symmetry in the layers of light grey and black atoms, at positions  $0, 0, \frac{1}{4}$  and  $0, 0, \frac{3}{4}$ , respectively. These inversion centres are lost in the antiferromagnetic phase because  $\mathcal{P}$  acting through one of these centres takes a spin pointing to the right onto a site where the spin points to the left. Note that the direction of a magnetic moment remains unchanged under  $\mathcal{P}$  but reverses under  $\mathcal{T}$ .

$\mathbf{k}$  and  $-\mathbf{k}$  are equivalent, so Kramers degeneracy is enforced by  $\mathcal{T}$ . In  $D$  dimensions the number of TRIM is given by  $2^D$ . Their locations are illustrated in Figure 12 for  $D = 2$  and 3.

### 5.5. Symmetry of Berry curvature under $\mathcal{P}$ and $\mathcal{T}$

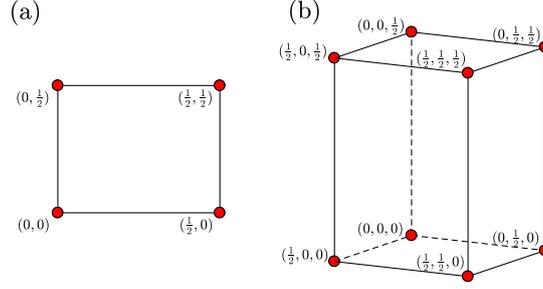
The cell-periodic Bloch functions transform under  $\mathcal{P}$  and  $\mathcal{T}$  according to  $\mathcal{P}u_{\mathbf{k}}(\mathbf{r}) = u_{-\mathbf{k}}(\mathbf{r})$  and  $\mathcal{T}u_{\mathbf{k}}(\mathbf{r}) = u_{-\mathbf{k}}^*(\mathbf{r})$ , see Section 5.2. By using these relations in Equations (19) and (24) one can show that for band  $n$ ,

$$\mathcal{P}\mathcal{B}_n(\mathbf{k}) = \mathcal{B}_n(-\mathbf{k}) \quad (49)$$

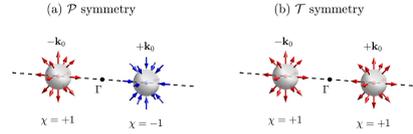
$$\mathcal{T}\mathcal{B}_n(\mathbf{k}) = -\mathcal{B}_n(-\mathbf{k}). \quad (50)$$

For a crystal with inversion symmetry  $\mathcal{P}\mathcal{B}_n(\mathbf{k}) = \mathcal{B}_n(\mathbf{k})$ , and so Equation (49) implies that  $\mathcal{B}_n(\mathbf{k}) = \mathcal{B}_n(-\mathbf{k})$ . Similarly, Equation (50) implies that for  $\mathcal{T}$  symmetry,  $\mathcal{T}\mathcal{B}_n(\mathbf{k}) = \mathcal{B}_n(\mathbf{k}) = -\mathcal{B}_n(-\mathbf{k})$ . In the latter case, the Berry curvature is an odd function of  $\mathbf{k}$ , which means that the integral of  $\mathcal{B}_n(\mathbf{k})$  over a BZ vanishes. Equations (49) and (50) also show that the combined symmetry  $\mathcal{P} \times \mathcal{T}$  maps  $\mathcal{B}_n(\mathbf{k})$  to  $-\mathcal{B}_n(\mathbf{k})$ , so if the crystal has  $\mathcal{P} \times \mathcal{T}$  symmetry, meaning that  $\mathcal{B}_n(\mathbf{k})$  is invariant under  $\mathcal{P} \times \mathcal{T}$ , then we must have  $\mathcal{B}_n(\mathbf{k}) = 0$  throughout the BZ.

The transformation properties of  $\mathcal{B}_n(\mathbf{k})$  place constraints on the location of Weyl nodes in  $\mathbf{k}$  space and their chirality. If a crystal with either inversion or time-reversal symmetry has a Weyl node at  $\mathbf{k}_0$ , then the relation  $E(\mathbf{k}) = E(-\mathbf{k})$  (see Section 5.4) requires that there must also be a Weyl node at  $-\mathbf{k}_0$ . In the case of inversion symmetry, the condition  $\mathcal{B}_n(\mathbf{k}) = \mathcal{B}_n(-\mathbf{k})$  means that the nodes at  $\pm\mathbf{k}_0$  will have opposite chirality, whereas for time-reversal symmetry,  $\mathcal{B}_n(\mathbf{k}) = -\mathcal{B}_n(-\mathbf{k})$ , and the nodes have the same chirality. These two situations are illustrated in Figure 13. If neither  $\mathcal{P}$  nor  $\mathcal{T}$  symmetry is present, then a Weyl node at  $\mathbf{k}_0$  does not in general have a partner at  $-\mathbf{k}_0$ .



**Figure 12.** Time-reversal invariant momenta (TRIM) for (a) two-dimensional, and (b) three-dimensional lattices. The TRIM are shown as red circles, and are the points in reciprocal space where  $-\mathbf{k} = \mathbf{k} + \mathbf{G}$ . Their coordinates are given in reciprocal lattice units, i.e.  $(h, k, l)$  stands for  $h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3$ . In a system with time-reversal symmetry, all states occur in degenerate pairs at the TRIM.



**Figure 13.** (a) Berry curvature field for the lower energy band in the vicinity of a pair of Weyl nodes in a crystal with inversion ( $\mathcal{P}$ ) symmetry but broken time-reversal ( $\mathcal{T}$ ) symmetry. The field satisfies the symmetry relation  $\mathcal{B}(\mathbf{k}) = \mathcal{B}(-\mathbf{k})$ . (b) The same but for a crystal with  $\mathcal{T}$  symmetry but broken  $\mathcal{P}$  symmetry, for which  $\mathcal{B}(\mathbf{k}) = -\mathcal{B}(-\mathbf{k})$ . The grey spheres represent a closed surface over which the surface integral in Equation (34) can be calculated in order to determine the Chern number,  $C_n$ . The sign of  $C_n$  for the lower energy band gives the chirality ( $\chi$ ) of the node, as indicated. The point  $\mathbf{k} = 0$  is marked  $\Gamma$ .

It follows from the fermion doubling theorem, Section 4.10, that the minimum number of Weyl nodes in a crystal with  $\mathcal{P}$  symmetry but broken  $\mathcal{T}$  symmetry is two, since in this case the chirality of the  $\pm\mathbf{k}_0$  nodes is opposite and sums to zero. On the other hand, the minimum number of nodes in a crystal with  $\mathcal{T}$  symmetry but broken  $\mathcal{P}$  symmetry is four, because the  $\pm\mathbf{k}_0$  nodes have the same chirality and so there must be at least one other pair of nodes with the opposite chirality to satisfy the theorem. The nodes with positive chirality will in general be at a different energy to the pair with negative chirality.

The properties of Weyl nodes under different combinations of  $\mathcal{P}$  and  $\mathcal{T}$  symmetry are summarised in Table 1

## 6. Two-dimensional topological band insulators and edge states

The first accounts of topological effects in electronic insulators on a lattice were based on model two-dimensional (2D) systems [11,12,28,43]. In 2D, there is only one closed surface in  $\mathbf{k}$  space and that is the entire Brillouin zone, which is closed because of the periodicity of the Bloch functions in reciprocal space (see Figure 3). Hence, each

isolated band (i.e. one that doesn't touch any other bands) has a Chern number given by

$$C_n = \frac{1}{2\pi} \int_{\text{BZ}} \mathcal{B}_n^{xy}(\mathbf{k}) d^2\mathbf{k}. \quad (51)$$

In 2D, the Berry curvature is a scalar [c.f. the 3D case, Equation (24)],

$$\mathcal{B}_n^{xy} = \frac{\partial A_n^y}{\partial k_x} - \frac{\partial A_n^x}{\partial k_y}. \quad (52)$$

A band in 2D can have a non-zero Chern number only if the Hamiltonian has broken time-reversal symmetry, because time-reversal symmetry requires  $\mathcal{B}_n^{xy}$  to be an odd function of  $\mathbf{k}$  (see Section 5.5).

### 6.1. Quantum anomalous Hall insulator

Two-dimensional electron systems with  $C_n \neq 0$  are called *Chern insulators*, or alternatively *quantum anomalous Hall (QAH) insulators*. This class of material was introduced by Haldane in 1988 [43] in order to show how the quantum Hall effect can arise on a lattice in the absence of an applied magnetic field. The key ingredients of Haldane's model are (i) a band inversion, and (ii) a type of magnetic ordering that is lattice periodic but has no net magnetic flux through the surface and creates an insulating gap. The QAH effect was first observed in thin films of  $(\text{Bi,Sb})_2\text{Te}_3$  doped with Cr to introduce ferromagnetism [44].

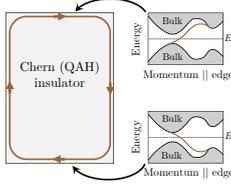
QAH insulators have a quantised Hall conductivity given by  $\sigma_{yx} = Ce^2/h$ , where  $C = \sum_{n \leq n_F} C_n$  is the total Chern number for the occupied bands. They have an edge band of states that crosses the gap between the bulk valence and conduction bands at the boundary where the gap closes and the topology changes, as described in Section 2.4. The edge states are topologically protected, which means that they are not eliminated by small perturbations, such as lattice distortions or non-magnetic impurities. Their existence is guaranteed as long as there is a change in topology on crossing the boundary, as illustrated in Figure 2. Moreover, because of the broken time-reversal symmetry, the currents associated with the edge mode flow around the boundary in one direction only, that direction being given by the gradient (group velocity) of the 1D surface band of states. For this reason, the edge modes are termed *chiral*. Note, however, that the edge modes are not, in general, fully spin-polarised.

Figure 14 is a schematic representation of a simple band of states on the edge of a QAH insulator. The band dispersion on the bottom edge is reversed relative to the

**Table 1.** Effect of space inversion ( $\mathcal{P}$ ) and time-reversal ( $\mathcal{T}$ ) symmetry on some properties of electronic bands and Weyl points (WPs).

Property	Type of symmetry present			
	$\mathcal{P}$	$\mathcal{T}$	$\mathcal{P}$ and $\mathcal{T}$	$\mathcal{P} \times \mathcal{T}^a$
Band dispersion	$E_n(\mathbf{k}) = E_n(-\mathbf{k})$	$E_n(\mathbf{k}) = E_n(-\mathbf{k})$	$E_n(\mathbf{k}) = E_n(-\mathbf{k})$	$E_n(\mathbf{k}) \neq E_n(-\mathbf{k})$
Berry curvature	$\mathcal{B}_n(\mathbf{k}) = \mathcal{B}_n(-\mathbf{k})$	$\mathcal{B}_n(\mathbf{k}) = -\mathcal{B}_n(-\mathbf{k})$	$\mathcal{B}_n(\mathbf{k}) = 0$	$\mathcal{B}_n(\mathbf{k}) = 0$
Chirality of WPs	$\chi(\mathbf{k}_0) = -\chi(-\mathbf{k}_0)$	$\chi(\mathbf{k}_0) = \chi(-\mathbf{k}_0)$	no WPs	no WPs
Min. no. of WPs	2	4	–	–

<sup>a</sup> Both  $\mathcal{P}$  and  $\mathcal{T}$  symmetries are broken, but their combination  $\mathcal{P} \times \mathcal{T}$  (or  $\mathcal{T} \times \mathcal{P}$ ) is present.



**Figure 14.** Left: chiral edge state on a Chern (QAH) insulator. Right: energy spectrum as a function of momentum along the top and bottom edges. The shaded regions are the projections of the bulk valence and conduction bands. A single edge band connects the bulk valence and conduction bands with a positive (negative) Fermi velocity on the top (bottom) edges.

top edge, corresponding to the opposite velocity. The displayed band crosses the Fermi energy  $E_F$  only once, but this needn't be the case. An edge band could in principle cross  $E_F$  multiple times, generating left- and right-moving states at  $E_F$ . In general, the number of times an edge band crosses the Fermi level is governed by the *bulk–boundary correspondence*, which can be expressed as

$$N_R - N_L = \Delta C_n, \quad (53)$$

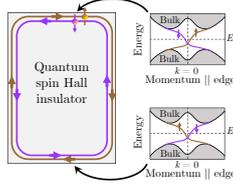
where  $N_R$  and  $N_L$  are the number of right- and left-moving modes, and  $\Delta C_n$  is the change in Chern number across the boundary. The band shown in Figure 14, therefore, corresponds to a change in  $C_n$  from 1 to 0.

## 6.2. Quantum spin Hall insulator

In the QAH insulator, the topological effects require magnetic order. In 2005, Kane and Mele introduced an important model which showed that topological effects can be present even without magnetic order [11,29]. Their model is similar to Haldane's but includes spin-orbit coupling (SOC) in place of magnetic order. The effect of SOC is to split the bands into pairs related by time-reversal symmetry, and to create a gap throughout the BZ. Remarkably, when the parameters of the model are tuned to produce a band inversion, pairs of spin-polarised, or *helical*, edge states appear which cross the gap between the bulk valence and conduction bands, see Figure 15. By *helical* we mean that the up and down spin states travel in opposite directions around the boundary. The associated currents are unaffected by non-magnetic perturbations because there is no conduction channel for electrons whose direction of motion is reversed but whose spin is unchanged,

The helical surface bands cross at the TRIM of the surface BZ, and the degeneracy is protected by time-reversal symmetry (Section 5.4). This ensures a metallic surface. If  $\mathcal{T}$ -invariance is broken, e.g. by magnetic order, then gaps open at the TRIM and the surface will become insulating if the gap is large enough to contain  $E_F$ .

The type of system described by the Kane–Mele model is termed a *quantum spin Hall (QSH) insulator*, or simply a *2D topological insulator*. From a topological point of view, the QSH insulator is interesting because the presence of time-reversal symmetry implies  $C_n = 0$  (Section 5.5), and so there must be another topological invariant that



**Figure 15.** Left: helical (spin-polarised) edge state on a Quantum spin Hall (QSH) insulator. Right: energy spectrum as a function of momentum along the top and bottom edges. The shaded regions are the projections of the bulk valence and conduction bands. A pair of edge band connects the bulk valence and conduction bands with opposite Fermi velocities. The band crossing at the TRIM ( $k = 0$ ) is protected by time-reversal symmetry.

distinguishes a QSH insulator from a normal band insulator.

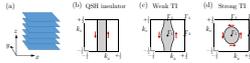
To see how this comes about, let us label the SOC-split bands  $\uparrow$  and  $\downarrow$ . Because time-reversal transforms  $\uparrow$  to  $\downarrow$ , the states in these bands are not time-reversal invariant and so have separate Chern numbers ( $C_{n\uparrow}, C_{n\downarrow}$ ) which do not have to be zero providing  $C_n = C_{n\uparrow} + C_{n\downarrow} = 0$ . It is found that when  $C_{n\uparrow}$  and  $C_{n\downarrow}$  are both odd integers the system is a QSH insulator, and when they are both even integers it is a normal insulator. This type of topology is characterised by a  $\mathbb{Z}_2$  topological invariant [29], where  $\mathbb{Z}_2$  represents the group of integers (0, 1) under addition modulo 2. It can be shown that the  $\mathbb{Z}_2$  index (0 or 1) measures the number of times (mod 2) an edge band dispersion crosses the Fermi level in half of the 1D surface BZ, i.e.  $0 \leq k \leq \pi/a$ , so it provides a simple test to distinguish an insulator which is topologically trivial ( $\mathbb{Z}_2$ -even) from one that is topologically non-trivial ( $\mathbb{Z}_2$ -odd). The edge bands shown in Figure 15 cross  $E_F$  once in the half BZ, corresponding to a topologically non-trivial insulator.

## 7. 3D topological insulators

Three-dimensional (3D) topological insulators (TIs) are insulators in the bulk but have conducting states on some or all of their surfaces (Section 2.4). They are closely related to the 2D insulators described in Section 6, in which time-reversal symmetry is either present (QSH insulator) or broken (QAH/Chern insulator).

### 7.1. Time-reversal symmetric TIs

Consider a 3D material built out of 2D QSH insulating layers stacked on top of one another in the  $z$  direction, like a deck of cards, as shown in Figure 16(a). An isolated 2D QSH layer has helical edge states as described in Section 6.2. In the simplest case (Figure 15), the bulk band gap is traversed by a single pair of Kramers-degenerate edge states,  $E(-k) = E(k)$ , which cross the Fermi level ( $E_F$ ) once in the half BZ  $k > 0$ . If the coupling between adjacent QSH layers in the stack is zero, then there is no dispersion in the inter-layer direction  $k_z$ , and the Fermi energy contour that separates the filled and empty surface states is a pair of straight lines running parallel to  $k_z$ , as shown in Figure 16(b).



**Figure 16.** (a) Stack of QSH insulators. (b)–(d) Surface Brillouin zone and Fermi energy contour for 3D TIs with increasing inter-layer couplings. The shaded regions indicate filled states, the red arrows represent the spin direction, and  $\Gamma_{1,2,3,4}$  are the TRIM. (b) zero coupling; (c) weak TI — the Fermi contour encloses two TRIM ( $\nu_0 = 0$ ); (d) strong TI — the Fermi contour encloses one TRIM ( $\nu_0 = 1$ ).

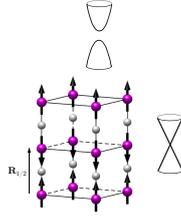
Now suppose that there is a small but non-zero inter-layer coupling. The surface bands may develop a weak dispersion along  $k_z$ , giving a slightly curved Fermi contour, Figure 16(c), which remains open, i.e. the contour connects to itself on opposite boundaries of the BZ. Finally, with a stronger inter-layer coupling we might obtain a closed Fermi contour like that sketched in Figure 16(d).

Figures 16(c) and (d) are illustrative of what are known as *weak* and *strong* TIs. The distinction can be understood if we consider how the bands connect the time-reversal invariant momenta (TRIM) in the BZ. Recall that there are four TRIM in the 2D BZ, labelled  $\Gamma_1$  to  $\Gamma_4$  in Figure 16, and that surface states, if they exist, must be Kramers degenerate at these points. Elsewhere, the degeneracy is lifted by spin–orbit coupling, so that the TRIM are the nodes of a 2D linear band crossing like the Dirac dispersion in graphene (see Section 8.3).

In Figure 16(c), a surface band crosses  $E_F$  once along the path from  $\Gamma_1$  to  $\Gamma_2$  ( $k_x$  direction) but does not cross  $E_F$  between  $\Gamma_1$  and  $\Gamma_3$  ( $k_z$  direction). In Figure 16(d), a band crosses  $E_F$  once along both paths. Therefore, situations (c) and (d) are not smoothly connected and represent distinct topological phases. In particular, in (c) the surface perpendicular to  $k_z$  is gapped and insulating, whereas in (d) it is metallic. The surfaces perpendicular to  $k_x$  and  $k_y$  are metallic in both cases.

More precisely, a weak TI is one in which the Fermi contour encloses an even number of TRIM, whereas in a strong TI it encloses an odd number. This classification is represented by a  $\mathbb{Z}_2$  topological index  $\nu_0$ , such that a weak TI has  $\nu_0 = 0$  (even integer, modulo 2), and a strong TI has  $\nu_0 = 1$  (odd integer, modulo 2). The value of  $\nu_0$  depends on symmetry properties of the Bloch functions at all eight TRIM of the bulk BZ. More generally, the Fermi contour may intersect the line joining any pair of TRIM an odd or an even number of times, and in order to describe all cases three more  $\mathbb{Z}_2$  indices,  $\nu_1$ ,  $\nu_2$  and  $\nu_3$  are required [30–32].

A magnetic field applied perpendicular to the conducting surface of a 3D TI will generate Landau levels, as for a conventional 2D metal, resulting in quantum oscillations and an associated quantum Hall effect. The Hall conductivity  $\sigma_{xy}$  is quantised in steps of  $e^2/h$ , as usual, but the first step is at  $e^2/2h$ . This half-quantisation is a result of the  $\pi$  Berry phase for adiabatic transport around a Dirac point in 2D (Section 4.8). It is important to appreciate, however, that the surface states of a 3D TI are fundamentally different from a topologically trivial 2D metal. In the latter, there are up and down spins at every point on the Fermi contour, whereas in a 3D TI, time-reversal symmetry requires that  $+\mathbf{k}$  and  $-\mathbf{k}$  states have opposite spin, so the spin must rotate with  $\mathbf{k}$  around the Fermi contour, Figure 16(c) and (d). This helical spin texture has been observed in ARPES experiments [45].



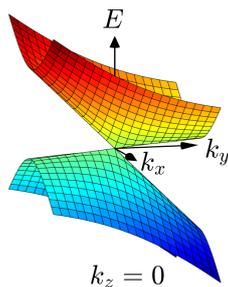
**Figure 17.** An antiferromagnetic spin structure which has  $S = \mathcal{T} \times \mathbf{R}_{1/2}$  symmetry, where  $\mathbf{R}_{1/2}$  is a half-lattice translation in the vertical direction which connects sites with opposite spin. A strong topological insulator with this AFM structure is an example of an axion insulator. As indicated, the Dirac cones on the top and bottom surfaces of the crystal are gapped, whereas those on the sides are gapless.

## 7.2. 3D magnetic TIs

Much of the research on 3D TIs has focussed on time-reversal invariant systems [46, 47]. Magnetic TIs, which combine non-trivial band topology and magnetic order, are expected to exhibit some features which are similar and some which are distinct from their non-magnetic counterparts [24, 48–51]. Magnetic TIs offer the added possibility of magnetic control of their topological states. The principal experimental techniques for determining magnetic structures are neutron diffraction and resonant inelastic x-ray scattering (see, for example, Refs. [52, 53]).

A simple way to create a 3D magnetic TI is to introduce ferromagnetism into a 3D  $\mathcal{T}$ -invariant TI, either by doping or a ferromagnetic coating layer. The magnetization  $\mathbf{M}$  couples to the surface states via an exchange interaction, and opens a gap in the Dirac surface spectrum of the host TI on the boundary surfaces perpendicular to  $\mathbf{M}$ . The sign of the Chern number  $C$  ( $= \pm 1$ ) depends on the direction of  $\mathbf{M}$ . On entering the material across the side surfaces parallel to  $\mathbf{M}$ ,  $C$  jumps from 0 to  $\pm 1$ , so a dissipationless chiral edge mode appears in accordance with the bulk–boundary correspondence, just as for the 2D QAH insulator. If  $E_F$  lies in the gap then  $\sigma_{xy}$  is half-quantised (nb: in a measurement that probes both top and bottom sample surfaces, the observed  $\sigma_{xy}$  will be  $\pm e^2/h$ , i.e.  $\pm e^2/2h$  from each surface).

One problem with doping is that it leads to inhomogeneous electronic properties. Recent research, therefore, has focussed on intrinsic (stoichiometric and well-ordered) magnetic compounds. Pioneering work reported in 2010 [54] showed that a strong TI with broken  $\mathcal{T}$  symmetry due to a type of antiferromagnetic (AFM) order is topological in a similar sense to  $\mathcal{T}$ -invariant TIs, but has some surfaces which are gapped and some which are not gapped. The AFM order considered in Ref. [54] has twice the periodicity of the crystal lattice but preserves the combined symmetry  $S = \mathcal{T} \times \mathbf{R}_{1/2}$ , where  $\mathbf{R}_{1/2}$  is a half-lattice translation which connects sites with opposite spin. A simple example is shown in Figure 17. Such an AFM crystal has two types of surfaces, those in which all the spins are aligned ferromagnetically (the top and bottom in Figure 17), and those in which there is AFM order within the surface (the sides). The latter has the  $S$  symmetry, which preserves the Kramers degeneracy of the surface states and preserves



**Figure 18.** Dispersion surface in the vicinity of an anisotropic and tilted band crossing at  $\mathbf{k} = 0$ .

an odd number of gapless Dirac cones, as for a strong TI. Ferromagnetic surfaces, on the other hand, break  $S$  symmetry because the  $\mathbf{R}_{1/2}$  translation is terminated at the surface, and so these surfaces are gapped and exhibit a half-integer  $\sigma_{xy}$ .

Magnetic TI materials whose bulk symmetries include  $\mathcal{T}$  combined with another crystalline symmetry to give topological properties similar to those just described are termed *axion insulators* [48,49]. The name comes from a particular type of quantised magnetoelectric coupling associated with the electrodynamics of TIs. The coupling, which adds a term  $\mathcal{L}_\theta \propto \theta \mathbf{E} \cdot \mathbf{B}$  to the electromagnetic Lagrangian of the material, where  $\theta = 0$  (trivial insulator) or  $\pm\pi$  (topological insulator), is analogous to the axion field introduced in high-energy physics to explain the absence of charge–parity violation in the strong interaction between quarks [55]. The first intrinsic AFM axion insulator material to be identified was  $\text{MnBi}_2\text{Te}_4$  [56].

## 8. Nodal semimetals

### 8.1. Weyl semimetals

Recall that a Weyl point (or node) is a touching point between two singly degenerate bands in a 3D crystal (Section 4.8). In Section 5.5, we showed that the existence of Weyl nodes requires either broken  $\mathcal{P}$  symmetry, or broken  $\mathcal{T}$  symmetry, or both. In Section 4.8, we analysed a simple model Hamiltonian that describes a Weyl node with a linear isotropic band dispersion. We assigned a chirality  $\chi = \pm 1$  to the Weyl point depending on whether the lower band is a source or sink of Berry curvature, and the results in Section 4.9 show that a Weyl point is topological because the Chern number  $C$  calculated for a closed surface surrounding the node is non-zero. The only way to eliminate Weyl points is for pairs with opposite chirality to annihilate one another in order to preserve zero net chirality, as required by the Nielsen-Ninomiya theorem, Equation (41). Annihilation requires the pair to move to the same  $\mathbf{k}$ , and so Weyl points are stable against small changes in the Hamiltonian as long as they remain separated in  $\mathbf{k}$  space.

In 3D materials, the dispersion around a touching point between two bands is often anisotropic and tilted, as illustrated in Figure 18. We can describe such cases with a more general two-band model whose effective Hamiltonian can be represented by the

two-by-two Hermitian matrix

$$\mathcal{H}(\mathbf{k}) = \begin{pmatrix} f_0 + f_3 & f_1 - if_2 \\ f_1 + if_2 & f_0 - f_3 \end{pmatrix} \quad (54)$$

$$= f_0\sigma_0 + f_1\sigma_1 + f_2\sigma_2 + f_3\sigma_3, \quad (55)$$

where  $f_0$  to  $f_3$  are real functions of  $\mathbf{k}$ . On the second line,  $\sigma_0$  is the  $2 \times 2$  identity matrix, and  $\sigma_1$ ,  $\sigma_2$  and  $\sigma_3$  are the Pauli matrices. The eigenvalues of (54) are  $E^\pm = f_0 \pm \sqrt{f_1^2 + f_2^2 + f_3^2}$ , so the band degeneracy occurs when  $f_1 = f_2 = f_3 = 0$ .

If, for example, we put  $(f_1, f_2, f_3) = \hbar(v_x k_x, v_y k_y, v_z k_z)$ , then we obtain an anisotropic linear dispersion, with speeds  $(v_x, v_y, v_z)$  along a set of orthogonal axes  $(k_x, k_y, k_z)$ . If, in addition, we put  $f_0 = \hbar\mathbf{v}_0 \cdot \mathbf{k}$ , then we obtain a tilt in the direction of  $\mathbf{v}_0$ . The linear isotropic Hamiltonian, Equation (28), is recovered when  $(f_0, f_1, f_2, f_3) = \hbar v(0, k_x, k_y, k_z)$ . An anisotropic and tilted linear dispersion can always be brought into diagonal isotropic form by an orthogonal transformation together with a scaling of the  $k_i$  axes. Hence, the topology is the same as the isotropic case, and the Berry curvature derives from monopoles of strength  $\pm\frac{1}{2}$  in the diagonal/scaled basis.

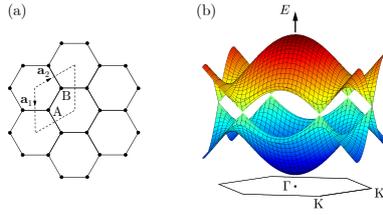
Weyl points are not restricted to any particular energy, but if they are located close to the Fermi energy then they can influence the excitations which are responsible for the low-energy properties of the material. Consider once again the case of a linear isotropic band crossing at  $\mathbf{k} = \mathbf{k}_0$  with the zero of energy at the Weyl node, and put  $\mathbf{q} = \mathbf{k} - \mathbf{k}_0$ . The second form of the effective Hamiltonian, given in Equation (55), may then be written

$$\mathcal{H}(\mathbf{q}) = \pm\hbar v \mathbf{q} \cdot \boldsymbol{\sigma}, \quad (56)$$

where  $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ , and the  $\pm$  sign describes the two different chiralities of Weyl point.

Equation (56) takes exactly the same form, up to the replacement of  $v$  by the speed of light  $c$ , as the Hamiltonian proposed by Hermann Weyl in 1929 for massless relativistic fermions [57,58]. Therefore, in a crystal whose electronic states are described by (56) the low energy quasiparticle excitations mimic Weyl fermions, except they travel at speeds  $v \ll c$ . We call them *emergent Weyl fermions*. It is important to note that in the case of elementary particles,  $\boldsymbol{\sigma}$  represents the spin of the particle, whereas for the emergent quasiparticles  $\boldsymbol{\sigma}$  is a *pseudospin* which acts in the two-band sub-space of Bloch functions and is not usually a pure spin degree of freedom. Equation (56) shows that the pseudospin is coupled to the reduced wavevector  $\mathbf{q}$ , a property known as *helicity*, and the sign determines whether the pseudospin is parallel or antiparallel to  $\mathbf{q}$ .

If the Weyl node is located right at the Fermi energy  $E_F$ , then we have what is known as a *Weyl semimetal* (WSM). A semimetal is characterised by a vanishing electronic density of states at  $E = E_F$  but non-zero density of states for  $E > E_F$  and  $E < E_F$ . If the node is just above or just below  $E_F$  then the material is termed a *Weyl metal*. A Weyl metal retains many of the characteristics of a WSM providing  $E_F$  is located on the linear part of the dispersion, and providing there are no trivial bands at  $E_F$  to mask the response of the Weyl fermions.



**Figure 19.** (a) Crystal structure of graphene, showing the primitive lattice vectors  $\mathbf{a}_1$  and  $\mathbf{a}_2$ . A and B label the two carbon atoms in the primitive unit cell. (b) Conduction and valence bands of graphene assuming a two-band tight-binding model with nearest-neighbour coupling  $t$ . The BZ is shown below. The band dispersion is given by  $E(\mathbf{k}) = E_F \pm t\{1 + 4 \cos(3k_x a/2) \cos(\sqrt{3}k_y a/2) + 4 \cos^2(\sqrt{3}k_y a/2)\}^{\frac{1}{2}}$ , with the Fermi level  $E_F$  at the nodes.

### 8.2. Dirac semimetals

A Dirac point forms at the crossing of two bands which are doubly degenerate, rather than singly degenerate as in the case of a Weyl point. A system in which the Dirac node occurs at the Fermi level is termed a *Dirac semimetal* (DSM). The conical band dispersion of an ideal DSM is the same as that of massless (relativistic) fermions described by the Dirac equation, hence the name.

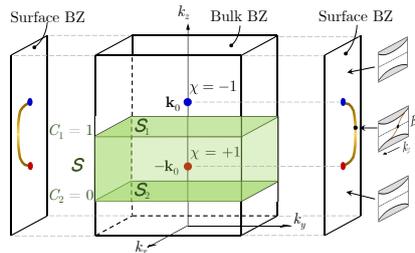
Dirac points require the combined  $\mathcal{P} \times \mathcal{T}$  symmetry in order to enforce double degeneracy throughout the BZ (Section 5.4), so  $\mathcal{P}$  and  $\mathcal{T}$  are either both present or both absent. The latter case makes it possible for Dirac points to exist in magnetic materials.

As explained in Section 5.5,  $\mathcal{P} \times \mathcal{T}$  symmetry means that the Berry curvature  $\mathcal{B}_n(\mathbf{k})$  is zero throughout the BZ (except at the degeneracy point itself, where  $\mathcal{B}_n(\mathbf{k})$  is undefined), and so the Chern number of any closed surface surrounding a Dirac point is zero. Hence, a Dirac point can be thought of as the merger of two Weyl nodes of opposite chirality. Nevertheless, Dirac points are topological because they involve band inversion, and there are other topological invariants which characterise them.

When a pair of doubly degenerate bands cross the states can mix, forming a gap in the spectrum at the node. In general, therefore, Dirac points do not form where there is an accidental band crossing at an arbitrary  $\mathbf{k}$ . Additional symmetries are required to prevent gap formation, for example rotational symmetry axes or high-symmetry points in the BZ, especially TRIM. As a result, Dirac nodes, unlike Weyl nodes, are sensitive to perturbations which remove the symmetry protection of the degeneracy.

### 8.3. Graphene

To a very good approximation, graphene realises emergent Dirac fermions in 2D, see Figure 19. The honeycomb arrangement of carbon atoms has  $\mathcal{P}$  and  $\mathcal{T}$  symmetry, which enforces Kramers degeneracy of the bands. The valence and conduction bands touch at the K and K' points on the BZ boundary, and near to these points the electronic dispersion is linear and isotropic, with a Fermi velocity  $v_F \simeq 10^6 \text{ ms}^{-1}$ . The touching points are not TRIM ( $\mathbf{K}$  and  $-\mathbf{K}$  are not related by a reciprocal lattice vector), but three-fold rotational symmetry around K and K' fixes the nodes at these points in the BZ. There is in reality a tiny gap ( $\sim 10 \mu\text{eV}$ ) due to spin-orbit coupling



**Figure 20.** Surface Fermi arcs on a Weyl semimetal. Two Weyl points with opposite chirality  $\chi = \pm 1$  (red and blue circles) are located at wavevectors  $\mp \mathbf{k}_0$ , respectively, in the bulk BZ. The closed surface  $\mathcal{S}$  is defined by the pale green planes, and includes the two slice planes  $\mathcal{S}_1$  and  $\mathcal{S}_2$  together with the side boundaries of the BZ. The Chern numbers of  $\mathcal{S}_{1,2}$  are  $C_{1,2}$ . Fermi arcs are shown on the 2D BZs for the right and left surfaces. On the right side, the energy spectrum is plotted as a function of  $k_x$  at three different  $k_z$ . When  $-k_0 < k_z < k_0$ , a single surface band connects the bulk valence and conduction bands (shaded). The Fermi arc is formed from the locus of  $k_x$  points where the surface band crosses  $E_F$ .

which turns graphene formally into a quantum spin Hall insulator [11], but for most practical purposes the gap can be neglected, and indeed graphene exhibits many of the phenomena predicted by quantum electrodynamics for massless fermions [59].

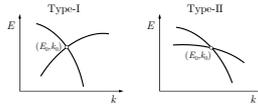
Because of  $\mathcal{P}$  and  $\mathcal{T}$  symmetry, the Berry curvature  $\mathcal{B}^{xy}(\mathbf{k})$  of pristine graphene is zero everywhere (except at the band touching points, where it is undefined). Nevertheless, according to Equation (33), the Berry phase for one circuit around a K or K' point is  $\pm\pi$ , and this has been verified experimentally [10]. Therefore, the Bloch states of graphene have non-trivial local geometrical features, but the global topology is trivial. If  $\mathcal{P}$  symmetry is broken with a staggered sublattice potential, making the two carbon sites in the unit cell inequivalent, then a gap opens at the K and K' points and a non-zero  $\mathcal{B}^{xy}(\mathbf{k})$  emerges, though the Chern number remains zero because as long as  $\mathcal{T}$  symmetry is retained  $\mathcal{B}^{xy}(\mathbf{k})$  is an odd function of  $\mathbf{k}$ .

#### 8.4. Surface states and surface Fermi arcs

We have seen that changes in bulk topology are accompanied by surface or interface states, a property known as the *bulk-boundary correspondence*. The surface states associated with Weyl and Dirac points take the form of Fermi arcs in the 2D surface BZ, and have been extensively studied by angle-resolved photoemission spectroscopy [22].

In order to understand the formation of Fermi arcs in more detail, we shall consider the particular case of a  $\mathcal{T}$ -broken type-I Weyl semimetal with a single pair of Weyl nodes located at wavevectors  $-\mathbf{k}_0$  and  $+\mathbf{k}_0$  with chiralities  $\chi = +1$  and  $-1$ , respectively, as shown in Figure 20. We shall assume that the two bands that touch at the Weyl points are separated by an energy gap everywhere else in the BZ, and that the Fermi energy lies at the Weyl nodes. For simplicity, we also assume that the crystal axes are orthogonal, and take  $\mathbf{k}_0 = (0, 0, k_0)$ . We showed in Section 4.8 that there is a Chern number associated with a closed surface  $\mathcal{S}$  surrounding a Weyl node, and that the Chern number of the lower band is equal to  $\chi$ .

Now consider  $\mathcal{S}$  as shown in Figure 20. It is formed by a pair of planes  $\mathcal{S}_1$  and



**Figure 21.** Band dispersion along one component of  $\mathbf{k}$  in the vicinity of Type-I and Type-II nodal points.

$\mathcal{S}_2$  perpendicular to the  $k_z$  axis, together with the side boundaries of the BZ. In the periodic gauge, any BZ slice plane is itself a closed surface (Section 3.2, Figure 3), and so has a Chern number in accordance with the Chern theorem, Equation (34). The Chern numbers  $C_{1,2}$  of the  $\mathcal{S}_{1,2}$  slices are defined with respect to a normal vector that points along  $k_z$ .

When  $\mathcal{S}$  encloses the Weyl point at  $-\mathbf{k}_0$ , the Chern number associated with  $\mathcal{S}$  for the lower band is  $C = 1$ . The side boundaries of the BZ do not contribute to  $C$  because the Berry curvature is periodic in the Brillouin zone and the outward normal of one side of the BZ is the inward normal of the opposite side. Hence,  $C = C_1 - C_2$ . If the bands are topologically trivial for  $k_z < -k_0$ , then  $C_2 = 0$  and  $C_1 = C = 1$ .

Since the slice  $\mathcal{S}_1$  has a non-zero Chern number, it is equivalent to a 2D Chern (QAH) insulator, which in the simplest case has a single edge band crossing  $E_F$  (Section 6.1). If the position of  $\mathcal{S}_1$  is varied,  $C_1$  remains constant as long as  $\mathcal{S}$  encloses only the  $-\mathbf{k}_0$  Weyl point. Hence,  $C_1 = 1$  for  $-k_0 < k_z < k_0$ , and  $C_1 = 0$  otherwise. The region between the two Weyl points, therefore, can be viewed as a stack of such Chern insulators, each of which has an edge state that crosses  $E_F$ .

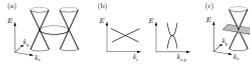
The detailed dispersion of the edge bands depends on the surface Hamiltonian and boundary conditions. Assuming these are known, one can construct the surface Fermi arcs as follows. Pick a  $k_z$ , and draw a dot on the surface BZ where the edge band for the slice  $\mathcal{S}(k_z)$  crosses  $E_F$ . Repeat for a series of  $k_z$  in the range  $-k_0 < k_z < k_0$ , and join the dots together to form the arc. The Fermi arc begins and ends on the projection of the Weyl points onto the surface BZ. This is because, firstly, the Weyl points are located at  $E_F$  and must therefore lie on the Fermi arc, and secondly, there can be no surface states for  $k_z < -k_0$  or  $k_z > k_0$  where the bands are topologically trivial. Note that the electronic dispersion on the bottom edge of a Chern insulator is the reverse of that on the top edge (Figure 14), with the result that the Fermi arcs on opposite surfaces curve in opposite directions (Figure 20).

### 8.5. Other characteristics of band crossings

Before concluding, I want to mention several other differences between the relativistic fermions envisaged in particle physics<sup>10</sup> and the emergent relativistic fermions which live in crystalline solids. These differences arise because electrons in crystal lattices are not constrained by the full Poincaré symmetry of free space.

First, according to Special Relativity, massless particles have a property known as *Lorentz invariance*, which means that they propagate at the speed of light in all

<sup>10</sup>Weyl fermions, incidentally, have not been observed in high-energy physics. Neutrinos were initially believed to be Weyl fermions, but the discovery of neutrino mass put an end to that possibility.



**Figure 22.** Schematic diagrams of (a) a nodal line, (b) a quadratic Weyl point, with a linear dispersion along  $k_z$  and quadratic dispersions along  $k_x$  and  $k_y$ , and (c) a three-fold band crossing.

directions. In real band structures, as already mentioned, the Dirac cone is usually tilted and anisotropic (graphene is a notable exception). If the cone is tilted, the group velocities in the directions  $+\mathbf{q}$  and  $-\mathbf{q}$  are different (see Figure 18). Hence, Lorentz invariance can be broken.

Second, Weyl and Dirac points can be divided into two classes, called type-I and type-II, depending on the degree of tilt, as illustrated in Figure 21. If the tilt is small, the density of states vanishes at the node, and constant-energy surfaces close to the node are ellipsoids. This is a type-I point. If, however, there exists a direction  $\mathbf{q}$  in which the tilt is sufficiently large that the sign of the gradient of both bands  $E^\pm(\mathbf{q})$  is the same, either both positive or both negative, then it is defined as a type-II point. Constant-energy surfaces near a type-II point are open, and as a result, the density of states does not vanish at the node. Type-I and type-II topological semimetals have different transport and thermodynamic properties [60].

Third, under certain circumstances two bands can touch along a line or closed loop in  $\mathbf{k}$  space, rather than at a point, as sketched in Figure 22(a). The curve along which the bands touch is called a nodal line, and if the nodal line energy coincides with  $E_F$ , the system is termed a *nodal line semimetal*. Both twofold (Weyl) and fourfold (Dirac) nodal lines are possible. There are several different symmetry combinations that can protect a nodal line, and experimental signatures of nodal line semimetals have been predicted [61,62].

Fourth, it is possible to create twofold degenerate band crossings which have a quadratic or cubic dispersion in two directions of  $\mathbf{k}$  space, and a linear dispersion in the third direction [63], Figure 22(b). These *multi-Weyl nodes* can occur on lines of symmetry with either 4-fold or 6-fold rotational symmetry in  $\mathbf{k}$  space. Closed surfaces surrounding quadratic and cubic Weyl nodes have Chern numbers  $C = \pm 2$  and  $\pm 3$ , respectively.

Finally, in addition to the two- or four-band crossings that define Weyl and Dirac points, three-, six- and eight-band crossings are also possible [64], as shown in Figure 22(c). The excitations associated with these points correspond to exotic fermionic quasiparticles with spin representations that differ from one-half and have no counterpart in high-energy physics.

## 9. Concluding remarks

This review presents the basic principles behind the new topological way of thinking about electronic structure in crystalline solids. The central ideas of Berry-ology — the Berry phase, connection and curvature — are outlined from their quantum-mechanical origins, and applied to the Bloch states which describe non-interacting

electrons in solids. This leads to a topological classification of bulk states in terms of the Chern number and  $\mathbb{Z}_2$  topological invariants, and hence to the bulk–boundary correspondence, the wonderful notion that at the boundary between two materials whose electronic states are described by different topological invariants there must inevitably exist conducting surface or edge states.

Due to the constraints of space and time, many important and contemporary topics have been omitted. The intrinsic anomalous Hall and Nernst effects, the chiral anomaly in Weyl semimetals, optically induced transport effects, as well as topological superconductivity and proposals for emergent Majorana fermions, are all conspicuous by their absence. The Berry phase theory of electrical polarisation and orbital magnetism is also missing, as is any meaningful discussion of magnetoelectric effects.

The field of topological materials is a vibrant one, and continues to be fuelled by a wealth of predictions of new quantum states and spin transport phenomena with potential applications in areas such as spintronics, topological electronics and quantum technology. Currently, many of the predictions are as-yet unverified experimentally, and a major effort is underway to identify and synthesise suitable compounds that realise the predicted characteristics. In recent years, this effort has been considerably assisted by the development of publicly available and searchable databases of topological materials [65–68].

## **Acknowledgements**

I am grateful to the Laboratory for Neutron Scattering and Imaging at the Paul Scherrer Institute, Switzerland, and the LINXS Institute of Advanced Neutron and X-ray Science, Sweden, for their hospitality and support during extended visits in 2023. I would like to thank Jian-Rui Soh, D. Prabhakaran and Yanfeng Guo for discussions and fruitful collaborations on magnetic topological materials.

## **Disclosure statement**

No potential conflict of interest was reported by the author.

## **Funding**

Support is acknowledged from the Oxford–ShanghaiTech Collaboration Project and from Oriel College, Oxford.

## **Notes on contributor**

*Andrew Boothroyd* is a Professor of Physics at Oxford University, and a Tutorial Fellow of Oriel College. Currently he is serving as Associate Head of Department. After completing his undergraduate and graduate degrees at the University of Cambridge, Andrew joined the University of Warwick. In 1992, he moved to Oxford University where he has served as Head of Condensed Matter Physics, Associate Head of Department, and as Vice-Provost of Oriel College. His research exploits neutron and synchrotron x-ray scattering methods to study crystalline, electronic and magnetic

order and dynamics in quantum materials. Recent work has investigated the interplay between magnetic structure and electronic topology in magnetic topological semimetals. In 2011, Andrew received the Institute of Physics' Brian Pippard Prize, and in 2017 he was recognised as an Outstanding Referee by the American Physical Society.

## ORCID

*Andrew Boothroyd*: <https://orcid.org/0000-0002-3575-7471>

## References

- [1] Kosterlitz JM. Nobel Lecture: Topological defects and phase transitions. *Rev Mod Phys.* 2017;89:040501. Available from: <https://doi.org/10.1103/RevModPhys.89.040501>.
- [2] Haldane FDM. Nobel Lecture: Topological quantum matter. *Rev Mod Phys.* 2017; 89:040502. Available from: <https://doi.org/10.1103/RevModPhys.89.040502>.
- [3] Kane CL, Moore JE. Topological insulators. *Physics World.* 2011;24:32–36. Available from: <https://doi.org/10.1088/2058-7058/24/02/36>.
- [4] Pachos JK. Manifestations of topological effects in graphene. *Contemp Phys.* 2009;50:375–389. Available from: <https://doi.org/10.1080/00107510802650507>.
- [5] Lancaster T. Skyrmions in magnetic materials. *Contemp Phys.* 2019;60:246–261. Available from: <https://doi.org/10.1080/00107514.2019.1699352>.
- [6] Aguado R, Kouwenhoven LP. Majorana qubits for topological computing. *Physics Today.* 2020;73:44–50. Available from: <https://doi.org/10.1063/PT.3.4499>.
- [7] Segev M, Bandres MA. Topological photonics: Where do we go from here? *Nanophotonics.* 2021;10:425–434. Available from: <https://doi.org/10.1515/nanoph-2020-0441>.
- [8] Gu Y, Zhao J, Johnson JA. A (macro)molecular-level understanding of polymer network topology. *Trends Chem.* 2019;1:318–334. Available from: <https://doi.org/10.1016/j.trechm.2019.02.017>.
- [9] Ardaseva A, Doostmohammadi A. Topological defects in biological matter. *Nat Rev Phys.* 2022;4:354–356. Available from: <https://doi.org/10.1038/s42254-022-00469-9>.
- [10] Novoselov KS, Geim AK, Morosov SV, et al. Two-dimensional gas of massless Dirac fermions in graphene. *Nature.* 2005;438:197–200. Available from: <https://doi.org/10.1038/nature04233>.
- [11] Kane CL, Mele EJ. Quantum spin Hall effect in graphene. *Phys Rev Lett.* 2005;95:226801. Available from: <https://doi.org/10.1103/PhysRevLett.95.226801>.
- [12] Bernevig BA, Hughes TL, Zhang SC. Quantum spin Hall effect and topological phase transition in HgTe quantum wells. *Science.* 2006;314:1757–1761. Available from: <https://doi.org/10.1126/science.1133734>.
- [13] König M, Wiedmann S, Brüne C, et al. Quantum spin Hall insulator state in HgTe quantum wells. *Science.* 2007;318:766–770. Available from: <https://doi.org/10.1126/science.1148047>.
- [14] Moessner R, Moore JE. *Topological phases of matter.* Cambridge, U.K.: Cambridge University Press; 2021.
- [15] Vanderbilt D. *Berry phases in electronic structure theory.* Cambridge, U.K.: Cambridge University Press; 2018.
- [16] Bernevig BA. *Topological insulators and topological superconductors.* Princeton, NJ: Princeton University Press; 2013.
- [17] Kane CL. Topological band theory and the  $\mathbb{Z}_2$  invariant. In: Marcel F, Molenkamp L, editors. *Topological insulators.* Amsterdam: Elsevier; 2013. p. 3–34. Available from: <https://doi.org/10.1016/B978-0-444-63314-9.00001-9>.

- [18] Cayssol J, Fuchs JN. Topological and geometrical aspects of band theory. *J Phys Mater.* 2021;4:034007. Available from: <https://doi.org/10.1088/2515-7639/abf0b5>.
- [19] Armitage NP, Mele EJ, Vishwanath A. Weyl and Dirac semimetals in three-dimensional solids. *Rev Mod Phys.* 2018;90:015001. Available from: <https://doi.org/10.1103/RevModPhys.90.015001>.
- [20] Xiao D, Chiang MC, Niu Q. Berry phase effects on electronic properties. *Rev Mod Phys.* 2010;82:1959–2007. Available from: <https://doi.org/10.1103/RevModPhys.82.1959>.
- [21] Yang SA. Dirac and Weyl materials: Fundamental aspects and some spintronics applications. *SPIN.* 2016;6:1640003. Available from: <https://doi.org/10.1142/S2010324716400038>.
- [22] Lv BQ, Qian T, Ding H. Experimental perspective on three-dimensional topological semimetals. *Rev Mod Phys.* 2021;93:025002. Available from: <https://doi.org/10.1103/RevModPhys.93.025002>.
- [23] Rachel S. Interacting topological insulators: a review. *Rep Prog Phys.* 2018;81:116501. Available from: <https://doi.org/10.1088/1361-6633/aad6a6>.
- [24] Tokura Y, Yasuda K, Tsukazaki A. Magnetic topological insulators. *Nat Rev Mater.* 2019; 1:126–143. Available from: <https://doi.org/10.1038/s42254-018-0011-5>.
- [25] Wehling TO, Black-Schaffer AM, Balatsky AV. Dirac materials. *Adv Phys.* 2014;63:1–76. Available from: <https://doi.org/10.1080/00018732.2014.927109>.
- [26] Geim AK, Novoselov KS. The rise of graphene. *Nature Mater.* 2007;6:183–191. Available from: <https://doi.org/10.1038/nmat1849>.
- [27] Sato M, Ando Y. Topological superconductors: a review. *Rep Prog Phys.* 2017;80:076501. Available from: <https://doi.org/10.1088/1361-6633/aa6ac7>.
- [28] Thouless DJ, Kohmoto M, Nightingale MP, et al. Quantized Hall conductance in a two-dimensional periodic potential. *Phys Rev Lett.* 1982;49:405–408. Available from: <https://doi.org/10.1103/PhysRevLett.49.405>.
- [29] Kane CL, Mele EJ.  $Z_2$  Topological order and the quantum spin Hall effect. *Phys Rev Lett.* 2005;95:146802. Available from: <https://doi.org/10.1103/PhysRevLett.95.146802>.
- [30] Fu L, Kane CL, Mele EJ. Topological insulators in three dimensions. *Phys Rev Lett.* 2007; 98:106803. Available from: <https://doi.org/10.1103/PhysRevLett.98.106803>.
- [31] Moore JE, Balents L. Topological invariants of time-reversal-invariant band structures. *Phys Rev B.* 2007;75:121306(R). Available from: <https://doi.org/10.1103/PhysRevB.75.121306>.
- [32] Roy R. Topological phases and the quantum spin Hall effect in three dimensions. *Phys Rev B.* 2009;79:195322. Available from: <https://doi.org/10.1103/PhysRevB.79.195322>.
- [33] Hsieh D, Qian D, Wray L, et al. A topological Dirac insulator in a quantum spin Hall phase. *Nature.* 2008;452:970–974. Available from: <https://doi.org/10.1038/nature06843>.
- [34] Xia Y, Qian D, Hsieh D, et al. Observation of a large-gap topological-insulator class with a single Dirac cone on the surface. *Nature Phys.* 2009;5:398–402. Available from: <https://doi.org/10.1038/nphys1274>.
- [35] Qi XL, Li R, Zang J, et al. Inducing a magnetic monopole with topological surface states. *Science.* 2009;323:1184–1187. Available from: <https://doi.org/10.1126/science.1167747>.
- [36] Qi XL, Hughes TL, Zhang SC. Topological field theory of time-reversal invariant insulators. *Phys Rev B.* 2008;78:195424. Available from: <https://doi.org/10.1103/PhysRevB.78.195424>.
- [37] Fu L, Kane CL. Superconducting proximity effect and majorana fermions at the surface of a topological insulator. *Phys Rev Lett.* 2008;100:096407. Available from: <https://doi.org/10.1103/PhysRevLett.100.096407>.
- [38] Bloch F. Über die quantenmechanik der elektronen in kristallgittern. *Z Phys.* 1929;52:555–600. Available from: <https://doi.org/10.1007/BF01339455>.
- [39] Berry MV. Quantal phase factors accompanying adiabatic changes. *Proc R Soc Lond A.* 1984;392:45–57. Available from: <https://doi.org/10.1098/rspa.1984.0023>.
- [40] Zak J. Berry’s phase for energy bands in solids. *Phys Rev Lett.* 1989;62:2747–2750. Available from: <https://doi.org/10.1103/PhysRevLett.62.2747>.

- [41] Nielsen HB, Ninomiya M. The Adler-Bell-Jackiw anomaly and Weyl fermions in a crystal. *Phys Lett B*. 1983;130:389–396. Available from: [https://doi.org/10.1016/0370-2693\(83\)91529-0](https://doi.org/10.1016/0370-2693(83)91529-0).
- [42] Schiff LI. *Quantum Mechanics*. 3rd edition. New York: McGraw–Hill; 1968.
- [43] Haldane FDM. Model for a quantum Hall effect without Landau levels: Condensed-matter realization of the “parity anomaly”. *Phys Rev Lett*. 1988;61:2015–2018. Available from: <https://doi.org/10.1103/PhysRevLett.61.2015>.
- [44] Chang CZ, Zhang J, Feng X, et al. Experimental observation of the quantum anomalous Hall effect in a magnetic topological insulator. *Science*. 2013;340:167–170. Available from: <https://doi.org/10.1126/science.1234414>.
- [45] Hsieh D, Xia Y, Qian D, et al. A tunable topological insulator in the spin helical Dirac transport regime. *Nature*. 2009;460:1101–1105. Available from: <https://doi.org/10.1038/nature08234>.
- [46] Hasan MZ, Kane CL. Topological insulators. *Rev Mod Phys*. 2010;82:3045–3067. Available from: <https://doi.org/10.1103/RevModPhys.82.3045>.
- [47] Ando Y. Topological insulator materials. *J Phys Soc Jpn*. 2013;82:102001. Available from: <https://doi.org/10.7566/JPSJ.82.102001>.
- [48] Nenno DM, Garcia CAC, Gooth J, et al. Axion physics in condensed-matter systems. *Nat Rev Phys*. 2020;2:682–696. Available from: <https://doi.org/10.1038/s42254-020-0240-2>.
- [49] Sekine A, Nomura K. Axion electrodynamics in topological materials. *J Appl Phys*. 2021;129:141101. Available from: <https://doi.org/10.1063/5.0038804>.
- [50] Bernevig BA, Felser C, Beidenkopf H. Progress and prospects in magnetic topological materials. *Nature*. 2022;603:41–51. Available from: <https://doi.org/10.1038/s41586-021-04105-x>.
- [51] Wang Y, Zhang F, Zeng M, et al. Intrinsic magnetic topological materials. *Frontiers of Physics*. 2023;18:21304. Available from: <https://doi.org/10.1007/s11467-022-1250-6>.
- [52] Soh JR, Tobin SM, Su H, et al. Magnetic structure of the topological semimetal YbMnSb<sub>2</sub>. *Phys Rev B*. 2021;104:L161103. Available from: <https://doi.org/10.1103/PhysRevB.104.L161103>.
- [53] Rahn MC, Soh JR, Francoual S, et al. Coupling of magnetic order and charge transport in the candidate Dirac semimetal EuCd<sub>2</sub>As<sub>2</sub>. *Phys Rev B*. 2018;97:214422. Available from: <https://doi.org/10.1103/PhysRevB.97.214422>.
- [54] Mong RSK, Essin AM, Moore JE. Antiferromagnetic topological insulators. *Phys Rev B*. 2010;81:245209. Available from: <https://doi.org/10.1103/PhysRevB.81.245209>.
- [55] Wilczek F. Two applications of axion electrodynamics. *Phys Rev Lett*. 1987;58:1799–1802. Available from: <https://doi.org/10.1103/PhysRevLett.58.1799>.
- [56] Otrokov MM, Klimovskikh II, Bentmann H, et al. Prediction and observation of an antiferromagnetic topological insulator. *Nature*. 2019;576:416–422. Available from: <https://doi.org/10.1038/s41586-019-1840-9>.
- [57] Weyl H. Gravitation and the electron. *Proc Natl Acad, Sci USA*. 1929;15:323–334. Available from: <https://doi.org/10.1073/pnas.15.4.32>.
- [58] Weyl H. Elektron und gravitation. I. *Z Physik*. 1929;56:330–352. Available from: <https://doi.org/10.1007/BF01339504>.
- [59] Castro Neto AH, Guinea F, Peres NMR, et al. The electronic properties of graphene. *Rev Mod Phys*. 2009;81:109–162. Available from: <https://doi.org/10.1103/RevModPhys.81.109>.
- [60] Soluyanov AA, Gresch D, Wang Z, et al. Type-II Weyl semimetals. *Nature*. 2015;527:495–498. Available from: <https://doi.org/10.1038/nature15768>.
- [61] Fang C, Weng H, Dai X, et al. Topological nodal line semimetals. *Chinese Phys B*. 2016;25:117106. Available from: <https://doi.org/10.1088/1674-1056/25/11/117106>.
- [62] Yang SY, Yang H, Derunova E, et al. Symmetry demanded topological nodal-line materials. *Advances in Physics: X*. 2018;3:1414631. Available from: <https://doi.org/10.1080/23746149.2017.1414631>.
- [63] Fang C, Gilbert MJ, Dai X, et al. Multi-Weyl topological semimetals stabi-

- lized by point group symmetry. *Phys Rev Lett.* 2012;108:266802. Available from: <https://doi.org/10.1103/PhysRevLett.108.266802>.
- [64] Bradlyn B, Cano J, Wang Z, et al. Beyond Dirac and Weyl fermions: Unconventional quasiparticles in conventional crystals. *Science.* 2016;353:558. Available from: <https://doi.org/10.1126/science.aaf5037>.
- [65] Zhang T, Jiang Y, Song Z, et al. Catalogue of topological electronic materials. *Nature.* 2019;566:475–479. Available from: <https://doi.org/10.1038/s41586-019-0944-6>.
- [66] Vergniory MG, Elcoro L, Felser C, et al. A complete catalogue of high-quality topological materials. *Nature.* 2019;566:480–485. Available from: <https://doi.org/10.1038/s41586-019-0954-4>.
- [67] Xu Y, Elcoro L, Song ZD, et al. High-throughput calculations of magnetic topological materials. *Nature.* 2020;586:702–707. Available from: <https://doi.org/10.1038/s41586-020-2837-0>.
- [68] Vergniory MG, Wieder BJ, Elcoro L, et al. All topological bands of all nonmagnetic stoichiometric materials. *Science.* 2022;376:816. Available from: <https://doi.org/10.1126/science.abg9094>.