Physics – Uspekhi **61** (11) 1116 – 1126 (2018)

CONFERENCES AND SYMPOSIA

PACS numbers: **73.20.–r**, 73.40.Lq, 73.43.Cd

Understanding surface states of topological insulators

O A Pankratov

DOI: https://doi.org/10.3367/UFNe.2017.12.038307

Contents

Abstract. **Topological electron states were theoretically predicted by B A Volkov and O A Pankratov in 1985 as interface states in an inverted contact between IV–VI semiconductors with their bands mutually inverted. As became clear later, the 'inverted' SnTe semiconductor is a topological insulator, and the inverted contact is an example of a topologically nontrivial interface. This paper discusses the key results of Volkov and Pankratov's 1985 work and examines the usefulness of the inverted contact model for explaining the close link between the topologically nontrivial bulk band structure and topological surface states. An advantage of the model for getting a deeper insight into this link is that it allows an analytical solution. An inhomogeneous semiconductor structure is described by an effective Dirac Hamiltonian, which was obtained analytically from a tight binding model for the band structure of IV–VI materials. This allows one to trace the relation between topological surface states and bands in the bulk. As a result, the spin texture of a topological state can be expressed explicitly in terms of the bulk characteristics. It turns out that the spin texture can be controlled by varying the surface band bending. Given the nontrivial spin polarization on the surface, it is interesting to take a look at the Ruderman–Kittel–Kasuya–Yosida. (RKKY) interaction between magnetic adatoms, which can serve to probe the spin distribution locally. This interaction shows a much more complex structure than the common RKKY coupling in a nonpolarized Fermi gas. The analytical theory provides an explicit relation between the RKKY interaction on the surface of a topological insulator and the parameters of the bulk spectrum.**

Keywords: topological insulators, topological surface states, inverted contact, spin texture

O A Pankratov University of Erlangen-Nürnberg, Staudstr, 7-B2, 91058 Erlangen, Germany
E-mail: Oleg.Pankratov@fau.de

Received 29 January 2018
Uspekhi Fizicheskikh Nauk **188** (11) 1226 – 1237 (2018)
DOI: https://doi.org/10.3367/UFNr.2017.12.038307
Translated by O Pankratov

1. Introduction: the inverted contact

Topological insulators are making headlines in modern solid-state research. It is probably not an exaggeration to say that their discovery is the most important advance in solid-state physics of the last decade. There are many reviews, books, and overview talks on this subject [1–3], and I do not intend to extend this list by one more item. What I am going to do is tell a story that connects some old developments with the present. To the best of my knowledge, Yu V Kopaev, in his last years, was very interested in this topic and gave lectures about topological insulators. It is exactly the kind of physics he liked so much, 'Kopaev's physics' indeed.

My story begins here, at the Lebedev institute, in the 1980s, when B A Volkov and myself published paper [4] in *JETP Letters*, which I briefly summarize below. Consider a Dirac electron with the relativistic energy spectrum

$$E = \pm\sqrt{m^2c^4 + c^2p^2}\,. \tag{1}$$

Apparently, the sign of the rest mass m is irrelevant here since it enters as a square in Eqn (1). However, if we imagine that the mass is a function of a position, and this function changes sign, we encounter a new phenomenon. Namely, if the mass is a function of z coordinate $m(z)$ and changes its sign at $z = 0$. then, regardless of the particular shape of the function $m(z)$, electron states which are bound to the plane $z = 0$ always exist. Of course, these states are plane waves along the $z = 0$ plane. They have a linear in-plane dispersion $\varepsilon(\mathbf{p}_\perp) = \pm cp_\perp$, where $\mathbf{p}_\perp = (p_x, p_y)$ stands for the in-plane momentum (Fig. 1). Interestingly, these states are not spin-degenerate, that is, the spin structure of the wave function is 'frozen'. What we have here are two-dimensional chiral massless particles — Weyl neutrinos.

An inverted semiconductor structure was a rather speculative suggestion in 1985. However, later on, the linear electron spectrum was discovered, first in graphene and then in topological insulators. In graphene, the analogy is not complete, since the spin remains 'free', i.e., decoupled from the orbital motion due to the extreme weakness of the spin-orbit interaction. On the contrary, the 2D states on the

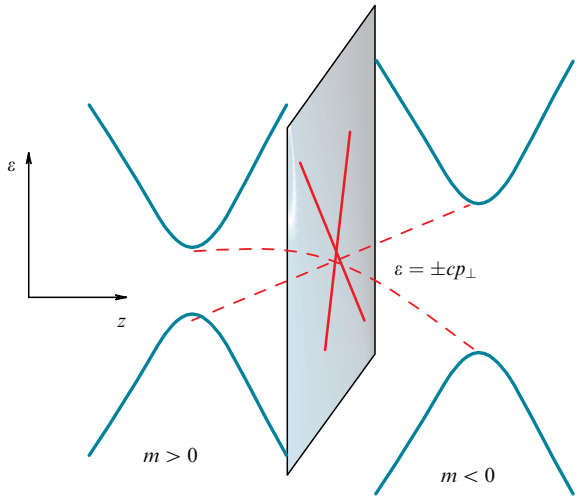

Figure 1. Inverted contact between Dirac spaces with negative and positive rest mass. Regardless of the particular shape of the transition region, the interface hosts spin-nondegenerate Weyl states with linear dispersion.

surface of a 3D topological insulator are indeed Weyl fermions.

Of course, we had in mind a particular realization of the variable Dirac mass. In a family of IV–VI semiconductors, there are materials with mutually inverted fundamental gap edges, i.e., with positive and negative Dirac mass. In these materials, a direct fundamental band gap is located in the L point of the Brillouin zone. The low energy electron spectrum is, to a very good approximation, described by the Dirac Hamiltonian. The gap edges are labeled by the parity, which is opposite for a conduction and a valence band. For example, in PbTe, the conduction band edge is 'odd', whereas the valence edge is 'even' (the parity is commonly defined choosing the origin of the coordinate system on a Pb lattice site). In contrast, in SnTe, the parities of the two bands are opposite. Therefore, in a solid solution $Pb_{1-x}Sn_xTe$, a gradual increase in Sn content should cause the gap to close and finally to invert (Fig. 2). By a spatial variation of the Sn content, one can achieve the band inversion in real space (Fig. 3), a situation that we call an inverted contact (this system is also referred to as a 'topological heterojunction' [5, 6]). As mentioned above, the universal property of such a structure is 2D Weyl fermionic states at the interface.

The inhomogeneous structure depicted in Fig. 3 is described by the Dirac equation with the variable mass, where mc^2 is replaced by some function $\Delta(z) = \Delta_0 f(z)$, which changes sign when crossing the plane $z = 0$. Of course, some material-dependent in-plane velocities $v_\parallel$ (parallel to z) and $v_\perp$ (perpendicular to z, i.e. in the junction plane) must replace the speed of light. As we found in [4], the Dirac equation

$$\widehat{H}(z, \mathbf{p}_\perp)\, \Psi = \varepsilon(\mathbf{p}_\perp)\, \Psi \tag{2}$$

always has at least one bound state, which is localized along z regardless of the particular shape of $f(z)$. The only condition imposed on $f(z)$ is the sign change. This solution possesses a linear in-plane dispersion

$$\varepsilon(\mathbf{p}_\perp) = \pm\, v_\perp p_\perp\,. \tag{3}$$

Whereas the bulk bands are spin-degenerate, the interface state occurs only in a specific spin configuration, i.e., it has a

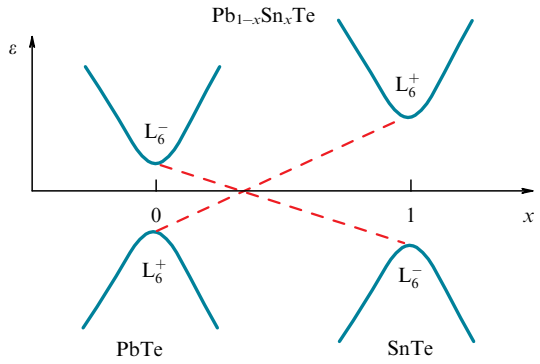

Figure 2. Inversion of the fundamental gap edges in $Pb_{1-x}Sn_xTe$ with an increase in Sn content.

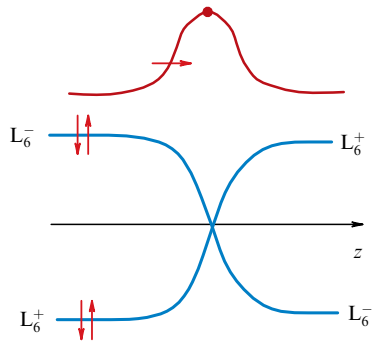

Figure 3. Real space energy diagram of an inverted contact. Here, we ignore the shift of the middle point of the band gap, the band bending. In addition, the band gaps on different sides of the junction are not necessarily equal. Whereas the bulk bands are spin-degenerate, the Weyl state, which is trapped at the interface, has a fixed spin structure.

fixed spin structure:

$$\Psi_\pm = \frac{1}{2}\begin{pmatrix} \exp\left[-\mathrm{i}\left(\frac{\theta}{2}+\frac{\pi}{4}\right)\right] \\ \pm\exp\left[\mathrm{i}\left(\frac{\theta}{2}+\frac{\pi}{4}\right)\right] \\ \exp\left[-\mathrm{i}\left(\frac{\theta}{2}-\frac{\pi}{4}\right)\right] \\ \pm\exp\left[-\mathrm{i}\left(\frac{\theta}{2}-\frac{\pi}{4}\right)\right] \end{pmatrix} \times \exp\left[-\frac{1}{\hbar v_\parallel}\int_0^z \Delta(z)\,\mathrm{d}z + \mathrm{i}\mathbf{p}_\perp\mathbf{r}\right]. \tag{4}$$

In Eqn (4), the plus/minus sign refers to the positive/negative energy branches and the phase factors are defined by the polar angle θ which controls the direction of the in-plane momentum $\mathbf{p}_\perp$. It is evident that wave function (4) is exponentially localized along z if $\Delta(z)$ changes sign. Otherwise, it becomes nonnormalizable and the bound state vanishes. In the plane $z = 0$, interface states are described by the effective Weyl Hamiltonian

$$\widehat{H}_\mathrm{W} = v_\perp \boldsymbol{\tau}_\perp \mathbf{p}_\perp \tag{5}$$

with the linear in-plane energy spectrum (3). The matrix-valued vector $\boldsymbol{\tau}_\perp = (\tau_x, \tau_y)$ is the vector of the pseudospin

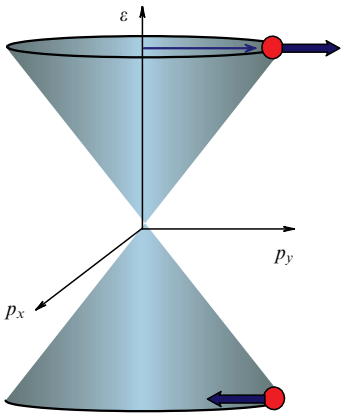

Figure 4. Dirac–Weyl cone: the in-plane energy dispersion of the interface states. The positive and the negative energy branches correspond to the pseudospin orientation parallel or antiparallel to the in-plane momentum, i.e., to positive or negative chirality.

Pauli matrices. The projection of $\boldsymbol{\tau}_\perp$ on $\mathbf{p}_\perp$ determines the chirality eigenvalues ± 1, which correspond to positive or negative energy branches in Eqn (3), Fig. 4. The most surprising thing about solution (4) is its robustness with respect to the shape of $f(z)$, alongside the absence of spin degeneracy. At the time we published paper [4], the origin of these unusual features was unclear. Now, more than 20 years later, we know the answer, and it is the topology of the underlying band structure.

2. Topological insulators

The band structure of an insulating material can be classified [1–3] according to the topology of the Hilbert space, which is characterized by the corresponding topological index. The index does not change upon any smooth deformation of the Hamiltonian, unless the energy levels cross. A change in the topological index is only possible via a level crossing that is via the emergent metallic state. This happens at the interface separating two topologically distinct materials or on the surface of a topologically nontrivial material, which is an interface with a vacuum. The metallic interface state must be spin-nondegenerate, possessing linear dispersion [1–3], i.e., with exactly the same properties as the Weyl state (4). It is not the purpose of this paper to review the topological band theory; I illustrate in Fig. 5 the change in the band structure topology by analogy with the topological classification of surfaces. For closed surfaces, the topological index characterizes their connectivity. By the astonishing theorem of Gauss–Bonnet, the integral (divided by 2π) of the curvature over a closed surface is an integer $2(1-g)$, where g is the number of holes in the surface (Fig. 6).

The integral of the curvature does not change upon an arbitrary deformation of the surface, unless the topology (characterized by the number of holes) changes. In the case of the band structure, the topology describes the 'connectivity' of the Hilbert space.

The change in the topological index signals the existence of interface/surface states, which are universal, i.e., which do not depend critically on the details of the interface. This is illustrated in Fig. 7. The left panel (a) shows 'ordinary' surface states, which come as Kramers pairs. Generally, if there is no inversion symmetry, the spin degeneracy is lifted, but the

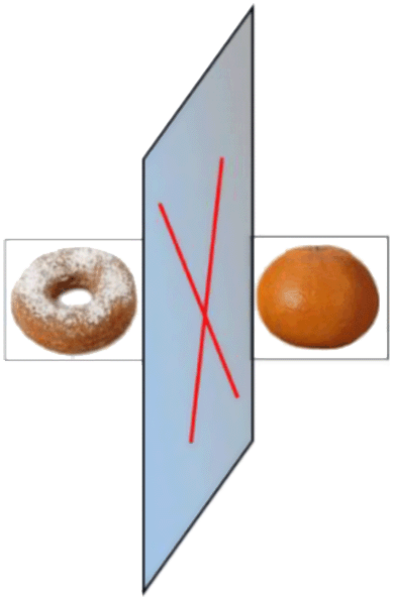

Figure 5. Dirac–Weyl interface state emerges as a metallic state connecting insulating materials (one of which may be a vacuum) with a different band topology.

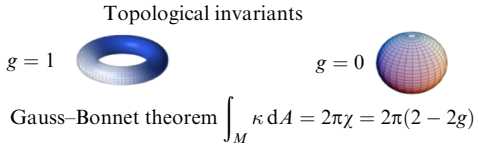

Figure 6. Topological classification of surfaces and the Gauss–Bonnet theorem. Here, κ is the surface curvature $\kappa = 1/r_1 r_2$ and g is the number of holes.

surface bands must cross at high symmetry points of the surface Brillouin zone as required by the Kramers theorem. Such states can be, in principle, removed from the band gap by the action of some surface potential. On the contrary, the states in panel (b) cannot be removed, since they form a continuous chain through the gap switching Kramers partners at high symmetry points. The intersection of these surface bands at a symmetry point forms a Weyl–Dirac cone. Figure 7 suggests a way to distinguish between the 'ordinary' and the 'topologically protected' states. To do this, it is enough to count the number of intersections of the surface bands with the line of constant energy. In the ordinary case, the number of intersections is even, whereas in the topological case it is odd.

As the spin structure of a surface state is 'frozen', every surface state at a particular in-plane momentum $\mathbf{p}_\perp = (p_x, p_y)$ has a certain expectation value of spin. The spin distribution in the (p_x, p_y) plane forms a spin texture, which is the most important hallmark of a topological surface state. Spin textures were measured by spin-resolved ARPES (Angle-Resolved Photoemission Spectroscopy) for a number of topological insulators. An example of the Bi_2Se_3 topological insulator is shown in Fig. 8. Here, we have a single Dirac–Weyl cone in the center of a surface Brillouin zone. The spin is locked perpendicularly to momentum, as one would expect from the symmetry arguments. However, in general, topological theory cannot predict the spin texture. This may seem paradoxical: on the one hand, the very existence of the

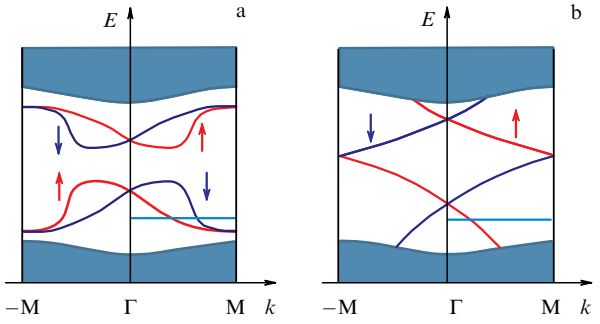

Figure 7. Adapted from [2]. Ordinary surface states shown in panel (a) come in pairs of Kramers-conjugated bands. The number of crossings with the line of constant energy is even, and any surface state can be pushed out of the energy gap by applying a proper surface potential. Panel (b): topological surface state. The number of crossings is odd. Such a state cannot be eliminated from the band gap.

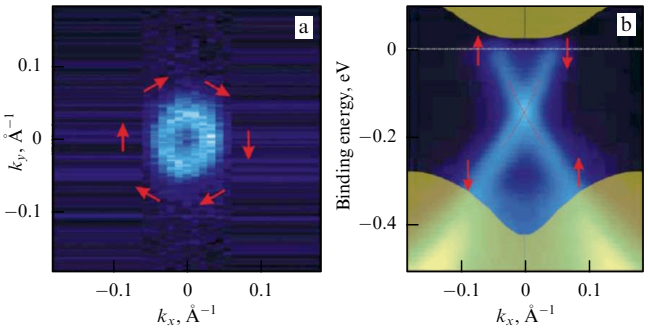

Figure 8. ARPES images of the topological surface states in Bi_2Se_3 [7]. There is a single Dirac–Weyl cone in the center on the surface Brillouin zone and the helical spin texture is circularly symmetric.

topological surface state is due to the nontrivial topology of the bulk. On the other hand, we cannot predict what happens to that state if we change the bulk, and we generally do not know what the spin texture looks like. To address this issue one needs an explicit connection between the bulk and the surface electronic states. Unfortunately, most topological insulators possess complicated crystal structures, which makes it impossible to make this connection explicit.

Ab initio calculations provide both the bulk and the surface states, but the link between the two remains hidden. Now, we turn to the inverted contact model, which offers a unique opportunity to study this link analytically.

3. Is inverted contact PbTe/SnTe a topological interface?

According to the criterion formulated in Fig. 7, it seems that the answer to this question should be 'no'. Indeed, the band extrema in IV–VI materials are located at L-points and there are four inequivalent L points in the Brillouin zone. They project in different ways on different surfaces, but their number is always even and, hence, the number of surface bands is even. It follows that these states are not protected by time-reversal symmetry, which commonly underlies topological classification [1–3]. Nonetheless, numerical calculations for SnTe (Fig. 9) show that the surface bands stemming from different L points do not form self-closed Kramers pairs, as in Fig. 7a, but extend through the whole band gap, as in Fig. 7b.

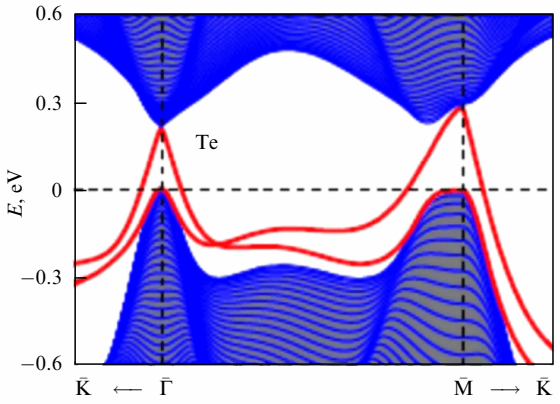

Figure 9. Numerically calculated surface states on the (111) Te-terminated surface of SnTe [8]. The two states that are derived from the different L points of the bulk spectrum belong to different symmetry classes.

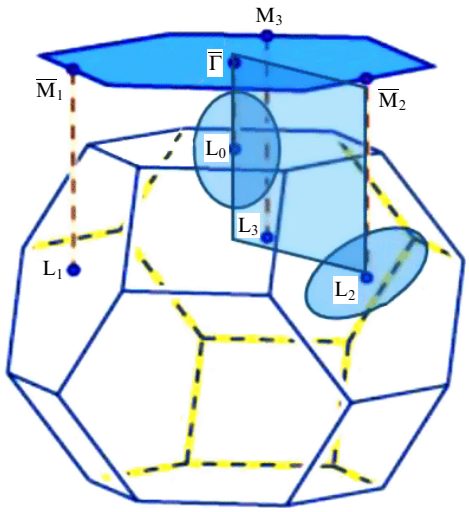

Figure 10. Brillouin zone for SnTe (after [15]). The band extrema are located at L points, four of which are inequivalent. On the (111) surface, the L points project onto the central Γ point and three inequivalent M points in the surface Brillouin zone. The (110) symmetry plane is common to the bulk and to the surface. The anisotropy of the energy spectrum at the L points is schematically shown.

Two surface states in Fig. 9, which are derived from the bulk states L_2 and L_0 (Fig. 10), do not interact at the crossing point, because they possess different symmetry with respect to reflection in the mirror plane (110) (Fig. 10, 11). This crystal symmetry protection replaces the time-reversal symmetry protection originally introduced for topological insulators. Hence, SnTe-type materials belong to the so-called crystalline topological insulators, where topological protection of the surface state is conditioned by a particular point group symmetry element, which is common to the bulk crystal and to the surface [9]. The mirror plane (110) is such a symmetry element.

The corresponding topological index, a mirror Chern number [10], is defined as the difference between the Chern numbers for energy bands with opposite mirror eigenvalues [11–14]:

$$n_{\mathrm{M}} = \frac{n_{+i} - n_{-i}}{2} \,. \qquad (6)$$

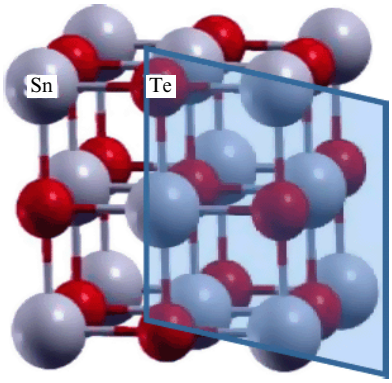

Figure 11. Rock salt crystal structure of SnTe (from [12]). The (110) symmetry plane is shown.

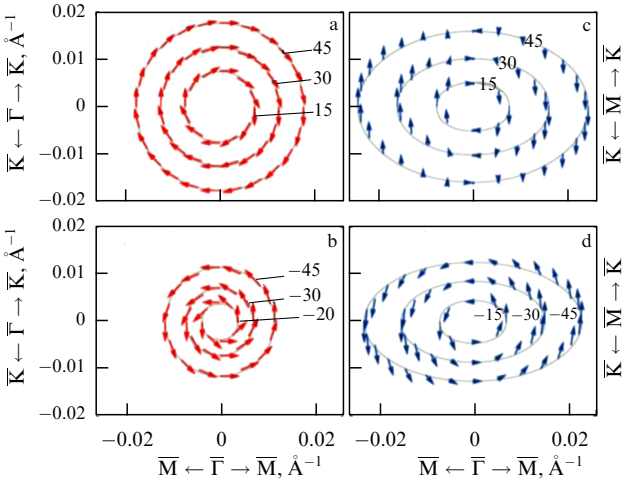

Figure 12. Ab initio calculations of SnTe spin textures [15]. The in-plane projection for upper (a, c) and lower (b, d) Dirac–Weyl state on the (111) surface of SnTe is shown. Panels (a, b) refer to the Γ state and panels (c, d) to the M state. Numbers are the energies of the constant energy lines in meV.

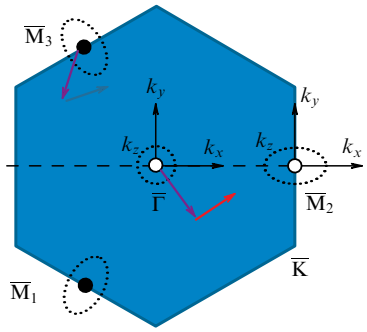

Figure 13. Surface Brillouin zone for the (111) surface of SnTe (from [15]).

For SnTe, $n_{\mathrm{M}} = -2$ [12]. In contrast, for PbTe, which has the 'normal' band ordering, $n_{\mathrm{M}} = 0$ [12]. The inverted contact is thus a topological interface where the topology change is reflected by the change of a mirror Chern number. Hence, the Dirac–Weyl interface state that we found in 1985 indeed has a topological origin. Therefore, it possesses all the hallmarks of the topological surface state: it spans the whole band gap and it is not spin-degenerate.

The spin texture for the SnTe (111) surface was calculated from first principles [15]; see Figs 12, 13. The surface states in Γ and M, which originate from 'direct' and 'oblique' L points, have profoundly distinct textures. At the Γ point, a cylindrical symmetry dictates 'standard' spin locking with the spin oriented perpendicularly to momentum.

In contrast, for M state, the symmetry constrains are released and the spin is no more locked normally to momentum. Moreover, there is a spin component normal to the surface, which is strictly zero at the Γ point. These results indicate that the surface spin texture contains information about the parent bulk states. However, without analytical theory, the link between the two remains concealed. At this point, we can benefit from the inverted contact model that allows, for IV–VI materials, this connection to be unraveled and thus making the bulk-boundary correspondence explicit.

To calculate the spin texture, we need to identify the way the real spin enters the effective Dirac Hamiltonian in Eqn (2). The Hamiltonian matrix is defined in the space of Kramers-conjugated states, which are not pure spin states. Calculation of the spin expectation values requires knowledge of the basis states that underlie the Dirac matrix. To elucidate that, we have to step back in time to the early 1980s, when B A Volkov and myself developed the analytical band structure theory for IV–VI materials [16, 17]. This work eventually led us to the concept of the band-inverted contact.

4. Where does the Dirac Hamiltonian come from? Analytical band structure theory

Our derivation of the effective Dirac Hamiltonian for electron states at the L point is based on understanding the IV–VI band structure as a whole. This understanding can be achieved by constructing tight binding Bloch states from atomic p-orbitals [16, 17]. The cubic symmetry of the crystal allows choosing the p-orbitals as functions of the x, y, z symmetry, transforming as Cartesian coordinates (Fig. 14). From these orbitals, we construct the tight binding Bloch functions $|x\rangle$, $|y\rangle$, $|z\rangle$.

In this approach, it is immediately visible why L points (Fig. 15) are so special in IV–VI materials. Indeed, at the L point, the tight binding wave function is a sum that runs either over the group IV (Pb, Sn, Ge) sites or over the group VI (Te, Se) sites. Due to the symmetry of the [111] axis, the $|x\rangle$, $|y\rangle$, $|z\rangle$ states enter on equal footing, and corresponding energy levels are degenerate if we disregard the crystal field and the spin-orbit interaction. This degeneracy is lifted by the mixing of the $|x\rangle$, $|y\rangle$, $|z\rangle$ states, induced by the cubic crystal field and by the spin-orbit coupling. The splitting drastically reduces the gap that separates the Sn- and Te-type triplets, which results in a minimal gap in the electron spectrum (Fig. 16). Since there are 6 valence p-electrons in the unit cell, the three lower levels are occupied, thus forming valence bands, whereas the three upper levels form conduction bands. Thus, group IV-derived levels normally form the conduction bands with odd parity and group VI states the valence bands with even parity. However, in some cases, for example in SnTe, the gap reduction is so strong that the edge states interchange their positions and an inverted band gap (as shown in Fig. 16) emerges.

The inverted band order is in contrast to the 'normal' one that occurs, for example, in PbTe, which may be considered topologically equivalent to a vacuum, since the 'normal'

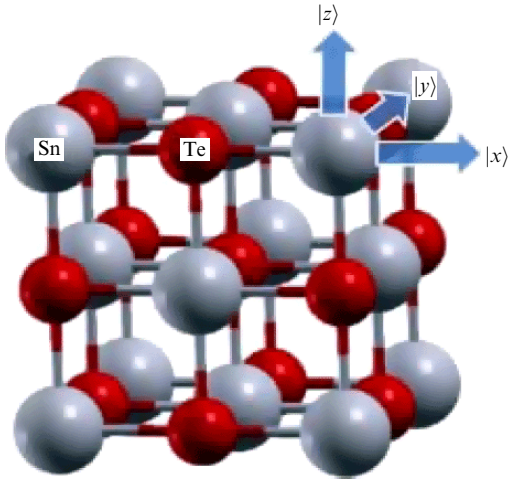

Figure 14. Symmetry of the atomic p-orbitals for tight binding Bloch functions [16, 17] (adapted from [12]).

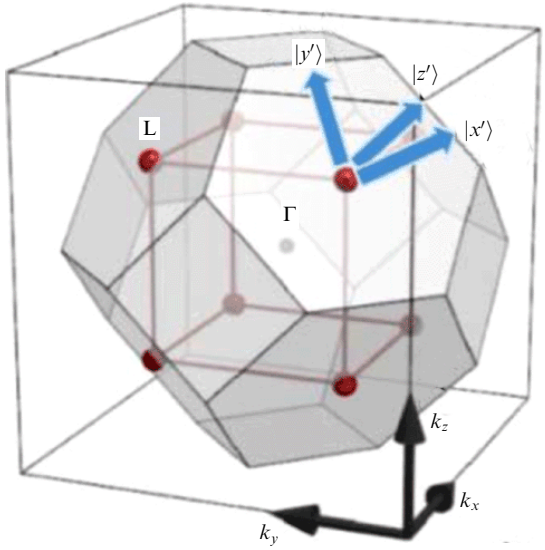

Figure 15. At the L point, the symmetry dictates a convenient choice of the p-Bloch functions in the rotated coordinate system with the z' axis along the [111] symmetry axis.

ordering with the higher cation (negative parity) state and lower anion (positive parity) level should be restored with the increase in the lattice constant, i.e., for a vanishing crystal field. The construction of the edge states according to the diagram in Fig. 16 provides the wave functions at the L-point

$$
\begin{aligned}
|\Phi_2^-\rangle &= -\sin\frac{\theta^-}{2}\,|\Phi_+^{-\downarrow}\rangle + \cos\frac{\theta^-}{2}\,|\Phi_0^{-\uparrow}\rangle\,,\\
\widehat{K}|\Phi_2^-\rangle &= -\sin\frac{\theta^-}{2}\,|\Phi_-^{-\uparrow}\rangle + \cos\frac{\theta^-}{2}\,|\Phi_0^{-\downarrow}\rangle\,,\\
|\Phi_1^+\rangle &= -\cos\frac{\theta^+}{2}\,|\Phi_+^{+\downarrow}\rangle + \sin\frac{\theta^+}{2}\,|\Phi_0^{+\uparrow}\rangle\,,\\
\widehat{K}|\Phi_1^+\rangle &= \cos\frac{\theta^+}{2}\,|\Phi_-^{+\uparrow}\rangle + \sin\frac{\theta^+}{2}\,|\Phi_0^{+\downarrow}\rangle\,,
\end{aligned}
\tag{7}
$$

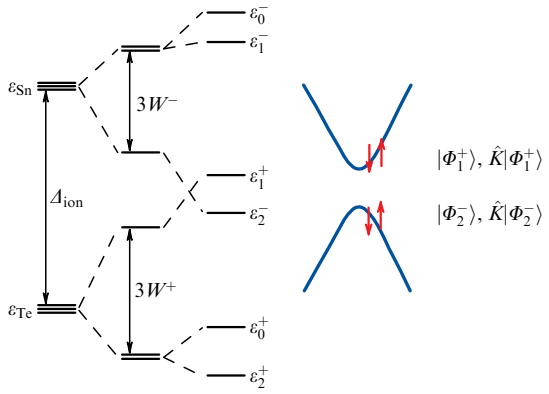

Figure 16. Formation of the fundamental band gap in the L point from atomic-like p-states, which form Sn-type and Te-type triplets initially separated by the difference between the atomic ionization potentials Δ_{ion}. This energy gap shrinks after crystal field splitting and spin-orbit splitting are turned on. Band inversion occurs if the splitting exceeds the initial separation of odd and even states [17].

where the Bloch functions $|\Phi_0^\pm\rangle$, $|\Phi_\pm^\pm\rangle$, which transform according to irreducible representations of the L-point small group D_{3d}, are combinations of the Bloch states $|x'\rangle$, $|y'\rangle$, $|z'\rangle$, which we denote as $|\Phi_{x'}^\pm\rangle$, $|\Phi_{y'}^\pm\rangle$, $|\Phi_{z'}^\pm\rangle$:

$$
\begin{aligned}
|\Phi_0^\pm\rangle &= |\Phi_{z'}^\pm\rangle\,,\\
|\Phi_\pm^\pm\rangle &= \frac{1}{\sqrt{2}}\left(|\Phi_{x'}^\pm\rangle \pm \mathrm{i}\,|\Phi_{y'}^\pm\rangle\right).
\end{aligned}
\tag{8}
$$

The last functions are obtained from the $|x\rangle$, $|y\rangle$, $|z\rangle$ basis functions by rotation of the cubic coordinate system to the L point coordinate system with the z' axis parallel to [111] (Fig. 15). The $\pm$ superscript in (7) and (8) refers to even and odd functions. The spin superscript in Eqn (7) bears on the z' component of the spin quantized along the [111] axis. As shown in Fig. 16, there are four states belonging to the gap edges: the functions $|\Phi_2^-\rangle$ and $|\Phi_1^+\rangle$ and their Kramers partners $\widehat{K}|\Phi_2^-\rangle$ and $\widehat{K}|\Phi_1^+\rangle$; see Eqn (7). The spin ad-mixture due to the spin-orbit coupling is controlled by the material-dependent *spin mixing angles* $\theta^\pm$ that depend on the ratio of the crystal field matrix elements $\langle x|\widehat{H}|y\rangle$ and the spin-orbit coupling strength. In Ref. [17], the values of $\theta^\pm$ are tabulated for all materials of the IV–VI family. The effective Dirac Hamiltonian at the L point in the basis of functions (7) reads

$$
H = \begin{pmatrix} \varepsilon_g/2 & v_\parallel\sigma_z p_z + v_\perp\boldsymbol{\sigma}_\perp\mathbf{p}_\perp \\ v_\parallel\sigma_z p_z + v_\perp\boldsymbol{\sigma}_\perp\mathbf{p}_\perp & -\varepsilon_g/2 \end{pmatrix},
\tag{9}
$$

with ε_g being the band gap $\mathbf{p}_\perp = (p_x, p_y)$ and $\boldsymbol{\sigma}_\perp = (\sigma_x, \sigma_y)$. It is instructive to note the physical meaning of $\cos\theta^\pm$. A simple calculation gives for the real spin $\mathbf{s} = (s_x, s_y, s_z)$ the average values

$$
\begin{aligned}
&\langle\Phi_2^-|s_z|\Phi_2^-\rangle = \cos\theta^-, \quad \langle\widehat{K}\Phi_2^-|s_z|\widehat{K}\Phi_2^-\rangle = -\cos\theta^-\,,\\
&\langle\Phi_2^-|s_{x,y}|\Phi_2^-\rangle = 0\,,\\
&\langle\Phi_1^+|s_z|\Phi_1^+\rangle = -\cos\theta^+, \quad \langle\widehat{K}\Phi_1^+|s_z|\widehat{K}\Phi_1^+\rangle = \cos\theta^+\,,\\
&\langle\Phi_1^+|s_{x,y}|\Phi_1^+\rangle = 0\,.
\end{aligned}
\tag{10}
$$

Hence, $\cos\theta^\pm$ gives the spin polarization of the bulk L-state along the [111] axis.

5. Topological surface states and spin texture

To introduce topological surface states, we, in analogy with the inverted contact, make use of a topological boundary condition [18, 19]. This condition requires band inversion on the surface while sending the band gap to infinity on the vacuum side (Fig. 17). In other words, the surface is described as inverted contact with the vacuum. We implement this by replacing ε_g with the function $\Delta(z)$ that changes sign on the surface. With this approach, we are in a position to calculate the surface spin texture analytically. We focus on the case of the (111) surface, which is the most transparent. There are two additional factors that one must take into account in considering a realistic surface [20].

First, one has to include band bending, i.e., the change in the work function close to the surface. This is done by introducing a scalar function $\varphi(z)$ on the diagonal of the Hamiltonian matrix (9). Second, we have to take into account that the surface normal is not necessarily aligned with the symmetry axis [111] of the parent L point. For the (111) surface, the alignment is present for a 'direct' point L_0 which generates the surface state at Γ, but not for other three L points $L_{1,2,3}$ that are responsible for topological states at M points (see Fig. 10). To implement this in Eqn (9), we replace the nondiagonal part in matrix (9) as follows:

$$v_{\parallel}\sigma_z p_z + v_{\perp}\boldsymbol{\sigma}_{\perp}\mathbf{p}_{\perp} \Rightarrow \boldsymbol{\sigma} R V R^{\dagger}(\mathbf{p}_{\perp} - \mathrm{i}\hbar\mathbf{e}_z\partial_z) \equiv \Lambda\,. \tag{11}$$

Here, $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$, V is the velocity matrix $V = \mathrm{Diag}\,(v_{\perp}, v_{\perp}, v_{\parallel})$ with velocities perpendicular or parallel to the [111] direction, $\mathbf{e}_z$ is the unit vector normal to the surface, and the rotation matrix R describes a coordinate transformation of the [111] aligned coordinate system into the new orientation, such that the z-axis coincides with the surface normal. This rotation must also be performed on the basis functions (7). For example, for the case of the M state, R is simply a rotation by the angle β between the two [111] cubic diagonals, where $\cos\beta = 1/3$. The Dirac equation takes the form

$$\begin{pmatrix} \Delta(z) + \varphi(z) & \Lambda \\ \Lambda^{\dagger} & \Delta(z) + \varphi(z) \end{pmatrix} \Psi = \varepsilon\Psi\,. \tag{12}$$

Equation (12) can be solved by a similar strategy that has been deployed in Refs [21, 22]. Under the assumption that the coordinate dependence for $\Delta(z) = \Delta_0 f(z)$ and for $\varphi(z) = \varphi_0 f(z)$ is given by the same sign-changing function $f(z)$, it can be shown [21, 22] that the consequence of the band bending is that the Weyl interface state becomes constrained in a certain energy window (in fact, this is always observed in numerical calculations for topological insulators).

Let us first consider a simpler case of a 'direct' L valley, which projects onto the surface Γ point, and disregard for a moment the band bending. In this case, there is no need to rotate the coordinate system, and the Λ operator in Eqn (12) is given by the left-hand side of Eqn (11). The 4-component wave function is given by Eqn (4):

$$\Psi_{\pm} = \frac{1}{2}\begin{pmatrix} \exp\left[-\mathrm{i}\left(\dfrac{\theta}{2} + \dfrac{\pi}{4}\right)\right] \\ \pm\exp\left[\mathrm{i}\left(\dfrac{\theta}{2} + \dfrac{\pi}{4}\right)\right] \\ \exp\left[-\mathrm{i}\left(\dfrac{\theta}{2} - \dfrac{\pi}{4}\right)\right] \\ \pm\exp\left[\mathrm{i}\left(\dfrac{\theta}{2} - \dfrac{\pi}{4}\right)\right] \end{pmatrix}, \tag{13}$$

where we omitted the z-dependent part, as well as the plane wave part. This expression provides the expansion coefficients of the full wave function $\Psi_{\pm}$ with the basis functions $|\Phi^i\rangle$ of the L point states (7)

$$\Psi_{\pm} = \sum_i \Psi^i_{\pm}|\Phi^i\rangle\,. \tag{14}$$

Now, we can calculate the mean values of the spin components:

$$\begin{aligned} &\langle\Psi_{\pm}|\mathrm{s}_x|\Psi_{\pm}\rangle = \mp s\sin\theta\,, \\ &\langle\Psi_{\pm}|\mathrm{s}_y|\Psi_{\pm}\rangle = \pm s\cos\theta\,, \\ &\langle\Psi_{\pm}|\mathrm{s}_z|\Psi_{\pm}\rangle = 0\,. \end{aligned} \tag{15}$$

The angle θ determines the direction of the in-plane momentum $\tan\theta = p_x/p_y$. The amplitude of the spin is

$$s = \frac{1}{4}(\cos\theta^{+} + \cos\theta^{-})\,. \tag{16}$$

We see that s is controlled by the bulk band structure parameters, the spin mixing angles $\theta^{\pm}$ that are directly related to the spin polarizations of the bulk states at the L-point; see Eqn (10). Depending on the spin mixing angles, the amplitude of the spin texture (16) can change sign and, for some parameter values, can vanish. From Eqn (15), it follows that the spin vector is locked perpendicular to the momentum; hence, we have a common helical spin texture (Fig. 18). However, this is true only for the Γ valley. In the more general case of the M valley, we find that the in-plane spin is tangent to some conical section contour in the (p_x, p_y) plane [20].

For a general orientation of the surface with respect to the L point symmetry axis and for nonzero band bending, a direct

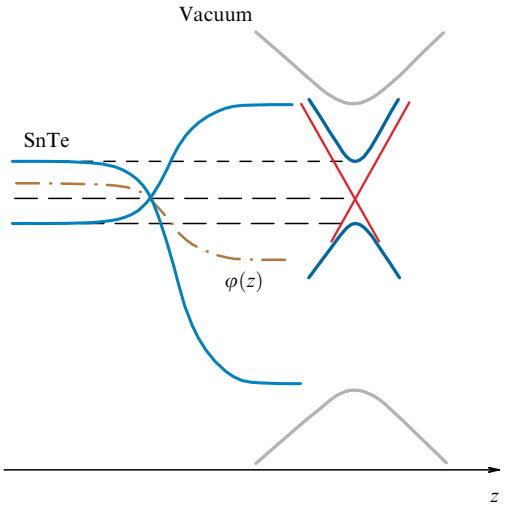

Figure 17. Topological boundary condition on an SnTe surface: the band inversion and the displacement of the fundamental gap center (band bending). On the vacuum side, the gap goes to infinity. The origin of the Dirac–Weyl cone is aligned with the zero energy point $\Delta(z) = \varphi(z) = 0$. Shown is the case of downward band bending, when the topological state is constrained from below.

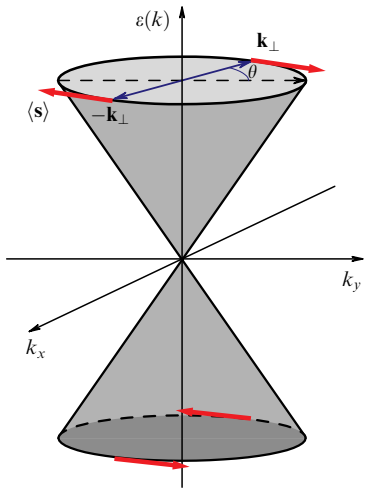

Figure 18. Spin-momentum locking for the surface Γ state. Energy branches with positive or negative energy/chirality have opposite spin directions.

calculation gives

$$\langle \Psi_\pm | \mathbf{s} | \Psi_\pm \rangle = \pm \begin{pmatrix} -a \sin\theta \\ b \cos\theta \\ m_z \sin\theta \end{pmatrix}, \tag{17}$$

where the values of a, b, and m_z depend on the orientation of the symmetry axis of the parent L point. For M states on the (111) surface, the symmetry axis of the parent L point is at an angle β with the surface normal [111], where $\cos\beta = 1/3$. For an arbitrary surface orientation, we have

$$\begin{aligned} a &= c\,\frac{v_x}{v_\parallel}\sin^2\beta + b\,\frac{v_x}{v_\perp}\cos^2\beta\,, \\ b &= s - \frac{\varphi_0}{2\Delta_0}\left(\cos^2\frac{\theta^-}{2} - \sin^2\frac{\theta^+}{2}\right), \\ c &= 2b + \frac{\varphi_0}{\Delta_0}\,, \\ m_z &= \sin(2\beta)\,\frac{1}{2}\left(c\,\frac{v_x}{v_\parallel} - b\,\frac{v_x}{v_\perp}\right). \end{aligned} \tag{18}$$

The Weyl cone is anisotropic,

$$\varepsilon(\mathbf{p}_\perp) = \gamma\sqrt{v_x^2 p_x^2 + v_\perp^2 p_y^2}\,, \tag{19}$$

with

$$v_x = \frac{v_\parallel v_\perp}{\sqrt{v_\perp^2\sin^2\beta + v_\parallel^2\cos^2\beta}} \tag{20}$$

and

$$\gamma = \sqrt{1 - \frac{\varphi_0^2}{\Delta_0^2}}\,. \tag{21}$$

From Eqn (19), it is obvious that the anisotropy of the Weyl cone is due to the anisotropy of velocities at the parent L point. In contrast, the spin texture anisotropy has its origin mainly in different spin polarizations at different L points. Indeed, for a symmetric Γ point, the spin texture is perfectly helical. The [111] symmetry axis, along which the spin polarization of the parent L states is aligned, coincides with the surface normal. For M points, the parent symmetry axes are tilted with respect to the surface normal. This leads to anisotropy and to the finite polarization component m_z, which is normal to the surface. An interesting feature of Eqns (17), (18) is a strong dependence of the spin texture on the band bending φ_0. It is obvious from Eqn (21) that, if $|\varphi_0|$ exceeds $|\Delta_0|$, the topological state vanishes. This is because the potential in the effective Schrödinger equation (which coincides with the equation of Witten's supersymmetric quantum mechanics [21, 22]) that follows from Dirac equation (12) loses the property of changing sign. For $|\varphi_0| < |\Delta_0|$, the Dirac–Weyl state exists, but it is constrained in energy; namely, it terminates at the points where the Weyl branches touch the bulk spectrum (the bulk bands should be taken at the zero momentum component, which is normal to the interface). For inverted contact between two semiconductors, where we have finite energy gaps on both sides of the junction, the Dirac–Weyl cone is constrained in energy in both the positive and negative energy ranges. The limiting energies are [21, 22]

$$\eta_\pm = -\left(\frac{\Delta_0^2}{\varphi_0^2} - 1\right)\varphi(\pm\infty)\,. \tag{22}$$

In the case of a surface, the effective gap is infinite on the vacuum side and the topological state has a cutoff at negative energies for downward band bending ($\varphi_0 > 0$; see Fig. 19) or at positive energies for upward band bending ($\varphi_0 < 0$) [6]. We define the sign of the band bending assuming $f(-\infty) = 1$ on the material side and $f(+\infty) \to -\infty$ on the vacuum side.

In Fig. 20, we show the evolution of the spin texture as a function of φ_0 [20]. The spin polarization was evaluated analytically for SnTe using the band structure parameters from [17] for four different values of φ_0/Δ_0. For a downward band bending, $\varphi_0 > 0$, which is seen in experiment [24], the spin textures at Γ and M are similar (panel (a)), except that for the Γ valley we have a perfect circular helix, but for the M valley it is slightly distorted. The winding numbers for

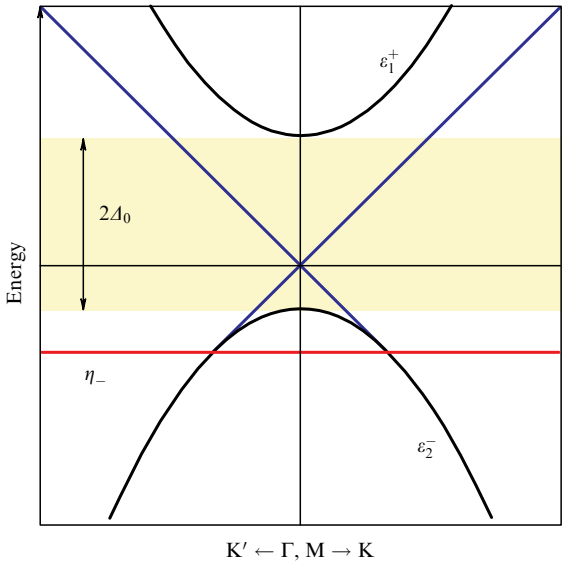

Figure 19. Dispersion of the Γ and M Dirac–Weyl topological surface states [23] for downward band bending. ε_1^+ and ε_2^- are projections of the bulk L bands onto the (111) surface. The topological surface band is cut off at the negative limiting energy η_-.

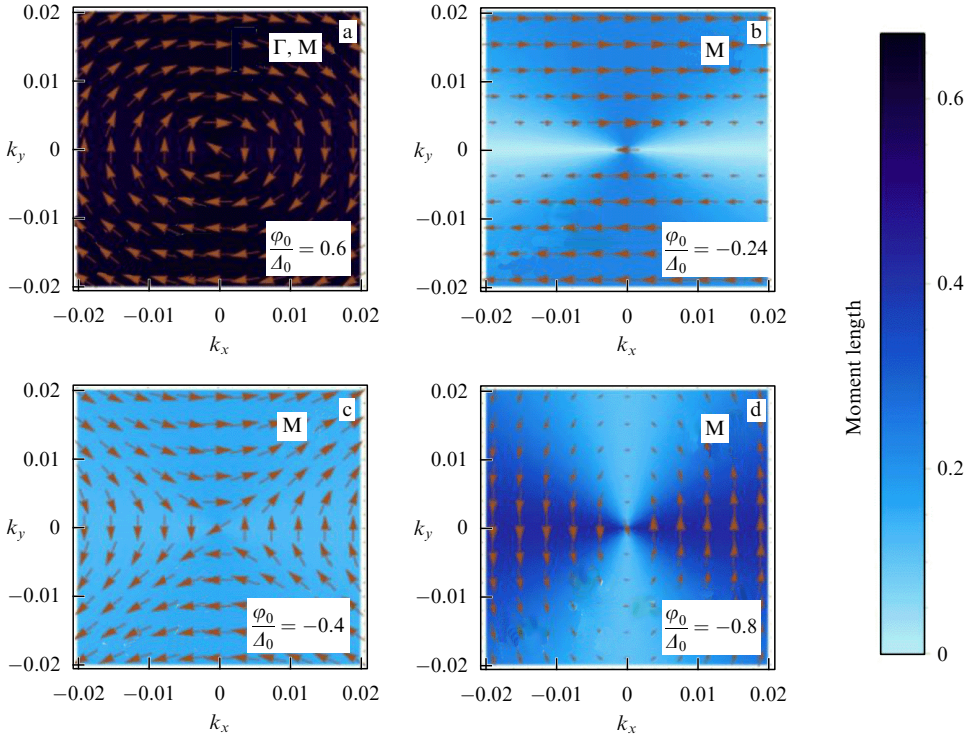

Figure 20. Conduction-band spin textures for topological surface states on an SnTe (111) surface [20]. The axes k_x, k_y are defined in the local M-point coordinate system (Fig. 13). The textures for Γ and M are similar for the downward band bending (a), except that for the Γ state the pattern is perfectly circular and for M it is slightly distorted (in the picture, the M-texture is shown). The difference becomes dramatic for the upward bending. Whereas the circular symmetry persists for the Γ state (yet at some point it flips the winding), the M texture goes through linear (b) and hyperbolic (c) distribution to the distorted circular (d) pattern with the opposite winding.

conduction (valence) branches are -1 ($+1$). The spin absolute value is $s \approx 0.8\mu_B$, in agreement with ab initio calculations. This represents a rather rigorous test for our model, as the magnitude of spin polarization is highly sensitive to surface band bending. With the change in φ_0, the spin texture in Γ does not alter its helical character, but the spin amplitude decreases, and for

$$\frac{\varphi_0}{\Delta_0} = -\frac{\sin^2(\theta_+/2) - \cos^2(\theta_-/2)}{\sin^2(\theta_+/2) + \cos^2(\theta_-/2)} \tag{23}$$

it vanishes altogether (with SnTe parameters from [17] we get an estimate $\varphi_0/\Delta_0 \approx -0.24$ for this critical value). With a further reduction in φ_0, the polarization changes its winding number to the opposite one. For the M point, the variation of φ_0/Δ_0 has much more dramatic consequences. As seen in Fig. 20, the spin texture evolves from helical with winding number -1 (panel (a)) through linear (b) and hyperbolic (c) textures, to finally a helical texture of winding number $+1$. The M spin polarization has a nontrivial out-of-plane component, which is in agreement with *ab initio* results for the M point [15]. While the magnitude of this component is approximately constant ($\approx 0.1\mu_B$), the in-plane moment is significantly larger for downward ($\varphi_0 > 0$), as opposed to upward ($\varphi_0 < 0$), band bending. The importance of these results is that, first, they show an explicit relation between the topological surface state and the parent bulk band structure, and, second, they open the way to control the surface spin polarization via changing the bulk parameters (e.g., in solid solutions) or via tuning the band bending with the top gate voltage.

6. Surface Ruderman–Kittel–Kasuya–Yosida interaction

As we have seen, the surface of a topological insulator hosts a chiral spin-polarized 2D electron gas. A natural test probe for a surface spin texture is a magnetic adatom. The magnetic moments, which are immersed in a Fermi gas, experience indirect interaction via virtual spin excitations, a mechanism known as the Ruderman–Kittel–Kasuya–Yosida (RKKY) interaction. In a 'normal' spin-unpolarized 3D electron gas, the interaction behaves as $\sin(2k_F r)/r^3$, decaying and oscillating between ferromagnetic and antiferromagnetic, where r is the separation of external spins and k_F is the Fermi wave number. It is clear that the spin locking in a topological state should have a pronounced effect on the RKKY interaction.

Using our analytical approach, we calculated this interaction for the (111) SnTe surface [23]. For simplicity, we considered the Γ valley contribution only. It is known experimentally [13] that in SnTe the Weyl–Dirac conical points for M and Γ are displaced in energy by as much as 170 meV. Hence, it should be possible to separate their contributions by choosing a proper Fermi energy.

To further simplify the derivation, we consider the case with no band bending, $\varphi_0 = 0$. Using Green's functions for the Weyl Hamiltonian (5), we find [23] that the interaction energy of two spins $\mathbf{S}_1$ and $\mathbf{S}_2$ contains various combinations of the spin-product terms $S_1^i S_2^j$. Let us choose the coordinate system such that the x-axis is aligned with the magnetic impurity separation vector, and the z-axis is normal to the surface. We find the Ising-type terms $S_1^x S_2^x$, $S_1^y S_2^y$, $S_1^z S_2^z$

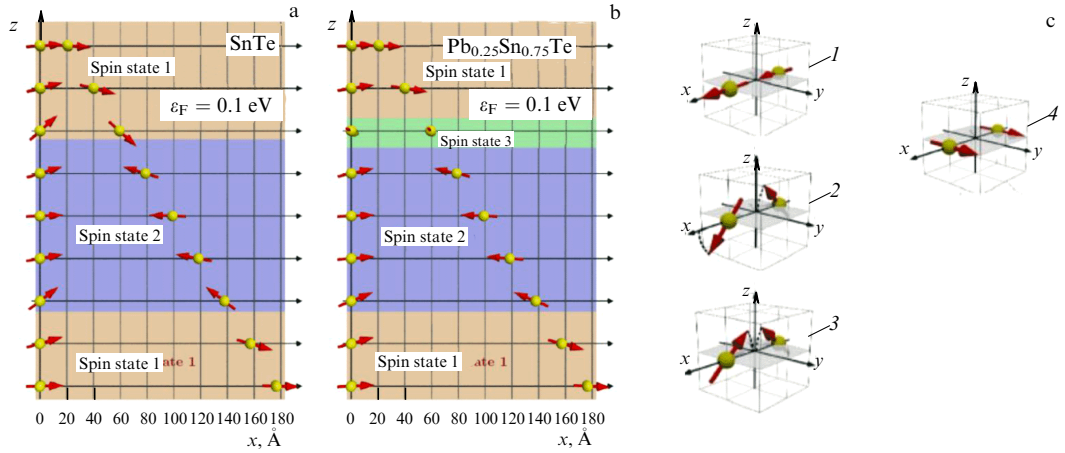

Figure 21. RKKY interaction on the tin terminated (111) surface of SnTe and $Pb_{0.25}Sn_{0.75}Te$. Equilibrium spin configurations are shown for different distances between the spins [23]. The Fermi energy is chosen at 0.1 eV. Whereas for SnTe there are two qualitatively different configurations (ferromagnetic with canting and antiferromagnetic with canting), for $Pb_{0.25}Sn_{0.75}Te$ there is a third spin configuration, which is ferromagnetic with spin alignment along the y-axis (see the sketch on the right).

which favor collinear ferromagnetic or antiferromagnetic spin alignments, as well as the Dzyaloshinskii–Moriya term $S_1^x S_2^z - S_2^x S_1^z$, which favors the two spins aligning in the xz plane perpendicularly to each other. Our microscopic model allows the calculation of the coefficients in front of all these terms, and hence we can determine which spin configuration is most favorable. When the Fermi energy runs through the Dirac–Weyl conical point (i.e., there are no free particles in the surface state), the RKKY interaction energy takes the form

$$E_{\varepsilon_F=0} = \frac{A}{r^3}\left[a(2S_1^x S_2^x - S_1^y S_2^y) + 2bS_1^z S_2^z\right], \qquad (24)$$

i.e., the interaction decays as $1/r^3$. Of course, coefficients a and b depend on the particular sublattice (Sn or Te) which hosts the spins. The Dzyaloshinskii–Moriya term appears when there are free carriers in the surface band:

$$E_{\varepsilon_F\neq 0} = \frac{\pi A}{r^2}\left[\sin(2k_F r)(aS_1^x S_2^x + bS_1^z S_2^z) - c\cos(2k_F r)(S_1^x S_2^z - S_2^x S_1^z)\right]. \qquad (25)$$

As we see in Eqns (24), (25), the RKKY interaction in a topological surface band is far more complex than in a spin-unpolarized Fermi gas. In Fig. 21, we show the mutual orientation of the two magnetic moments separated by the distance x due to the RKKY coupling on the Sn terminated (111) surface. The left panel (a) shows the case of SnTe and the right panel (b) is for $Pb_{1-x}Sn_xTe$ with $x = 0.25$. For the Fermi energy, we take $\varepsilon_F = 0.1$ eV. Figure 21c shows schematically typical configurations. Configurations 1 and 2 shown in Fig. 21c are referred to as 'spin state 1' in Fig. 21a, b and configurations 3 and 4 — as 'spin states 2 and 3' respectively. It is remarkable that a relatively small change in the parameter values between SnTe and $Pb_{0.25}Sn_{0.75}Te$ results in the appearance of a new spin state 3, where the two spins are aligned parallel to the y-axis (configuration 4 in Fig. 21c).

7. Conclusions

The analytical model of the inverted contact, which we introduced in 1985, is the first example of a 3D topological insulator. As understood later, SnTe-type semiconductors with an inverted band gap, which we proposed for that structure, are crystalline topological insulators.

The inverted contact model offers useful insight into the connection between a topologically nontrivial bulk band structure and topological surface states. SnTe-type materials provide a unique opportunity to trace this link analytically and to uncover how the surface state changes upon changing the bulk parameters. This is possible because the effective Dirac Hamiltonian allows an analytical derivation from the tight binding model of the bulk bands. Phenomenologically, the idea of inverted contact was recently exploited in the form of a 'topological boundary condition' [18] for derivation of the topological surface states.

However, the informative bulk-surface connection can be achieved only on the basis of a full description of the bulk bands. Our microscopic understanding of the bulk-boundary correspondence reveals the possibility of tailoring the surface spin texture by varying the bulk parameters or by adjusting the band bending. Using the microscopically derived picture, we calculated the RKKY interaction of magnetic impurities on the surface of an SnTe-type topological insulator. We found that this interaction is far more complex than in a 'normal' Fermi gas, and we proposed a direct connection between this interaction and the parameters of the bulk bands. This opens the way for a concerted design of complex surface magnetic structures using topologically protected spin polarized states.

Acknowledgements

I am deeply thankful to my teachers and coworkers, the late Yu V Kopaev and the late B A Volkov, who shared with me their knowledge of and admiration for physics. This talk is devoted to their memory. I am also indebted to Sam Shallcross and Nicolas Klier for our joint work.

References

1. Hasan M Z, Kane C L *Rev. Mod. Phys.* **82** 3045 (2010)
2. Ando Y *J. Phys. Soc. Jpn.* **82** 102001 (2013)
3. Qi X-L, Zhang S-C *Rev. Mod. Phys.* **83** 1057 (2011)
4. Volkov B A, Pankratov O A *JETP Lett.* **42** 178 (1985); *Pis'ma Zh. Eksp. Theor. Fiz.* **42** 145 (1985)

5. Tchoumakov S et al. *Phys. Rev. B* **96** 201302(R) (2017)
6. Inhofer A et al. *Phys. Rev. B* **96** 195104 (2017)
7. Hsieh D Y et al. *Nature* **460** 1101 (2009)
8. Liu J, Duan W, Fu L *Phys. Rev. B* **88** 241303(R) (2013)
9. Fu L *Phys. Rev. Lett.* **106** 106802 (2011)
10. Teo J C Y, Fu L, Kane C L *Phys. Rev. B* **78** 045426 (2008)
11. Tanaka Y et al. *Nature Phys.* **8** 800 (2012)
12. Hsieh T H et al. *Nature Commun.* **3** 982 (2012)
13. Tanaka Y et al. *Phys. Rev. B* **88** 235126 (2013)
14. Yan C et al. *Phys. Rev. Lett.* **112** 186801 (2014)
15. Shi Y et al. *Phys. Rev. B* **90** 235114 (2014)
16. Volkov B A, Pankratov O A *Sov. Phys. JETP* **48** 687 (1978); *Zh. Eksp. Theor. Fiz.* **75** 1362 (1978)
17. Volkov B A, Pankratov O A, Sazonov A V *Sov. Phys. JETP* **58** 809 (1983); *Zh. Eksp. Theor. Fiz.* **85** 1395 (1983)
18. Zhang F, Kane C L, Mele E J *Phys. Rev.* **86** 081303(R) (2012)
19. Buczko R, Cywiński L *Phys. Rev. B* **85** 205319 (2012)
20. Klier N, Sharma S, Pankratov O, Shallcross S, arXiv:1702.08784
21. Pankratov O A, Pakhomov S V, Volkov B A *Solid State Commun.* **61** 93 (1987)
22. Pankratov O A *Semicond. Sci. Technol.* **5** S204 (1990)
23. Klier N, Sharma S, Pankratov O, Shallcross S, arXiv:1711.10760
24. Taskin A A et al. *Phys. Rev. B* **89** 121302(R) (2014)