An Introduction to the Quantum Spin Hall Effect

In a strong magnetic field, two dimensional (2D) electron systems exhibit the integer quantum Hall effect, which consists of an insulating state and edge modes with Hall conductance $\sigma_{xy} = ne^2/h$, where n is an integer. The quantum Hall state is fundamentally different from that of an ordinary insulator in that the electronic structure of a quantum Hall insulator cannot be continuously deformed into that of a "trivial insulator" without closing the energy gap. The quantum spin Hall (QSH) effect is also an insulating state with conducting edge channels. However, unlike the quantum Hall effect, the QSH effect occurs in the absence of a magnetic field. This paper will discuss the QSH effect, starting with the Haldane model for a quantum Hall system. We will show that the Hall conductivity for the Haldane model is nonzero, and we will apply the Haldane model to Kane and Mele's model for a QSH insulator.

1   The Berry Phase

Let's first review the concept of a Berry phase. When a vector is parallel transported in a loop on a curved surface, the vector will acquire an anholonomy angle. Analogously, in quantum mechanics, a state indexed by a continuous set of parameters $\mathbf{R}$ that is adiabatically evolved along a closed parametrization of $\mathbf{R}$ will acquire a Berry phase [1]

$$\gamma_n = i \oint \langle n, \mathbf{R}|\nabla_{\mathbf{R}}|n, \mathbf{R}\rangle \cdot d\mathbf{R} \tag{1}$$

To understand the curved surface on which $|n, \mathbf{R}\rangle$ is being parallel transported would require the concept of a fiber bundle, which is outside the scope of this paper. The quantity in the integrand $\mathbf{A}_n(\mathbf{R}) = i\langle n, \mathbf{R}|\nabla_{\mathbf{R}}|n, \mathbf{R}\rangle$ is known as the Berry connection, which is not gauge invariant. We can construct a gauge invariant quantity known as the Berry curvature

$$\mathbf{\Omega}_n(\mathbf{R}) = i\nabla_{\mathbf{R}} \times \langle n, \mathbf{R}|\nabla_{\mathbf{R}}|n, \mathbf{R}\rangle \tag{2}$$

Then, using Stoke's Theorem, we can write the Berry phase as

$$\gamma_n = \int_S \mathbf{\Omega}_n(\mathbf{R}) \cdot d\mathbf{S} \tag{3}$$

where S is a surface whose boundary is the closed curve parameterized by $\mathbf{R}$. Given eigenstates of energy $E_n(\mathbf{R})$ of a Hamiltonian $H(\mathbf{R})$ parameterized by $\mathbf{R}$, the Berry curvature can easily be rewritten in another form [2]

$$\mathbf{\Omega}_n(\mathbf{R}) = -\text{Im} \sum_{m \neq n} \frac{\langle n, \mathbf{R}|(\nabla_\mathbf{R} H)|m, \mathbf{R}\rangle \times \langle m, \mathbf{R}|(\nabla_\mathbf{R} H)|n, \mathbf{R}\rangle}{(E_m(\mathbf{R}) - E_n(\mathbf{R}))^2} \tag{4}$$

This form is useful because one can immediately see that the sum of the Berry curvatures of all energy levels is zero.

We can gauge away the Berry phase if we replace $i\nabla_\mathbf{R}$ by $i\nabla_\mathbf{R} + \mathbf{A}_n(\mathbf{R})$ everywhere in the Hamiltonian [3]. In this sense, $\mathbf{A}_n(\mathbf{R})$ is similar to the magnetic vector potential, and since $\mathbf{\Omega}_n(\mathbf{R}) = \nabla_\mathbf{R} \times \mathbf{A}_n(\mathbf{R})$, the Berry curvature acts like a magnetic field in the parameter space of $\mathbf{R}$. If we imagine a crystalline solid, we can let the parameter $\mathbf{R}$ be the crystal momentum $\mathbf{k}$ of Bloch states. Then, we can think of the Berry curvature as a k-space magnetic field, and so the semiclassical equations of motion are [4]

$$\hbar \frac{d\mathbf{k}}{dt} = -e\left[\mathbf{E} + \mathbf{v} \times \mathbf{B}\right] \tag{5}$$

$$\hbar\mathbf{v} = \nabla_\mathbf{k} E_n(\mathbf{k}) + e\mathbf{E} \times \mathbf{\Omega}_n(\mathbf{k}) \tag{6}$$

where $e = |e|$, $\mathbf{v}$ is velocity, $\mathbf{E}$ is an applied electric field, and $\mathbf{B}$ is a magnetic field. To get the current, we sum $\mathbf{v}$ over all filled states. For a full band, summing $\nabla_\mathbf{k} E_n(\mathbf{k})$ will yield zero, but the "anomalous velocity" $e\mathbf{E} \times \mathbf{\Omega}_n(\mathbf{k})$ does not necessarily add to zero. Also notice that the anomalous velocity is perpendicular to $\mathbf{E}$, so it contributes to σ_{xy}.

To be more exact, we can compute the Hall conductance for a 2D insulator directly using the Kubo-Greenwood Formula [5]. The Kubo-Greenwood Formula is almost of the same form as the Berry curvature given in Eq. (4), so it does not take much effort to show that the Hall conductance is the integral of the Berry curvature over all filled states

$$\sigma_{xy} = \frac{e^2}{h} \frac{1}{2\pi} \sum_n \int_{BZ} dk_x dk_y \Omega_{xy}(\mathbf{k}) \tag{7}$$

$$\Omega_{xy}(\mathbf{k}) = \frac{\partial A_y(\mathbf{k})}{\partial k_x} - \frac{\partial A_x(\mathbf{k})}{\partial k_y} \tag{8}$$

where BZ is the Brillouin zone. The Berry curvature $\Omega_{xy}(\mathbf{k})$ depends on the band index n, and the sum over n is taken only over filled bands. Another way to derive this relationship between Hall conductance and Berry curvature is to treat the applied electric field as a vector potential $-\mathbf{E}t$ in time-dependent perturbation theory [6].
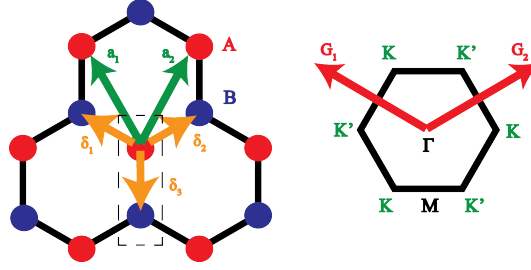
2 A Honeycomb Lattice

Figure 1: Honeycomb Lattice and Brillouin Zone

The Haldane model involves a honeycomb lattice, so lets introduce the theory of electrons on such a lattice [7]. The honeycomb lattice is made of two interpenetrating triangular sublattices, hereby denoted as the A and B sublattices (see Fig. 1). Each triangular sublattice can be constructed by the primitive lattice vectors

$$\mathbf{a}_1 = a\left(-\frac{\sqrt{3}}{2}, \frac{3}{2}\right) , \quad \mathbf{a}_2 = a\left(\frac{\sqrt{3}}{2}, \frac{3}{2}\right) \tag{9}$$

where a is the distance between a point on the A sublattice and its nearest neighbors on the B sublattice. The vectors between a point and its nearest neighbors are

$$\boldsymbol{\delta}_1 = a\left(-\frac{\sqrt{3}}{2}, \frac{1}{2}\right) , \quad \boldsymbol{\delta}_2 = a\left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right) , \quad \boldsymbol{\delta}_3 = a\left(0, -1\right) \tag{10}$$

Define a unit cell to encompass a point on A and its nearest neighbor on B at $\boldsymbol{\delta}_3$. The reciprocal lattice vectors are

$$\mathbf{G}_1 = \frac{2\pi}{a}\left(-\frac{1}{\sqrt{3}}, \frac{1}{3}\right) , \quad \mathbf{G}_2 = \frac{2\pi}{a}\left(\frac{1}{\sqrt{3}}, \frac{1}{3}\right) \tag{11}$$

The Brillouin Zone for a triangular lattice is a hexagon with high symmetry points K and K' on its corners:

$$\mathbf{K} = \frac{2\pi}{a}\left(\frac{2}{3\sqrt{3}}, 0\right) , \quad \mathbf{K}' = \frac{2\pi}{a}\left(-\frac{2}{3\sqrt{3}}, 0\right) \tag{12}$$

Lets now consider a tight binding model with nearest neighbor hopping amplitude t

$$H = U\sum_i a_i^\dagger a_i - U\sum_j b_j^\dagger b_j - t\sum_{\langle i,j\rangle}\left(a_i^\dagger b_j + \text{h.c.}\right) \tag{13}$$

where $a_i^\dagger$ and a_i are the creation and annihilation operators for site i on sublattice A (and similarly, $b_j^\dagger$ and b_j are for B). U is half the energy difference between an electron sitting on an

atom of sublattice B as opposed to A.

Let $|A, \mathbf{r}\rangle$ represent a Wannier function localized on the A site in the unit cell at $\mathbf{r}$. Then,

$$H|A, \mathbf{r}\rangle = U|A, \mathbf{r}\rangle - t\left(|B, \mathbf{r} + \mathbf{a}_1\rangle + |B, \mathbf{r} + \mathbf{a}_2\rangle + |B, \mathbf{r}\rangle\right) \tag{14}$$

And similarly for $|B, \mathbf{r}\rangle$ localized at B in the unit cell at $\mathbf{r}$:

$$H|B, \mathbf{r}\rangle = -U|B, \mathbf{r}\rangle - t\left(|A, \mathbf{r} - \mathbf{a}_1\rangle + |A, \mathbf{r} - \mathbf{a}_2\rangle + |A, \mathbf{r}\rangle\right) \tag{15}$$

Define Bloch states

$$|A, \mathbf{k}\rangle = \sum_{\mathbf{r}} e^{i\mathbf{k}\cdot\mathbf{r}}|A, \mathbf{r}\rangle \tag{16}$$

$$|B, \mathbf{k}\rangle = \sum_{\mathbf{r}} e^{i\mathbf{k}\cdot\mathbf{r}}|B, \mathbf{r}\rangle \tag{17}$$

Applying the Hamiltonian on these Bloch states,

$$H|A, \mathbf{k}\rangle = U|A, \mathbf{k}\rangle - t\left(e^{-i\mathbf{k}\cdot\mathbf{a}_1} + e^{-i\mathbf{k}\cdot\mathbf{a}_2} + 1\right)|B, \mathbf{k}\rangle \tag{18}$$

$$H|B, \mathbf{k}\rangle = -U|B, \mathbf{k}\rangle - t\left(e^{i\mathbf{k}\cdot\mathbf{a}_1} + e^{i\mathbf{k}\cdot\mathbf{a}_2} + 1\right)|A, \mathbf{k}\rangle \tag{19}$$

Clearly, the Hamiltonian for each $\mathbf{k}$ subspace can be represented by the matrix

$$H(\mathbf{k}) = -t\begin{pmatrix} U & e^{-i\mathbf{k}\cdot\mathbf{a}_1} + e^{-i\mathbf{k}\cdot\mathbf{a}_2} + 1 \\ e^{i\mathbf{k}\cdot\mathbf{a}_1} + e^{i\mathbf{k}\cdot\mathbf{a}_2} + 1 & -U \end{pmatrix} \tag{20}$$

We can rewrite the Hamiltonian in terms of the Pauli matrices σ_x, σ_y, and σ_z by defining

$$X(\mathbf{k}) = -t\left[1 + \cos\left(\frac{\sqrt{3}}{2}k_x a - \frac{3}{2}k_y a\right) + \cos\left(-\frac{\sqrt{3}}{2}k_x a - \frac{3}{2}k_y a\right)\right] \tag{21}$$

$$Y(\mathbf{k}) = t\left[\sin\left(\frac{\sqrt{3}}{2}k_x a - \frac{3}{2}k_y a\right) + \sin\left(-\frac{\sqrt{3}}{2}k_x a - \frac{3}{2}k_y a\right)\right] \tag{22}$$

$$Z(\mathbf{k}) = U \tag{23}$$

so that $H(\mathbf{k}) = X(\mathbf{k})\sigma_x + Y(\mathbf{k})\sigma_y + Z(\mathbf{k})\sigma_z$ and $E(\mathbf{k}) = \pm\sqrt{X(\mathbf{k})^2 + Y(\mathbf{k})^2 + Z(\mathbf{k})^2}$ for all $\mathbf{k}$ in the Brillouin Zone. We can obtain the low energy electronic structure by expanding $X(\mathbf{k})$ and $Y(\mathbf{k})$ around the K point (i.e., let $\mathbf{k} = \mathbf{K} + \mathbf{q}$): $X(\mathbf{k}) \approx \frac{3at}{2}q_x$, $Y(\mathbf{k}) \approx \frac{3at}{2}q_y$. Thus,

$$H(\mathbf{K} + \mathbf{q}) = \hbar v_F\left(q_x\sigma_x + q_y\sigma_y\right) + U\sigma_z \tag{24}$$

where $v_F = \frac{3at}{2\hbar}$. This is the two-dimensional Dirac Equation with mass U and the speed of light replaced by v_F. For a similar expansion near K': $H(\mathbf{K}' + \mathbf{q}) = \hbar v_F\left(-q_x\sigma_x + q_y\sigma_y\right) + U\sigma_z$.

Graphene is similar to this tight binding model in that the electrons in graphene hop between the out-of-plane p orbitals [8]. For graphene, there is no energy difference between A and B, so $U = 0$. Thus, $E = \pm\hbar v_F q$, so graphene is not gapped. On the other hand, $U \neq 0$ for hexagonal boron nitride (h-BN), so it has a gap at K and K'.

3 A Trivial Insulator

Let $\mathbf{d}(\mathbf{k}) = (d_x, d_y, d_z) = (X(\mathbf{k}), Y(\mathbf{k}), Z(\mathbf{k}))$. Parameterize $\mathbf{d}(\mathbf{k})$ in spherical coordinates $\mathbf{d}(\mathbf{k}) = |\mathbf{d}(\mathbf{k})|(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\phi)$, where θ and ϕ depend on $\mathbf{k}$, and $|\mathbf{d}(\mathbf{k})| = E(\mathbf{k})$. Then, the Hamiltonian is

$$H(\mathbf{k}) = \mathbf{d}(\mathbf{k}) \cdot \boldsymbol{\sigma} = |\mathbf{d}(\mathbf{k})| \begin{pmatrix} \cos\theta & \sin\theta e^{-i\phi} \\ \sin\theta e^{i\phi} & -\cos\theta \end{pmatrix} \tag{25}$$

The eigenvectors of this matrix are [6]

$$|\mathbf{k}, -\rangle = \begin{pmatrix} \sin\frac{\theta}{2}e^{-i\phi} \\ -\cos\frac{\theta}{2} \end{pmatrix}, \quad |\mathbf{k}, +\rangle = \begin{pmatrix} \cos\frac{\theta}{2}e^{-i\phi} \\ \sin\frac{\theta}{2} \end{pmatrix} \tag{26}$$

The state $|\mathbf{k}, -\rangle$ has lower energy than $|\mathbf{k}, +\rangle$. We can calculate the Berry curvature with respect to parameters θ and ϕ. We only need to calculate this for the lower energy state because the curvatures must sum to zero:

$$A_\theta = \langle \mathbf{k}, -|\frac{\partial}{\partial\theta}|\mathbf{k}, -\rangle = 0 \tag{27}$$

$$A_\phi = \langle \mathbf{k}, -|\frac{\partial}{\partial\phi}|\mathbf{k}, -\rangle = \sin^2\frac{\theta}{2} \tag{28}$$

$$\Omega_{\theta,\phi} = \frac{\partial}{\partial\theta}A_\phi - \frac{\partial}{\partial\phi}A_\theta = \frac{1}{2}\sin\theta \tag{29}$$

We can convert this to a Berry curvature with respect to k_x and k_y by multiplying by the Jacobian of the transformation between (θ, ϕ) and (k_x, k_y):

$$\Omega_{k_x,k_y} = \Omega_{\theta,\phi}\left[\frac{\partial\theta}{\partial k_x}\frac{\partial\phi}{\partial k_y} - \frac{\partial\phi}{\partial k_x}\frac{\partial\theta}{\partial k_y}\right] \tag{30}$$

Thus, the Hall conductance will be

$$\sigma_{xy} = \frac{e^2}{h}\frac{1}{4\pi}\int_{BZ} dk_x dk_y \sin\theta \left[\frac{\partial\theta}{\partial k_x}\frac{\partial\phi}{\partial k_y} - \frac{\partial\phi}{\partial k_x}\frac{\partial\theta}{\partial k_y}\right] \tag{31}$$

After another change of variables [9]

$$\sigma_{xy} = \frac{e^2}{h}\frac{1}{2\pi}\int_{BZ}\frac{1}{2}\frac{\mathbf{d}(\mathbf{k})}{|\mathbf{d}(\mathbf{k})|^3} \cdot \left(\frac{\partial\mathbf{d}(\mathbf{k})}{\partial k_x} \times \frac{\partial\mathbf{d}(\mathbf{k})}{\partial k_y}\right) dk_x dk_y \tag{32}$$

How do we interpret the above expression? If $H = \mathbf{d}(\mathbf{k}) \cdot \boldsymbol{\sigma}$, then $\nabla_{\mathbf{d}(\mathbf{k})} H = \boldsymbol{\sigma}$. Plugging into Eq. (4), for $\mathbf{R} = \mathbf{d}(\mathbf{k})$ and the lower energy state

$$\boldsymbol{\Omega}_-(\mathbf{d}(\mathbf{k})) = \frac{1}{2} \frac{\mathbf{d}(\mathbf{k})}{|\mathbf{d}(\mathbf{k})|^3} \tag{33}$$

The rest of the integrand is the unit normal to S, the surface in $\mathbb{R}^3$ parametrized by $\mathbf{d}(\mathbf{k})$ as $\mathbf{k}$ sweeps over the entire Brillouin zone. In other words, S is the image of the map $\mathbf{k} \to \mathbf{d}(\mathbf{k})$. We will discuss the importance of this later, but as an aside, the Gauss-Bonnet Theorem asserts that the surface integral of a curvature is a topological invariant. For now, lets return to Eq. (31), which can be rewritten as

$$\sigma_{xy} = -\frac{e^2}{h} \frac{1}{4\pi} \int_{BZ} dk_x dk_y \left[\frac{\partial \left(\cos\theta \frac{\partial\phi}{\partial k_y} \right)}{\partial k_x} - \frac{\partial \left(\cos\theta \frac{\partial\phi}{\partial k_x} \right)}{\partial k_y} \right] \tag{34}$$

Naively, this integral is zero because the Brillouin zone has the topology of a closed torus. However, that is not necessarily true because the integrand contains $\nabla_{\mathbf{k}}\phi$, which has singularities at $\theta = 0$ and $\theta = \pi$, the north and south poles of a Bloch sphere. Thus, using Green's Theorem, the Hall conductance can be written as the sum of line integrals around points in the Brillouin zone where $\theta = 0$ and $\theta = \pi$:

$$\sigma_{xy} = -\frac{e^2}{h} \frac{1}{4\pi} \lim_{\epsilon \to 0} \sum_{\theta=0,\pi} \oint_{C(\epsilon)} \cos(\theta) \nabla_{\mathbf{k}}\phi \cdot d\mathbf{k} \tag{35}$$

where $C(\epsilon)$ is an infinitesimally small circle around the singularity. If we were to do a more careful analysis, we would find that a nonzero Hall conductance comes from an obstruction of Stokes' Theorem over the Brillouin zone due to the fact that we cannot define a smooth gauge for all the wavefunctions.

Does our model for a honeycomb lattice have a nonzero Hall conductivity? We will now show that the answer is no. The angles $\theta = 0$ and $\theta = \pi$ occur at the K and K' points. We can write the Hamiltonian as $H = \xi q_x \sigma_x + q_y \sigma_y + M \sigma_z$, where we let $M = Z = U$, $\xi = \pm 1$ depending on whether we are at K or K', and $\hbar v_F = 1$ for convenience. We can also shift q_x and q_y to k_x and k_y, so the Hamiltonian becomes $H = \xi k_x \sigma_x + k_y \sigma_y + M \sigma_z$. Then, $\cos\theta = M/|\mathbf{d}(\mathbf{k})| = M/\sqrt{k_x^2 + k_y^2 + M^2}$. Since $\epsilon \to 0$, $\cos\theta = \text{sign}(M)$. The closed loop integral of $\nabla_{\mathbf{k}}\phi$ can easily be evaluated explicitly, but there is no need because ϕ is an angle whose direction clockwise or counterclockwise is controlled by ξ. Thus, $\oint \nabla_{\mathbf{k}}\phi \cdot d\mathbf{k} = 2\pi\xi$. Putting it

6

all together,

$$\sigma_{xy} = -\frac{e^2}{h} \sum_{\theta=0,\pi} \frac{\xi}{2} \operatorname{sign}(M) = \frac{e^2}{h} \left[\frac{1}{2} \left(\operatorname{sign}(Z(\mathbf{k} = \mathbf{K}')) - \operatorname{sign}(Z(\mathbf{k} = \mathbf{K})) \right) \right] \qquad (36)$$

Since $Z(\mathbf{k} = \mathbf{K}) = Z(\mathbf{k} = \mathbf{K}') = M$ in our model, $\sigma_{xy} = 0$. The system described by the Hamiltonian in Eq. (13) is an example of a trivial insulator. A trivial insulator is a system whose Hamiltonian can be continuously deformed into the vacuum without closing the energy gap, and thus, a trivial insulator does not have quantized conductive edge states.

4 The Haldane Model

Now we can finally discuss the Haldane model. Haldane considers a honeycomb lattice in a spatially modulated magnetic field [10]. The magnetic field is arranged such that there is zero net flux through each unit cell. We model the effect of the magnetic field by a $\int \mathbf{A} \cdot d\mathbf{r}$ phase in the next nearest neighbor hopping. Suppose the amplitude for hopping clockwise and counterclockwise around the hollow site to a next nearest neighbor is $t'e^{i\alpha}$ and $t'e^{-i\alpha}$, respectively. In Fig. 2, an electron can hop to its equivalent atom in each of its six adjacent unit cells. Whether the strength of the hopping is $t'e^{i\alpha}$ or $t'e^{-i\alpha}$ depends on whether it hops with or against the orange loops. The Hamiltonian is

$$H = -t \sum_{\langle i,j \rangle} \left(a_i^\dagger b_j + \text{h.c.} \right) - t' \sum_{\langle\langle i,j \rangle\rangle} \left(e^{i\nu_{ij}\alpha} a_i^\dagger a_j + \text{h.c.} \right) - t' \sum_{\langle\langle i,j \rangle\rangle} \left(e^{i\nu_{ij}\alpha} b_i^\dagger b_j + \text{h.c.} \right) \qquad (37)$$

where $\nu_{ij} = \pm 1$ depending on whether the electron hops clockwise or counterclockwise to its next nearest neighbor. The first term is regular nearest neighbor hopping. The second term is hopping between atoms on the A sublattice (denoted by purple dotted line in Fig. 2), and the third term is hopping on the B sublattice (denoted by the green dotted line). Note that we set $U = 0$.
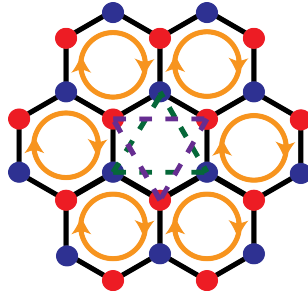

Figure 2: The Haldane Model for Next Nearest Neighbor Hopping

Doing the same analysis as for the model in Eq. (13),

$$H|A, \mathbf{r}\rangle = - t \left(|B, \mathbf{r} + \mathbf{a}_1\rangle + |B, \mathbf{r} + \mathbf{a}_2\rangle + |B, \mathbf{r}\rangle \right)$$

$$- t'(e^{i\alpha}|A, \mathbf{r} + \mathbf{a}_1\rangle + e^{-i\alpha}|A, \mathbf{r} + \mathbf{a}_2\rangle +$$

$$e^{i\alpha}|A, \mathbf{r} + \mathbf{a}_2 - \mathbf{a}_1\rangle + e^{-i\alpha}|A, \mathbf{r} - \mathbf{a}_1\rangle +$$

$$e^{i\alpha}|A, \mathbf{r} - \mathbf{a}_2\rangle + e^{-i\alpha}|A, \mathbf{r} - \mathbf{a}_2 + \mathbf{a}_1\rangle) \tag{38}$$

$$H|B, \mathbf{r}\rangle = - t \left(|A, \mathbf{r} + \mathbf{a}_1\rangle + |A, \mathbf{r} + \mathbf{a}_2\rangle + |A, \mathbf{r}\rangle \right)$$

$$- t'(e^{-i\alpha}|B, \mathbf{r} + \mathbf{a}_1\rangle + e^{i\alpha}|B, \mathbf{r} + \mathbf{a}_2\rangle +$$

$$e^{-i\alpha}|B, \mathbf{r} + \mathbf{a}_2 - \mathbf{a}_1\rangle + e^{i\alpha}|B, \mathbf{r} - \mathbf{a}_1\rangle +$$

$$e^{-i\alpha}|B, \mathbf{r} - \mathbf{a}_2\rangle + e^{i\alpha}|B, \mathbf{r} - \mathbf{a}_2 + \mathbf{a}_1\rangle) \tag{39}$$

Apply H to the Bloch states defined in Eg. (16) and Eq. (17), and define $S(\mathbf{k}) = e^{i\mathbf{k}\cdot\mathbf{a}_1} + e^{i\mathbf{k}\cdot\mathbf{a}_2} + 1$ and $T(\mathbf{k}) = e^{-i\mathbf{k}\cdot\mathbf{a}_1} + e^{-i\mathbf{k}\cdot(\mathbf{a}_2 - \mathbf{a}_1)} + e^{i\mathbf{k}\cdot\mathbf{a}_2}$. Then, put H in a matrix form

$$H(\mathbf{k}) = \begin{pmatrix} -t'[e^{i\alpha}T(\mathbf{k}) + e^{-i\alpha}T^*(\mathbf{k})] & -tS^*(\mathbf{k}) \\ -tS(\mathbf{k}) & -t'[e^{-i\alpha}T(\mathbf{k}) + e^{i\alpha}T^*(\mathbf{k})] \end{pmatrix} \tag{40}$$

Using $T(\mathbf{k} = \mathbf{K}) = -\frac{3}{2} + \frac{3i\sqrt{3}}{2}$ and $S(\mathbf{k} = \mathbf{K} + \mathbf{q}) \approx -\frac{3}{2}a(q_x + iq_y)$

$$H(\mathbf{K} + \mathbf{q}) = \begin{pmatrix} 3\sqrt{3}t' \sin\alpha & 3at(q_x - iq_y)/2 \\ 3at(q_x + iq_y)/2 & -3\sqrt{3}t' \sin\alpha \end{pmatrix} \tag{41}$$

or expressed in terms of Pauli matrices

$$H(\mathbf{K} + \mathbf{q}) = \hbar v_F (q_x \sigma_x + q_y \sigma_y) + (3\sqrt{3}t' \sin\alpha)\sigma_z \tag{42}$$

$$H(\mathbf{K}' + \mathbf{q}) = \hbar v_F (-q_x \sigma_x + q_y \sigma_y) + (-3\sqrt{3}t' \sin\alpha)\sigma_z \tag{43}$$

The coefficient of σ_z is frequently referred to as the mass. That the Haldane masses at K and K' have different sign is related to a lack of time reversal symmetry.

What is the Hall conductance? For the trivial insulator, the mass terms at K and K' have the same sign, so by Eq. (36), the Hall conductance was zero. For the Haldane mode, the masses at K and K' have opposite sign. Thus, if $\sin\alpha \neq 0$, Eq. (36) implies that $\sigma_{xy} = \pm e^2/h$, where the sign of σ_{xy} depends on the value of α.

5   The Edge State

For both the Haldane model (henceforth referred to as a Haldane insulator) and the trivial insulator, we expanded the Hamiltonian matrix around K and K'. Whether we are near K or K' is called the "valley" degree of freedom. If we define τ_z as the third Pauli matrix acting on the valley degree of freedom (and I as identity), we can unite the Hamiltonian expansions around K and K' into a single expression. For the trivial insulator:

$$H = \hbar v_F(q_x \sigma_x \otimes \tau_z + q_y \sigma_y \otimes I) + U\sigma_z \otimes I \tag{44}$$

and for the Haldane insulator:

$$H = \hbar v_F(q_x \sigma_x \otimes \tau_z + q_y \sigma_y \otimes I) + M_H \sigma_z \otimes \tau_z \tag{45}$$

where $M_H = 3\sqrt{3}t' \sin\alpha$ is the Haldane mass.

Now, suppose we put a Haldane insulator next to a trivial insulator. In the x-y plane, let $x < 0$ be the trivial insulator and $x > 0$ be the Haldane insulator. We want to solve $H\Psi = E\Psi$. Since we have translational symmetry in the y-direction, we can let $\Psi(x,y) = \psi(x)e^{ik_y y}$. Also, for simplicity, let $\hbar v_F = 1$ and $M = M_H = U$. Then [9]

$$\left[-i\sigma_x \otimes \tau_z \partial/\partial x + q_y \sigma_y + M\left(\Theta(-x)\sigma_z + \Theta(x)\sigma_z \otimes \tau_z\right)\right]\psi(x) = E\psi(x) \tag{46}$$

where $\Theta(x)$ is the Heaviside step function. The identity matrix I has been suppressed. For the moment, lets look for mid-gap solutions by setting $q_y = 0$ and $E = 0$.

For $x > 0$, $-i\sigma_x \otimes \tau_z \partial\psi/\partial x = -M\sigma_z \otimes \tau_z \psi$. Multiply both sides by $i\sigma_z \otimes \tau_z$ to get $\partial\psi/\partial x = -M\sigma_y \psi$. There exists two solutions to this differential equation that are localized on the boundary between the Haldane insulator and trivial insulator:

$$\psi_{\mathbf{K}}(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-M|x|}, \ \psi_{\mathbf{K'}}(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{-M|x|} \tag{47}$$

The corresponding differential equation for $x < 0$ is $\partial\psi/\partial x = -M\sigma_y \otimes \tau_z \psi$. Plugging both solutions into this equation reveals that only one of these solutions makes sense:

$$\psi_{\mathbf{K'}}(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{-M|x|} \tag{48}$$

This solution is also an eigenvector of $q_y \sigma_y$, so we deduce that the energy-momentum dispersion

relation for the edge state is $E = \hbar v_F q_y$, or allowing different signs of the Haldane mass at K,

$$E_{\text{Edge State}} = \text{sign}(M_H)\hbar v_F q_y \tag{49}$$

6   Winding Numbers

What makes the Haldane insulator different from the trivial insulator? Lets return to Eq. (32), which can be written as

$$\sigma_{xy} = n\frac{e^2}{h} \tag{50}$$

$$n = \frac{1}{4\pi}\int_{BZ}\hat{\mathbf{d}}(\mathbf{k}) \cdot \left(\frac{\partial\hat{\mathbf{d}}(\mathbf{k})}{\partial k_x} \times \frac{\partial\hat{\mathbf{d}}(\mathbf{k})}{\partial k_y}\right)dk_x dk_y \tag{51}$$

where $\hat{\mathbf{d}}(\mathbf{k}) = \mathbf{d}(\mathbf{k})/|\mathbf{d}(\mathbf{k})|$ is the unit vector in the $\mathbf{d}(\mathbf{k})$ direction, and n is an integer that goes by many names: the degree of continuous mapping, the first Chern number, the TKNN invariant, and the winding number [11]. The image of the map $\mathbf{k} \to \mathbf{d}(\mathbf{k})$ is a closed surface (denoted S) embedded in $\mathbb{R}^3$. The winding number n is the number of times S encloses the origin:

$$n = \frac{1}{4\pi}\int_{BZ}\frac{\mathbf{d}(\mathbf{k})}{|\mathbf{d}(\mathbf{k})|^3} \cdot \left(\frac{\partial\mathbf{d}(\mathbf{k})}{\partial k_x} \times \frac{\partial\mathbf{d}(\mathbf{k})}{\partial k_y}\right)dk_x dk_y = \int_S\frac{\hat{\mathbf{d}}(\mathbf{k})}{4\pi|\mathbf{d}(\mathbf{k})|^2} \cdot d\mathbf{S} \tag{52}$$

This familiar integral is just the flux of a monopole centered at the origin. If S does not enclose the origin, $n = 0$. If S encloses the origin once, $n = 1$ (or $n = -1$ depending on the orientation of S, but we will not worry about orientation), and so on.

The trivial insulator has Hamiltonian $H = \xi q_x\sigma_x + q_y\sigma_y + M\sigma_z$, where $\xi = 1$ for K and $\xi = -1$ for K'. If M is positive, then the vector $\mathbf{d}(\mathbf{k})$ lies on the positive z-axis of $\mathbb{R}^3$ (the north pole) at K and K'. It is easy to see that $\mathbf{d}(\mathbf{k})$ does not lie on the z-axis for any other $\mathbf{k}$. What surface S crosses the north pole twice but never crosses the south pole (the negative z-axis)? The surface S clearly does not enclose the origin, and thus $n = 0$ and $\sigma_{xy} = 0$.

The Haldane insulator has Hamiltonian $H = \xi q_x\sigma_x + q_y\sigma_y + \xi M\sigma_z$. Now, $\mathbf{d}(\mathbf{K}) = (0, 0, M)$, $\mathbf{d}(\mathbf{K}') = (0, 0, -M)$, and $(d_x, d_y) \neq (0, 0)$ everywhere else in the Brillouin zone. Thus, the surface S crosses the north pole once and the south pole once. A closed surface that crosses the north and south pole each exactly once must contain the origin exactly one time! We conclude that $|n| = 1$ and $|\sigma_{xy}| = 1$.

Now, suppose we adiabatically evolved the Haldane insulator into the trivial insulator. In order to continuously deform a surface that encloses the origin into a surface that does not enclose the origin, the surface must cross the origin. This implies that $\mathbf{d}(\mathbf{k}) = 0$ at some point during the adiabatic evolution, so the energy gap closes! We would say that the Haldane insulator and trivial insulator are different topological phases. This is why we have a conducting edge state when we put a Haldane insulator and trivial insulator next to each other. When you cross the boundary between the two insulators, the gap must close.

7 The Kane and Mele Model

Now we finally introduce a model for the quantum spin Hall (QSH) effect. Kane and Mele proposed the following Hamiltonian for a honeycomb lattice with spin-orbit coupling [12]

$$H = \left[-t \sum_{\langle i,j \rangle \alpha} a^\dagger_{i\alpha} b_{\alpha j} - it' \sum_{\langle\langle i,j \rangle\rangle} \nu_{ij} \left(s_z\right)_{\alpha\beta} a^\dagger_{i\alpha} a_{j\beta} - it' \sum_{\langle\langle i,j \rangle\rangle \alpha\beta} \nu_{ij} \left(s_z\right)_{\alpha\beta} b^\dagger_{i\alpha} b_{j\beta} \right] + \text{h.c.} \qquad (53)$$

where s_z is the third Pauli matrix acting on the spin degree of freedom, α and β are spin indices (up and down), and $\nu_{ij} = \pm 1$ depending on whether the electron hops clockwise or counterclockwise to its next nearest neighbor.

The Kane and Mele Hamiltonian is just two copies of the Haldane model, with $\alpha = \pi/2$ for spin up and $\alpha = -\pi/2$ for spin down. Therefore, we can rewrite Eq. (53) as

$$H = \hbar v_F (q_x \sigma_x \otimes \tau_z + q_y \sigma_y) + \Delta_{SO} \sigma_z \otimes \tau_z \otimes s_z \qquad (54)$$

where $\Delta_{SO} = 3\sqrt{3} t'$. Since the mass term is positive for spin up and negative for spin down, Eq. (49) implies that there are two counter-propagating edge modes. The spin of an edge mode is locked to its momentum (spin up for $+q_y$, spin down for $-q_y$). This had to be true because the Kane and Mele model (unlike the Haldane model) obeys time reversal symmetry and therefore has a Kramers degeneracy.

The winding numbers for spin up and spin down have opposite sign, so total $\sigma_{xy} = 0$. However, the Kane and Mele model is still different from the trivial insulator because we can define a new topological invariant: the spin Chern number $C_{spin} = (n_\uparrow - n_\downarrow)/2$, where $n_\uparrow$ and $n_\downarrow$ are the winding numbers for spin up and down, respectively. The spin Chern number describes the spin current, which is nonzero because spin and momentum are locked. The Kane and Mele

11

model of a QSH insulator has $|C_{spin}| = 1$.

An astute reader will notice that the Kane and Mele model conserves s_z, which may not happen with a different spin-orbit coupling term. The actual topological invariant that characterizes time reversal symmetric insulators is the so-called $\mathbb{Z}_2$ invariant, which takes values 0 (for the trivial phase) and 1 (for the nontrivial phase) [13]. If s_z is conserved, the $\mathbb{Z}_2$ invariant is equal to C_{spin} mod 2. We will not define the $\mathbb{Z}_2$ invariant more generally because it requires additional mathematical machinery, such as the Pfaffian of a matrix. Although it is too long to include in this paper, the theory of the $\mathbb{Z}_2$ invariant is quite interesting and describes the inability to define a smooth gauge convention that obeys time reversal symmetry. We will also note that, for systems with inversion symmetry, the $\mathbb{Z}_2$ invariant can be calculated through some product of parity eigenvalues at certain high symmetry points in the Brillouin zone (specifically $\mathbf{k}$ that are mapped onto themselves under time reversal) [14].

8   Protection from Backscattering

Even without discussing the $\mathbb{Z}_2$ invariant, we can give a heuristic argument for the protection of the QSH edge states [15]. As long as the bulk remains gapped, time reversal symmetric perturbations cannot gap the edge states. If we define $k = q_y$ in Eq. (49), then $E_{edge,\alpha} = (s_z)_{\alpha\alpha}\hbar v_F k$. We can write a Hamiltonian just for the edge modes

$$H = \hbar \begin{pmatrix} v_F k & 0 \\ 0 & -v_F k \end{pmatrix} \tag{55}$$

A pair of counter-propagating edge modes is called a Dirac cone. The two edge states meet at the "Dirac point" at $k = 0$. If we apply a time reversal symmetric perturbation, will the two edge states still cross at the Dirac point? The most general time reversal symmetric perturbation that we can write is

$$H' = \sum_{ni} \left[a_{ni} k^{2n+1} s_i + b_n k^{2n} I \right] \tag{56}$$

where s_i are the three Pauli matrices and I is the identity matrix. At $k = 0$, $H' = b_0 I$, so the two edge modes are still degenerate at $k = 0$. Therefore, no gap opens, and the edge states remain metallic. This would not be true if H' disobeyed time reversal symmetry.

Suppose we had two copies of the Kane and Mele model. If we had two Dirac cones, would our metallic edge states be protected by time reversal symmetry? We can represent two Dirac

cones by the Hamiltonian

$$H = \hbar \begin{pmatrix} v_1 k & 0 \\ 0 & -v_1 k \end{pmatrix} \oplus \begin{pmatrix} v_2 k & 0 \\ 0 & -v_2 k \end{pmatrix} = \hbar \begin{pmatrix} v_1 k & 0 & 0 & 0 \\ 0 & -v_1 k & 0 & 0 \\ 0 & 0 & v_2 k & 0 \\ 0 & 0 & 0 & -v_2 k \end{pmatrix} \quad (57)$$

where v_1 and v_2 are the velocities of the edge states. For simplicity, we will consider the case $v_F = v_1 = v_2$. We can express the time reversal operator as UK, where K takes numbers to their complex conjugates, and U is given by

$$U = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \quad (58)$$

Now, lets consider a particular perturbation of the system

$$H' = \begin{pmatrix} 0 & 0 & 0 & \lambda \\ 0 & 0 & -\lambda & 0 \\ 0 & -\lambda & 0 & 0 \\ \lambda & 0 & 0 & 0 \end{pmatrix} \quad (59)$$

It is easy to check that $(UH'U^{-1})^* = H'$, so H' is time reversal symmetric. However, if we diagonalize the full Hamiltonian $H + H'$, we get the energy eigenvalues $E = \pm\sqrt{(\hbar v_F k)^2 + \lambda^2}$. Thus, the edge states become gapped by the perturbation.

More generally, time reversal symmetric perturbations destroy pairs of Dirac cones. Therefore, if there are an odd number of Dirac cones, the metallic edge states are protected against time reversal invariant perturbations. If there are an even number of Dirac cones, these edge modes are not protected. Since the Kane and Mele QSH insulator has one Dirac cone, its edge channels are robust, so long as the bulk remains gapped.

References

[1] D. J. Griffiths, *Introduction to Quantum Mechanics* (2004)

[2] B. A. Bernevig and T. L. Hughes, *Topological Insulators and Topological Superconductors* (2013)

[3] Cohen and Louie, Chapter 5

[4] R. Karplus and J. M. Luttinger, *Physical Review* **95**, 1154 (1954).

[5] D. J. Thouless, M. Kohmoto, M. P. Nightingale, M. den Nijs, *Physical Review Letter* **49**, 405 (1982)

[6] D. Xiao, M. C. Chang, and Q. Niu, *Reviews of Modern Physics* **82**, 1959 (2010)

[7] M. I. Katsnelson, *Graphene: Carbon in Two Dimensions* (2012)

[8] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, A. K. Geim, *Reviews of Modern Physics* **81**, 109 (2009)

[9] J. Cayssol, *Comptes Rendus Physique* **14** 760-778 (2013)

[10] F. D. M. Haldane, *Physical Review Letters* **61**, 2015 (1988)

[11] M. Z. Hasan and C. L. Kane, *Reviews Modern Physics* **82**, 3045 (2010)

[12] C. L. Kane and E. J. Mele, *Physical Review Letters* **95**, 226801 (2005)

[13] C. L. Kane and E. J. Mele, *Physical Review Letters* **95**, 146802 (2005)

[14] Liang Fu and C. L. Kane *Physical Review B* **76**, 045302 (2007)

[15] Sun Kai's lecture notes, University of Michigan

[16] B. A. Bernevig, T. L. Hughes, S. C. Zhang, *Science* **314** 1757 (2006)

[17] M. König *et al.*, *Science* **318**, 766 (2007)

[18] J. Maciejko, T. L. Hughes, S. C. Zhang, *Annual Review of Condensed Matter Physics* **2** 31 (2011)

[19] C. Brüne *et al.*, *Nature Physics* **8**, 485 (2012)

[20] Liang Fu, C. L. Kane, and E. J. Mele *Physical Review Letters* **98**, 106803 (2007)

[21] D. Hsieh *et al.*, *Nature* **452**, 970 (2008)

[22] D. Hsieh *et al.*, *Science* **323**, 919 (2009)

[23] P. Roushan *et al.*, *Nature* **460**, 1106 (2009)

[24] Y. L. Chen *et al.*, *Science* **325**, 178 (2009)

[25] Y. L. Chen *et al.*, *Science* **329**, 659 (2010)

[26] C. Z. Chang *et al.*, *Science* **340**, 167 (2013)

[27] Z. K. Liu *et al.*, *Science* **343**, 864 (2014)