More on topological insulators: the experimental situation

The simple model explored in lecture 23 allows us to draw some important qualitative conclusions concerning the phenomenon of topological insulators. First, let's consider the relation to the QHE. This may not be immediately obvious, since most discussions of the QHE neglect the existence of a crystalline lattice (except in so far as it leads to the replacement of the real electron mass m by an effective mass m^*); this is usually experimentally realistic since when both the inter-electron distance and the magnetic length are very large compared to the lattice spacing, and thus the relevant electrons never get anywhere near the edge of the FBZ and a single-band picture appears to be quite adequate. On the other hand, the difference between a TI and a standard band insulator appears to depend essentially on the behavior of the band structure over the whole of the FBZ. Thus, at first sight there is not much relationship between the two phenomena.

However, it is perfectly possible to analyze the (integral) QHE taking into account completely the effects of the periodic crystalline lattice, and this was done in a famous 1982 paper[1] by Thouless and co-workers (usually known as TKNN). They found that under certain conditions the effect of the magnetic field is to split the energy levels of the system into sub-bands, with each band being characterized by a topological quantum number similar (but not identical) in form to (23.25). They further sketched an argument (filled out in more detail a few years later by Hatsugai[2]) that in this case there would automatically be chiral edge states (that is, states which can propagate in one direction only as in the "free-space" QHE) which would interpolate between the bands. The TKNN analysis neglects the spin degree of freedom, but it was subsequently realized that in the case of a system in zero magnetic field, the spin-orbit coupling could in effect play the role of an external field whose "sense" however depends on the spin. Indeed the model analyzed in lecture 23 can be viewed in precisely this way: the "up" and "down" spin electrons are completely decoupled, and each behaves effectively as an independent QHE system, with a Chern number ("TKNN number") which is +1 for the up spins and −1 for the down ones. Since time reversal (TR) inversion is not broken, there can be no charge current flowing around the edges of the system, but nothing prevents a spin current from doing so. (Recall that the expression for the spin current, $\sum_i \sigma_i p_i/m$, is *even* under TR).

We must briefly discuss the question of the "protection" of the edge modes by time-reversal invariance (TRI). In the simple model of lecture 23, this is rather trivial: In fig. 23.10, we see that two modes cross the gap completely, thus interpolating between the valence and conduction bands and allowing metallic behavior at the surface. Suppose now that some perturbation mixes the two modes, then by the usual "level-repulsion" arguments we would expect the picture to change to that shown in fig. 1, so that there is now a nonzero gap and the surface states behave as an insulator. The crucial point, now, is that because the modes have different values of the (real) spin projection s_z, no perturbation which com-

[1] Thouless et al., PRL **49**, 405 (1982)

[2] Hatsugai, PRL **71**, 3697 (1993)

mutes with s_z (e.g. scattering by nonmagnetic impurities) can mix the states in this way. To mix them would require, for example, a magnetic field, in the direction parallel to the plane (hence perpendicular to s_z) and this then suggests that such a field should have a dramatic qualitative effect.

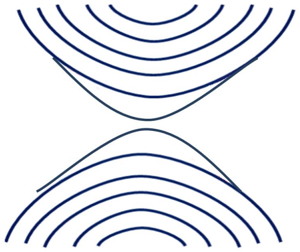


Fig. 1

Before moving beyond the simple model of lecture 23, let's examine the rather delicate question of the symmetry and periodicity of the Bloch states. Consider the Hamiltonian $\hat{\mathcal{H}}(\boldsymbol{k})$ associated with the Bloch state $\boldsymbol{k}$, which according to (23.19) is a matrix in the pseudospin space (for the moment we suppress the spin degree of freedom):

$$\hat{H}(\boldsymbol{k}) = -\hat{\boldsymbol{\sigma}}_k \cdot \boldsymbol{\mathcal{H}}_k \tag{1}$$

with $\boldsymbol{\mathcal{H}}_h$ given by the expression (23.20). If we extend $\boldsymbol{k}$ beyond the FBZ, then we see that the crystal periodicity implies that if $\boldsymbol{K}$ is any reciprocal lattice vector then

$$\hat{H}(\boldsymbol{k}+\boldsymbol{K}) \equiv \hat{H}(\boldsymbol{k}) \tag{2}$$

and in particular both for each of the four states $\boldsymbol{k}_i, i = 1, 2, 3, 4$ lying at the center of an FBZ edge and for the four "corner" states ($i = 5, 6, 7, 8$) we have

$$\hat{H}(\boldsymbol{k}_i) \equiv \hat{H}(-\boldsymbol{k}_i) \tag{3}$$

We now come to a delicate point: In the groundstate, at each of the eight points $\boldsymbol{k}_i(i = 1, ..8)$ the pseudospin $\boldsymbol{\sigma}$ is along the negative z-axis, and according to the argument given above the phase of the corresponding wave function u_{kn} (actually v_k) can be unambiguously defined. Moreover, according to the argument below eqn. (23.25) this phase should (up to an arbitrary additive constant) be simply equal to the angle of the $\boldsymbol{k}_i$ in question relative to (say) the x-axis:

$$\varphi(\boldsymbol{k}_i) = \theta_i \tag{4}$$

so that, explicitly, $\varphi = 0, \pi/2, \pi, 3\pi/2$ for $i = 1, ..4$ and $\varphi = \pi/4, 3\pi/4, 5\pi/4, 7\pi/4$ for $i = 5..8$. Thus, we see that according to this argument $\varphi(-\boldsymbol{k}_i) = \varphi(\boldsymbol{k}_i) + \pi$. On the other hand, for each of the $\boldsymbol{k}_i$ there exists a reciprocal lattice vector $\mathbf{K}$ such that $-\boldsymbol{k}_i = \boldsymbol{k}_i + \mathbf{K}$, and thus according to eqn. (7) of lecture 22 (periodicity of the Bloch wave functions in $\mathbf{K}$) we should have $\varphi(-\boldsymbol{k}_i) \equiv \varphi(\boldsymbol{k}_i)$! Thus we appear to reach a contradiction.

At first sight it is tempting to argue that the "absolute" phase φ of the Bloch wave function u_{km} is anyway physically meaningless: all that matters is the **relative** phase of the "up" and "down" componets u_k and v_k, which is what determines transverse components of $\langle\boldsymbol{\sigma}\rangle$ according to the standard textbook formula $\langle\sigma_{kL}\rangle = Re(u_k^* v_k), \langle\sigma_{yk}\rangle = lm(u_k^* v_k)$. Since $\langle\boldsymbol{\sigma}\rangle$ is well-defined (not just for the special points $\boldsymbol{k}_i$ but across the whole FBZ) this means that we need not worry about any ambiguity in the phases of the u_k and v_k individually.

We return to this argument below, but note that while it may give us an escape from the contradiction (at least as regards the bulk states) for a single (real)-spin state (say $S_z = +1$) it comes back as soon as we include the $S_z = -1$ component, since the expectation value of (e.g.) the Fourier-transformed operator S_{xk} is certainly physically meaningful and, while it does not depend on the absolute phase of $v_{k\uparrow}$ and $v_{k\downarrow}$, does depend on their **relative** phase. Hence the latter, at least, must be physically meaningful.

But now we note an interesting point: Since the $L_z = -1$ state has (k_x, k_y) replaced by $(k_x, -k_y)$, the phase of the down-(real)-spin state $S_z = -1$ has its phase reversed with respect to that of $S_z = +1$, and thus the relative phase of the $S_z = +1, -1$ states is just twice the "absolute" phase of the $S_z = +1$ state: $\varphi_{rel}(\boldsymbol{k}_i) = 2\theta_i$. Thus $\varphi_{rel}(-\boldsymbol{k}_i) = +\varphi_{rel}(\boldsymbol{k}_i)$ and we are no longer in contradiction with the **K**-periodicity given by eqn. (7) of lecture 22! This argument suggests that, more generally, various ambiguities occurring in the theory of topological insulators may be clarified if rather than focusing on the behavior of a single spin state, we **compare** the behavior of the $S_z = \pm 1$ states.

We still need to discuss the question: for say $S_z = +1$, is the phase of the u_k and v_k individually (as distinct from their relative phase) meaningful? Consider the simple case of a single filled band described by (spinless, pseudospinless) Bloch states $u_{\boldsymbol{k}}(r)$. The many-body GSWF is the Slater determinant

$$\Psi(r_1, r_2, ..., r_N)) = \mathcal{A} \prod_{k \in FBZ} u_{\boldsymbol{k}}(\boldsymbol{r}_i) \text{exp } i\boldsymbol{k} \cdot \boldsymbol{r}_i \tag{5}$$

Suppose now we multiply each $u_{\boldsymbol{k}}(r)$ by some arbitrary phase factor exp $i\varphi(\boldsymbol{k})$ (where $\varphi(\boldsymbol{k})$ need not be a continuous function of $\boldsymbol{k}$). We see that the effect is simply to multiply the many-body wave function (5) by the overall factor $\prod_{k \in FBZ} \text{exp } i\varphi(k) \equiv \text{exp } iA, A \equiv \sum_{k \in FBZ} \varphi(\boldsymbol{k})$, which clearly can have no physical significance! A similar argument should prima facie apply to the case considered in lecture 23 provided each spinor is multiplied by an overall phase factor.

This argument is not quite as strong as it looks. What is the physical situation described by eqn (5)? It has to be a system confined, in real space, to an appropriate box, *with the single-particle wave functions subjected* to *periodic boundary conditions*. For such a "theoretical" system, the conclusion that the overall phases of the u_{kn}'s do not matter is probably correct. But in real life we are interested in physical systems which have real surfaces (edges) and now the (relative) phases $\varphi(\boldsymbol{k})$ do matter – close to the edges the Slater determinant of Bloch waves given by eqn. (5) is misleading; we still of course have a scalar determinant, but generically of a linear superposition of Bloch waves, when the relative phases matter. So we cannot shrug off the problem quite so easily...

Returning, then, to the dilemma posed above, there seem to be two obvious possible ways of resolving it. One, which seems to be the majority preference, is to require that eqn. (4) be maintained; since if we also require $u_n(\boldsymbol{k})$ to be continuous throughout the

FBZ, we then get an unwanted singularity at the origin, we then need to define $u_n(\boldsymbol{k})$ to be continuous but different in overlapping regions of the FBZ and then "glue them together"; this is the procedure favored eg by Bernevig in his book. The alternative procedure would seem to be simply to abandon the constraint of periodicity in $\mathbf{K}$ (which is in some sense a special case of the former approach, with the "glueing" occurring along the FBZ edges). To my mind this is the simpler convention, and I will follow it. Note that none of the above is relevant to the definition of the Chern number, which is entirely in terms of the direction of $\boldsymbol{\sigma}$.

We now consider how to generalize the simple model of lecture 23. An important point, particularly when we come to 3D generalization, is that in the general case s_z does not commute with $\hat{H}$. This means that we can no longer treat the $s_z = \pm 1$ states separately and assign them individual Chern numbers: however, time reversal invariance may well still hold (we shall assume it does[3]) and may then be expected to produce somewhat similar results.

Indeed, a good deal of information can be obtained from invariance, if it occurs, under space inversion $\hat{P}$ and time reversal $\hat{T}$. For this purpose, it is convenient to combine the separate spinors for $s_z = \pm 1$ into a single four-component spinor whose components are labeled by $s_z = \pm 1$, $\sigma_z = \pm 1$. The parity operation $\hat{P}$ then simply changes $\boldsymbol{k}$ to $-\boldsymbol{k}$, without any effect on the structure of the spinor, while time reversal $\hat{T}$ also changes $\boldsymbol{k}$ to $-\boldsymbol{k}$ but at the same time effects the operation $-i\hat{s}_y$ on the (real) spin degree of freedom, plus complex conjugation. The Hamiltonian is now, for general $\boldsymbol{k}$, a 4×4 matrix; however, in the case of the four points $\boldsymbol{k}_i$ eqn. (3) plus TRI inplies that it separates into two 2×2 components describing two Kramers doublets.

How to distinguish a B1 from a T1? A first shot might involve the observation that for the TI of lecture 23 the phases of the $(\sigma_z = -1)$ components of the spinor are, for $s_z = +1, 0, \pi/2, \pi, 3\pi/2$ whereas for $s_z = -1$ they are $0, -\pi/2, -\pi, -3\pi/2$ (or equivalently $0, 3\pi/2, \pi, \pi/2$) i.e. the "chirality" is reversed. By contrast, for the band insulator the $(\sigma_z = +1)$ component for both s_z all have phase 0 (or constant). Hence one might try to take the quantity (where ψ_i denotes the orbital component only)

$$\sum_{i=1}^{4} (\psi_i^{(s_z=+1)}, \psi_i^{(s_z=-1)}) \tag{6}$$

which takes the value +1 for the BI and 0 for the TI, as an indicator of the differences. However, using the form of the TR operator $\hat{T}$ (which, recall, reverses $\boldsymbol{k}$, operates with $-i\hat{s}_y$ on the spinor and complex conjugates) we see that (6) is equivalent to

$$Q \equiv \sum_{i=1}^{4} (\psi_i^{(1)} \hat{T} \psi_i^{(2)}) \tag{7}$$

[3] Note in particular that while inversion invariance automatically fails near a surface, TRI may well still hold there.

where $\psi_i^{(1)}, \psi_i^{(2)}$ are the two occupied eigenstates. It is therefore plausible to regard the expression Q as an appropriate indicator of the TI/BI distinction: $Q = 1$ for a BI, 0 for a TI. Or equivalently we can define

$$\nu = 1 - Q$$

so that $\nu = 0$ for a BI and 1 for a TI. As far as I know this possible definition has not been explored in the literature.

In any case, a proper characterization of TI's versus band insulators needs to somehow reflect the fact that the pseudospin configuration changes non-trivially between the origin ($\boldsymbol{k} = 0$, "Γ-point") and the edge of the FBZ. Thus we expect that we need to involve somehow not only the TRI points on the edge of the FBZ but also those at $\boldsymbol{k} = 0$. To motivate the standard approach[4], let us proceed as follows: Consider the four TRI points (i.e. the points such that $\hat{\mathcal{H}}(\boldsymbol{k}) = \hat{\mathcal{H}}(-\boldsymbol{k})$ where $\hat{\mathcal{H}}$ is a matrix in pseudospin space):

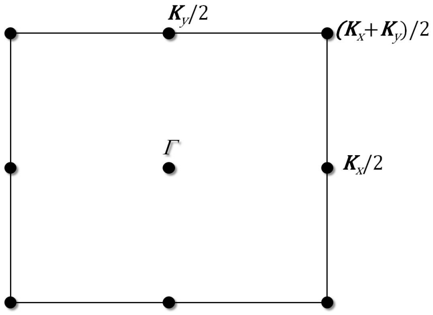


Fig. 2

$\Gamma_1 : \boldsymbol{k} = 0$

$\Gamma_2 : \boldsymbol{k} = \boldsymbol{G}_x/2$

$\Gamma_3 : \boldsymbol{k} = (\boldsymbol{G}_x + \boldsymbol{G}_y)/2$

$\Gamma_4 : \boldsymbol{k} = \boldsymbol{G}_y/2$

with $\boldsymbol{G}_x \equiv (G/2)\hat{\boldsymbol{x}}, \boldsymbol{G}_y \equiv (G/2)\hat{\boldsymbol{y}}, G \equiv 2\pi/a$

Staying for the moment with the model of lecture 23, let's consider the argument of the quantity $\mu_i \equiv \psi_\uparrow^* \psi_\downarrow(\Gamma_i)$, where $\psi_{\uparrow,\downarrow}$ are the orbital were functions associated with the (real) spin states $s_z = \pm 1$. Since at Γ_1 we can (by convention) choose the phases of $\psi_{\uparrow,\downarrow}$ to be identical, and at the other 3 Γ-points $\psi_{\uparrow\downarrow} \propto \exp \pm i\theta$ where θ is the angle made by a line to the $\boldsymbol{k} = 0$ point with (say) the x-axis, we have for $\lambda_i \equiv \arg \mu_i$

$$\lambda_1 = 0,\ \lambda_2 = 0,\ \lambda_3 = \pi/2,\ \lambda_4 = \pi \tag{8}$$

By contrast, for a band insulator all four λ_i are zero. How to express this difference quantitatively?

The answer given in the literature[4] goes as follows:

For each point TRI point Γ_i $(i = 1, ...4)$ consider the 2D matrix

$$w_{\mu\nu}^{(i)} \equiv \langle u_\mu(\boldsymbol{k}_i)|\hat{T}|u_\nu(-\boldsymbol{k}_i)\rangle$$

where $u_\mu(\mu = 1, 2)$ labels the two spinor states occupied in the groundstate, and $\hat{T}$ is the time reversal operator. Since $\hat{T}^2 = -1$ for fermions, and the points Γ_i are TR invariant,

[4] e.g. Fu and Kane, PR B **7**4. 195312 (2006)

$w_{\mu\upsilon}$ is an antisymmetric matrix. We now define

$$\delta_i \equiv w^{(i)}_{\mu\nu}/\sqrt{w^{(i)2}_{\mu\nu}} \tag{9}$$

so that δ_i can take only the values ±1. The tricky part is the treatment of the square root: we can define this so as to stay on the same branch throughout the whole of the FBZ, and, if we do this for the case of the model of lecture 23, then three of the four δ_i's will be $+1$ and the fourth -1 [not obvious; further discussion in the lecture]. (For the band insulator it is clear that all four δ_i's can be chosen to be $+1$). The general definition of a "Z_2 invariant" which will distinguish TI's and BI's is then

$$(-1)^\nu \equiv \prod_{i=1}^{4} \delta_i \tag{10}$$

so that $\nu = 0$ for a band insulator and 1 for a topological insulator.

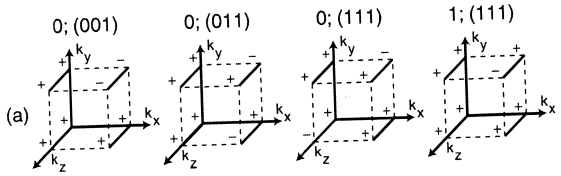


Fig. 3

How then to generalize this idea to three dimensions? The simplest scheme goes as follows: We add four more TRI points in the (now 3D) FBZ by translating each of the four Γ_i by $\boldsymbol{G}_z/2$ where $\boldsymbol{G}_z \equiv G\hat{\boldsymbol{z}}$. In this way the 8 Γ_i's form a cube, and by arguments similar to those for the 2D case we can assign a number $\delta_i = \pm 1$ to each of them: see fig 3 (taken from Fu and Kane 2006). We can now define (a) a Z_2 quantum number for each pair of faces separately (call these ν_1, ν_2, ν_3) but also a fourth quantum number $\nu_o = \pm 1$ which has no 2D analog:

$$\nu_o \equiv \prod_{i=1}^{8} \delta_i$$

We note that if $\nu_o = 0$ (the case of a so-called "weak 3D TI") then the quantities $\nu_i (i = 1, 2, 3)$ are formally independent of which faces of the cube we consider, whereas in the case $\nu_0 = 1$ ("strong 3D TI") they are different; we then say that $\nu_i = 1$ by convention (cf. (a - d) in fig. 3).

In general it turns out [not obvious!] that the nature of the surface states is strongly dependent on the value of ν_o. For the case $\nu_o = 0$ (weak TI) either there are no edge states bridging the gap, or they exist in the ideal case but are not robust against scattering by e.g. magnetic impurities. For the $\nu_o = 1$ (strong TI) the surface states are much more interesting: on each surface there exists one (or in general case an odd number) of "Dirac cones" with an energy related to momentum $\boldsymbol{k}$ and (real) spin $\boldsymbol{s}$ by the formula

$$E(\boldsymbol{k}) = \boldsymbol{s} \cdot \boldsymbol{k} \times \hat{\boldsymbol{n}}$$

where $\hat{\boldsymbol{n}}$ is the normal to the surface: note that this expression is correctly invariant under both P and T. Thus, we get "spin-momentum locking" as shown in fig. (9) below.

Let's now turn to the experimental evidence for topological insulators. This, like the theory, is already a vast subject, and I can only scratch the surface. By definition, a TI, whether 2D or 3D, is insulating in bulk, and it is usually difficult to see anything spectacular. One exception is the quantum Hall effect (in 2D systems); because of the mixing of conduction- and valence-band-states in a TI (as for that matter in a band insulator with substantial spin-orbit interaction) the filling of the Landau levels as a function of doping is anomalous. This will not be discussed here. The experiments done[5] on the surfaces of (putative) TI's are mostly of two types, surface transport and ARPES. The original 2D system tested (5) for TI-type behavior (according to the predictions of Bernevig et al.) is a quantum well heterostructure consisting of states of CdTe and HgTe: see fig (4). In the experiments the total width d_{tot} was always $\leq 40\ nm$, so since the effective mass m^* is not very different from m, the first "transverse" excited state has excitation energy $E \sim 1K$, so at the temperature of the experiment ($\sim 50\ mK$) it should be legitimate to treat the whole system as effectively 2D. However, crudely speaking, CdTe is a band insulator whereas HgTe has TI-like characteristics: more quantitatively. Bernevig et al. predicted that for d less than a critical value d_c which they estimated as $\sim 6\ nm$, the heterostructure should behave as a simple band insulator, whereas for $d > d_c$ it should be a 2D TI. In either case we can tune the Fermi energy E_F by varying the gate voltage and hence the carrier density n; of course, when E_F lies within either the valence or the conduction band we would expect the system to behave as a metal independently of its TI/BI character, so the interesting case is when E_F lies in the bulk gap; in that case the BI should show insulating behavior, and indeed the experiments done on a sample width $d_c = 4.5\ nm$ ($< d_c$) show a (longitudinal) resistance R_{xx} several orders of magnitude greater than h/e^2 (and consistent with ∞).

What do we expect in the TI phase? On inspection of fig. (5), we see that there is exactly *one* state of the correct chirality on each edge, so according to the Landauer formula we would predict $I = (2e^2/h)V$ or $R_{xx} = h/2e^2 \approx 12k\Omega$ And, lo and behold, for $d = 8\ nm$ ($> d_c$) and E_F in the bulkless gap, a substance close to this volume is seen in the experiments!

Is this a fluke coincidence? Two circumstantial pieces of evidence are (a) that R_{xx} is independent of the sample width, which indicates that it (or rather the conductance $R_{xx}^{(-1)}$) is likely a surface effect (b) the fact that the conclusion $\sum_{xx}$ is

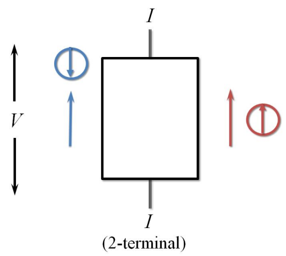


Fig. 5

[5] König et al., Science 319, 766 (2007)

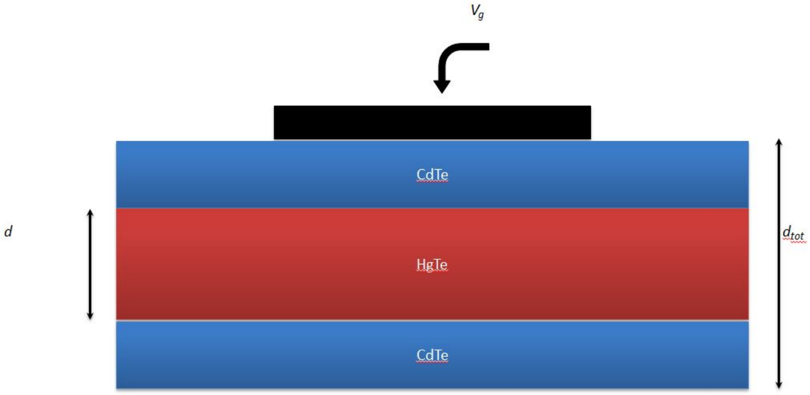


Fig. 4

strongly supported by even small magnetic fields $\sim 0{\cdot}05T$ (see fig. 6). This strongly suggests that the suppression is a consequence of the breaking of time reversal invariance. Moreover, the fact that the suppression is a strong function of the direction of the field, being essentially absent for fields parallel to the 2D surface, indicates that the effect does not have to do with the Zeeman coupling but is a consequence of the *orbital* effect of the field.

An even more convincing surface conductance experiment[6] by the same (Würzburg) group used the 6-terminal geometry shown in fig. 7. For this kind of setup the Landauer-Büttiker theory gives a relation between the currents I_i in the i-th lead and the voltages V_j in the j-th lead:

$$I_i = (e^2/h) \sum_j (T_{ji} V_i - T_{ij} V_j) \tag{11}$$

when T_{ij} is the total transmission amplitude ($T_{ij} = \sum_\nu T_{ij}^{(\nu)}$ where ν is a channel label) from lead i to lead j. For the quantum hall effect (a "chiral" case) the quantity $T_{i,i+1}$ is nonzero but $T_{i+1,i} = 0$, and we can work out the consequences. For the TI ("QSH") case, by contrast, there is exactly one channel of *each* chirality (with of course opposite spins), so we have a quite different result:

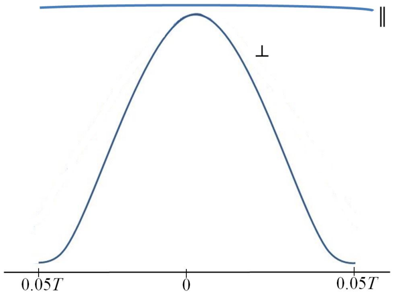


Fig. 6

$$T_{i,i+1} = T_{i+1,i} = 1 \;\; \text{; all other } T_{ij} \text{ zero.}$$

We can substitute this result into (11) and solve to find the currents I_{ij} between leads i and j in terms of the voltages V_{kl} applied between leads k and l. In particular we find for

[6] Roth et al., Science **325**, 294 (2009)

a 4-terminal measurement (quite differently from the predictions for a simple QH state) for $R_{ij,kl} \equiv V_{kl}/I_{ij}$

$$R_{14,14} = 3h/e^2, \quad R_{14,23} = h/2e^2$$

This is well verified in the experiment.

The other major class of existing experiments uses ARPES. For reasons of adequate intensity, this has to be done on surfaces rather than edges, i.e., we require a bulk 3D system; the simplest example to date is Bi_2Se_3. One can and does measure the energy spectrum, thereby confirming the "level-crossing" picture of lecture 23, fig. 10 (see fig. 8). However, even more interestingly, it turns out that by measuring the spin polarization of the ejected photo electrons one can infer the average spin polarization of the surface states from which they were ejected. The result for $\langle S_i \rangle (i = x, y, z)$ as a function of k_y is shown schematically in fig. (9). This confirms the prediction (cf. above) that the spin of the surface states is perpendicular both to the wave vector $\boldsymbol{k}$ and to the surface normal $\hat{\boldsymbol{n}} \equiv \hat{\boldsymbol{z}}$: thus the picture is, as predicted theoretically, that shown in fig. (9a) for a view perpendicular (in spin space) to the surface, and that of fig. (9b) for a view from the directions parallel to the surface.

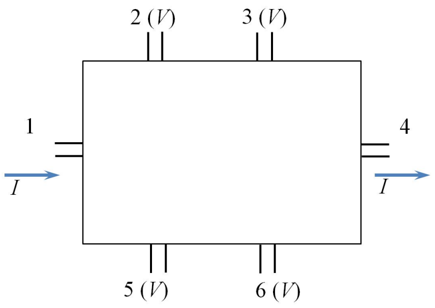


Fig. 7

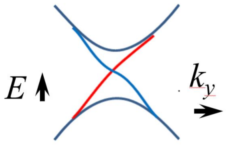
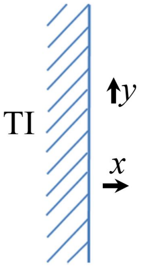


Fig. 8



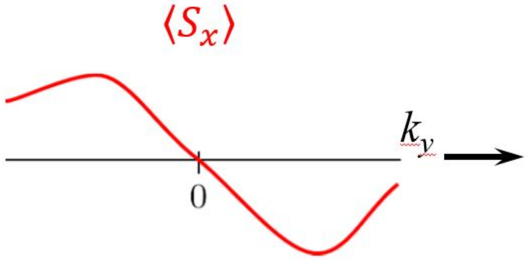


Fig. 9a

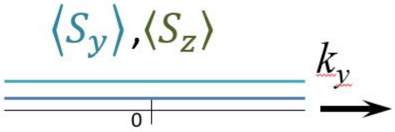


Fig. 9b