TOPICAL REVIEW— Topological insulator

Spin Chern numbers and time-reversal-symmetry-broken quantum spin Hall effect

Sheng Li(盛 利)[a)†], Li Hui-Chao(李会超)[a)], Yang Yun-You(杨运友)[a)],
Sheng Dong-Ning(盛冬宁)[b)†], and Xing Ding-Yu(邢定钰)[a)‡]

[a)] *National Laboratory of Solid State Microstructures and Department of Physics, Nanjing University, Nanjing* 210093, *China*
[b)] *Department of Physics and Astronomy, California State University, Northridge, California* 91330, *USA*

(Received 17 April 2013)

The quantum spin Hall (QSH) effect is considered to be unstable to perturbations violating the time-reversal (TR) symmetry. We review some recent developments in the search of the QSH effect in the absence of the TR symmetry. The possibility to realize a robust QSH effect by artificial removal of the TR symmetry of the edge states is explored. As a useful tool to characterize topological phases without the TR symmetry, the spin-Chern number theory is introduced.

Keywords: spin-polarized transport, quantum spin Hall effect, surface state, edge state, topological insulator

PACS: 72.25.–b, 73.43.–f, 73.20.At, 73.50.–h **DOI:** 10.1088/1674-1056/22/6/067201

1. Introduction

The quantum spin Hall (QSH) effect, a quantum state of matter, has attracted much attention in recent years because of its fundamental interest and potential applications in spintronic devices. The QSH effect was first predicted theoretically by Kane and Mele[1] and by Bernevig and Zhang[2] independently. Soon after, the QSH effect was observed experimentally in HgTe quantum wells,[3] following the theoretical prediction.[4] A key ingredient to the QSH effect is a strong intrinsic spin–orbit coupling, which acts as spin-dependent magnetic fluxes coupled to the electron momentum. In the ideal case, where the electron spin is conserved, the two spin sectors of a QSH system behave like two independent quantum Hall (QH) systems without Landau levels.[5] They contribute opposite quantized Hall conductivities when the electron Fermi level is inside the bulk band gap, so that the total Hall conductivity vanishes but the spin Hall conductivity is quantized. On a sample edge, two counterpropagating gapless edge modes with opposite spin polarizations exist in the bulk band gap, which can transport spin currents without dissipation of energy.

When the spin conservation is destroyed, e.g., by the Rashba-like spin–orbit coupling, the spin Hall conductivity deviates from the quantized value.[6] However, the edge transport can remain dissipationless,[1,7] provided that the time-reversal (TR) symmetry is present and the bulk band gap is not closed. In this case, the QSH system can no longer be divided into two QH systems, and the existence of the gapless edge states is attributed to the nontrivial topological properties of bulk energy bands. The nontrivial bulk band topology of the QSH systems is usually described by the Z_2 index[8] or the spin Chern numbers.[9–11] These topological invariants reveal the fundamental distinction between a QSH insulator and an ordinary band insulator. Prodan proved[10] that the spin-Chern numbers are topological invariants, as long as the energy gap and the spectrum gap of the projected spin operator $P\hat{\sigma}_z P$ stay open in the bulk, where P is the projection operator onto the subspace of the occupied bands, and $\hat{\sigma}_z$ is the Pauli matrix for the electron spin. Unlike the Z_2 invariant, the robust properties of the spin-Chern numbers remain unchanged when the TR symmetry is broken.

Extension of the idea of the QSH effect to higher dimensions has led to the proposals of three-dimensional (3D) topological insulators (TIs).[12–17] The research of the 3D TIs has been fruitful in recent years, both theoretically and experimentally. A 3D TI has a bulk band gap and gapless surface states on the sample boundary. A simple 3D TI consists of layers of 2D QSH systems. However, such a 3D TI is not stable in the presence of disorders due to the interlayer scattering, and is called the weak TI. A strong TI is topologically nontrivial in any direction, and the surface states are characterized by an odd number of Dirac cones. The metallic surface states of the 3D TIs provide a unique platform for realizing some exotic physical phenomena, such as Majorana fermions[18] and topological magnetoelectric effect.[19,20] The existence of topological surface states in the 3D TIs has been experimentally

*Project supported by the National Basic Research Program of China (Grant Nos. 2009CB929504, 2011CB922103, and 2010CB923400), the National Natural Science Foundation of China (Grant Nos. 11225420, 11074110, 11174125, 11074109, and 91021003), the Priority Academic Program Development of Jiangsu Higher Education Institutions, China, the US NSF (Grant Nos. DMR-0906816 and DMR-1205734), and Princeton MRSEC (Grant No. DMR-0819860).

†Corresponding author. E-mail: shengli@nju.edu.cn
‡Corresponding author. E-mail: dyxing@nju.edu.cn

© 2013 Chinese Physical Society and IOP Publishing Ltd

confirmed in $Bi_{1-x}Sb_x$, Bi_2Te_3, and Bi_2Se_3 materials,[15,21–26] which evokes a great surge of research interest in this field.

While the TR symmetry was often considered to be a prerequisite to the QSH effect, its role is two-sided. In a TR-invariant QSH system, the two oppositely moving edge states at the Fermi energy are connected to each other under TR, and so have opposite spin orientations. The elastic backscattering from a nonmagnetic random potential is forbidden. On the other hand, the two opposite movers have identical spatial probability distributions. Turning on small TR-symmetry-breaking perturbations immediately couples the two edge states, giving rise to backscattering. This makes the QSH effect fragile in realistic environments,[27] where perturbations violating the TR symmetry are usually unavoidable. Experimentally, two-terminal conductance close to the predicted value $2(e^2/h)$ was observed only for small QSH samples with dimensions of about $(1\ \mu\text{m} \times 1\ \mu\text{m})$[3] in contrast to the traditional QH effect, where the Hall conductivity can be precisely quantized on macroscopic scales. The search for QSH effect as robust as the QH effect has so far been illusive.

An open question is whether we can have a QSH-like phase in a system where the TR symmetry is broken. In this article, we will review some recent developments in the search of QSH effect that is not protected by the TR symmetry or other symmetries. Such a QSH effect is expected to be more robust in realistic environments. As a useful tool to classify the topological phases without the TR symmetry, we will introduce the spin-Chern number theory, together with its application to the study of topological properties of the TIs.

2. Spin-Chern number theory

2.1. A brief history

In 2005 and 2006, Kane and Mele,[1] and Bernevig and Zhang[2] proposed two different models for the QSH effect. We carried out a numerical work[6] with the Kane–Mele model to study the stability of the QSH effect to disorder and spin-mixing perturbation. We found that the charge transport through the counterpropagating spin-polarized edge channels is well quantized. We therefore developed the idea that the robust QSH effect is associated with a new type of topological invariant of the system.[6] The concrete formulation of the topological invariant, namely, the spin-Chern numbers was given in a subsequent work,[9] while Kane and Mele suggested the Z_2 invariant.[8] Consider a finite-size lattice system consisting of $L_x \times L_y$ unit cells, and impose spin-dependent boundary conditions to the many-body wavefunction $\boldsymbol{\Phi}(\cdots \boldsymbol{r}_{i_\alpha} + \boldsymbol{L}_j \cdots) = \mathrm{e}^{\mathrm{i}\theta_j^\alpha}\boldsymbol{\Phi}(\cdots \boldsymbol{r}_{i_\alpha} \cdots)$, where $j = x, y$, $\boldsymbol{L}_x = L_x\boldsymbol{a}_1$, $\boldsymbol{L}_y = L_y\boldsymbol{a}_2$ with $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ being the two primitive vectors of the Bravais lattice, $\boldsymbol{r}_{i_\alpha}$ is a composite index representing the coordinate and spin of the i-th electron ($\alpha = +$ and $-$ corresponding to spin up and spin down), and $0 \leq \theta_j^\alpha < 2\pi$ is the twisted boundary phase. Through a unitary transformation

$$\Psi = \exp\left[-\mathrm{i}\sum_\alpha\sum_{i_\alpha}\left(\frac{\theta_x^\alpha}{L_x}x_{i_\alpha} + \frac{\theta_y^\alpha}{L_y}y_{i_\alpha}\right)\right]\Phi,$$

where the summation runs over all electrons and both spins, Ψ becomes periodic on a torus. One can then define the topological Chern number matrix as

$$C_{\alpha,\beta} = \frac{\mathrm{i}}{4\pi}\int \mathrm{d}\theta_x^\alpha \mathrm{d}\theta_y^\beta \left[\left\langle \frac{\partial\Psi}{\partial\theta_x^\alpha}\middle|\frac{\partial\Psi}{\partial\theta_y^\beta}\right\rangle - \left\langle \frac{\partial\Psi}{\partial\theta_y^\beta}\middle|\frac{\partial\Psi}{\partial\theta_x^\alpha}\right\rangle\right]. \quad (1)$$

By this definition, $C_{\alpha,\beta}$ with $\alpha, \beta = +, -$ are four integer-quantized topological invariants. We demonstrated that the nontrivial QSH phase is identified by the nonzero diagonal matrix elements of the Chern number matrix, namely, the spin-Chern numbers, which are conserved in the presence of finite disorder scatterings and spin nonconserving perturbations. Through the Laughlin's gedanken experiment,[28] we also numerically showed that for the TR-invariant Kane–Mele model, as long as the quantization of the spin-Chern numbers is intact, the gapless nature of the edge states remains robust, indicating the inherent connection between the spin-Chern numbers and the edge states. When the strength of disorder exceeds certain critical value, the spin-Chern numbers become fluctuating, and energy gaps appear in the edge state spectrum.

Once there were some debates on the spin-Chern numbers defined in Eq. (1). Let us assume $C_{++} = -C_{--} = 1$ for a given spin quantization axis. Fu and Kane[29] argued that one can rotate the spin quantization axis continuously and flip it, which does not close the band gap, but changes the signs of the spin-Chern numbers, yielding $C_{++} = -C_{--} = -1$. Later, based on the noncommutative theory for Chern numbers, Prodan[10] showed that the spin-Chern numbers are true bulk topological invariants, whose robustness against disorders and smooth deformations of the Hamiltonian has important physical consequences. In contrast to the Z_2 invariant, which is applicable only to systems with the TR symmetry, the properties of the spin-Chern numbers remain valid even when the TR invariance is broken. Prodan suggested to partition the subspace of the occupied valence band into two nontrivial spin sectors by using the projected spin operator $P\hat{\sigma}_z P$, where P is the projection operator onto the occupied subspace. Then the Chern numbers of the individual spin sectors are just the spin-Chern numbers. Prodan showed that the spin-Chern numbers are protected by two spectrum gaps, namely, the energy gap between the conduction and valence bands and the spin spectrum gap. The spin spectrum gap refers to the gap in the eigenspectrum of the projected spin operator $P\hat{\sigma}_z P$, which can be understood as follows. When spin $\hat{\sigma}_z$ is conserved, the eigenvalues of $P\hat{\sigma}_z P$ can take only two values 1 and -1. When the

spin conservation is destroyed by some small perturbations, the eigenvalues of $P\hat{\sigma}_z P$ will deviate from 1 and -1, forming two islands near 1 and -1, but a finite spectrum gap usually still exists between the two islands, which is the so-called spin spectrum gap.

The issue of the spin-Chern numbers pointed out by Fu and Kane[29] can then be easily resolved. The rotation of the spin quantization axis not only changes the Hamiltonian, but also modifies the operator $P\hat{\sigma}_z P$. When the spin quantization axis is flipped, the two spin sectors must exchange, which implies that the spin spectrum gap must close during the rotation, and the spin-Chern numbers are allowable to change signs, leading to an equivalent description in different reference systems. Here, we should point out that Prodan's argument is suited only to noninteractive electron systems. How to extend the spin-Chern numbers to interactive electron systems is an open question. The original definition Eq. (1) is formulated for a many-body wavefunction, and can be directly applied to interactive electron systems. However, a hiding quantity which plays a similar role to the spin spectrum gap in the many-body scenario is still worth to be explored in the future.

2.2. Analytical method

Inspired by the work of Prodan,[10] we made a progress in analytical calculation of the spin-Chern numbers in 2009.[11] We took thin films of 3D TI Bi_2Se_3 as an example, where the nontrivial topological structures appear in the isospin sectors. Here, isospin is used to represent the bonding and antibonding of the electron states on the two surfaces. We analytically calculated the isospin-Chern numbers in the general case, where the isospin operator $\hat{\tau}_z$ is not conserved. An explicit connection between the isospin-Chern number and the Z_2 index was established, the two invariants were found to yield an equivalent description for TR-invariant QSH systems. In the following, we introduce the analytical approach[11] to the spin-Chern numbers.

We consider the effective Hamiltonian for a thin film of Bi_2Se_3

$$\begin{aligned} H(\boldsymbol{k}) &= E_0 - Dk^2 + \left(M_0 - Bk^2\right)\hat{\tau}_z\hat{\sigma}_z \\ &\quad + v_{\mathrm{F}}(k_y\hat{\sigma}_x - k_x\hat{\sigma}_y) + \gamma U\hat{\tau}_x, \end{aligned} \tag{2}$$

where $\hat{\sigma}_\alpha$ with $\alpha = x, y, z$ are the Pauli matrices for electron spin, and $\hat{\tau}_\alpha$ for bonding and antibonding of the surface states on the two surfaces of the thin film. For simplicity, we shall call $\hat{\tau}_\alpha$ the isospin. Here v_{F}, E_0, D, B, and M_0 are the model parameters, and $\gamma U\hat{\tau}_x$ stands for a gate voltage applied between the two surfaces of the film.

The following calculation is revised from the original derivation,[11] and the same result is obtained. By using a unitary transformation $\mathscr{H} = S^\dagger HS$, where $S = \frac{1}{2}(1+\hat{\sigma}_z) + \frac{1}{2}(1-\hat{\sigma}_z)\hat{\tau}_x$, the Hamiltonian is transformed into

$$\begin{aligned} \mathscr{H}(\boldsymbol{k}) &= E_0 - Dk^2 + (M_0 - Bk^2)\hat{\tau}_z \\ &\quad + [v_{\mathrm{F}}(k_y\hat{\sigma}_x - k_x\hat{\sigma}_y) + \gamma U]\hat{\tau}_x\,. \end{aligned} \tag{3}$$

The transformed Hamiltonian Eq. (3) can now be easily solved by diagonalizing the operator in the square bracket first, and the eigenenergies are obtained as

$$E_\pm^{\mathrm{c(v)}}(\boldsymbol{k}) = E_0 - Dk^2 + (-)\sqrt{(M_0 - Bk^2)^2 + (\gamma U \pm v_{\mathrm{F}}k)^2}, \tag{4}$$

where c (v) stands for the conduction (valence) bands. The energy bands are two-fold degenerated at $\gamma U = 0$, and the degeneracy is lifted when γU is nonzero. The γU term plays a role similar to the Rashba spin–orbit coupling in the 2D electron systems. When $B^2 > D^2$, the valence bands are inverted. Under this condition, if M_0 and B have opposite signs, i.e., $M_0/B < 0$, an energy gap always exists between the conduction and valence bands. If M_0 and B have the same sign, i.e., $M_0/B > 0$, an energy gap opens at small $|U|$. When $|U|$ reaches a critical value

$$U_0 = \frac{v_{\mathrm{F}}}{\gamma}\sqrt{\left|\frac{M_0}{B}\right|},$$

the energy gap closes. With further increasing $|U|$, the gap reopens. The wavefunctions of the two valence bands can be written as a direct product

$$\varphi_\pm(\boldsymbol{k}) = \phi_\pm(\boldsymbol{k}) \otimes \chi_\pm(\boldsymbol{k})\,. \tag{5}$$

Here, $\phi_\pm(\boldsymbol{k}) = (1/\sqrt{2})(\mathrm{i}\,\mathrm{e}^{-\mathrm{i}\theta_k}, \pm 1)^{\mathrm{T}}$ and

$$\chi_\pm(\boldsymbol{k}) = [\sin\alpha_k^\pm, -\mathrm{sgn}(\gamma U \pm v_{\mathrm{F}}k)\cos\alpha_k^\pm]^{\mathrm{T}}$$

are wavefunctions for spin and isospin, respectively, where $\mathrm{e}^{-\mathrm{i}\theta_k} = (k_x - \mathrm{i}k_y)/k$, and $2\alpha_k^\pm = \mathrm{ctg}^{-1}[(M_0 - Bk^2)/|\gamma U \pm v_{\mathrm{F}}k|]$.

The definition of the isospin-Chern numbers relies on a smooth decomposition of the occupied valence band into two sectors through the diagonalization of the operator $\hat{\tau}_z$ (in the representation of H) in the subspace of the valence band. Since $\hat{\tau}_z$ commutes with momentum, the decomposition can be done for each $\boldsymbol{k}$ separately. By expressing $S^\dagger\hat{\tau}_z S = \hat{\tau}_z\hat{\sigma}_z$ as a 2×2 matrix on the basis of $\varphi_+(\boldsymbol{k})$ and $\varphi_-(\boldsymbol{k})$ in the representation of $\mathscr{H}$

$$\begin{pmatrix} \langle\varphi_+|\hat{\tau}_z\hat{\sigma}_z|\varphi_+\rangle & \langle\varphi_+|\hat{\tau}_z\hat{\sigma}_z|\varphi_-\rangle \\ \langle\varphi_-|\hat{\tau}_z\hat{\sigma}_z|\varphi_+\rangle & \langle\varphi_-|\hat{\tau}_z\hat{\sigma}_z|\varphi_-\rangle \end{pmatrix} = \begin{pmatrix} 0 & t(k) \\ t(k) & 0 \end{pmatrix}, \tag{6}$$

where

$$t(k) = \begin{cases} -\cos(\alpha_k^+ + \alpha_k^-), & (k \le |\gamma U/v_{\mathrm{F}}|), \\ \cos(\alpha_k^+ - \alpha_k^-), & (k > |\gamma U/v_{\mathrm{F}}|), \end{cases} \tag{7}$$

and diagonalizing the matrix Eq. (6), we obtain the eigenvalues of $\hat{\tau}_z$ (strictly speaking, $P\hat{\tau}_z P$) as $\lambda(k) = \lambda_\pm(k) \equiv \pm|t(k)|$. The corresponding eigenfunctions of $\hat{\tau}_z$ are

$$\psi_\pm(\boldsymbol{k}) = \frac{1}{\sqrt{2}}[\varphi_+(\boldsymbol{k}) \pm \mathrm{sgn}(t(k))\varphi_-(\boldsymbol{k})]\,. \tag{8}$$

Clearly, the spectrum of $\hat{\tau}_z$ consists of two branches, namely, $\lambda_+(k)$ and $\lambda_-(k)$. We find that for $M_0/B<0$, there is always a finite gap between the two branches around zero. For $M_0/B>0$, there is a gap for $|U|<U_0$. However, if $|U|\geq U_0$, $\lambda_+(k)$ and $\lambda_-(k)$ touch each other, and the gap vanishes. As long as the gap between the two branches of the $\hat{\tau}_z$ spectrum exists, the electron states in the valence band are naturally partitioned into two sectors, namely, $\psi_+(k)$ and $\psi_-(k)$. One can define a Chern number for each isospin sector

$$C_\pm = \frac{1}{2\pi}\int \mathrm{d}^2k\, F_\pm(k)\,, \tag{9}$$

where $F_\pm(k)=\mathrm{i}\hat{e}_z\cdot[\nabla_k\times\langle\psi_\pm(k)|\nabla_k|\psi_\pm(k)\rangle]$.

By using the expression for the gradient operator in the polar coordinate system

$$\nabla_k = \hat{e}_k\frac{\partial}{\partial k}+\hat{e}_{\theta_k}\frac{1}{k}\frac{\partial}{\partial\theta_k},$$

and noticing $\partial\hat{e}_{\theta_k}/\partial k=0$ and $\partial\hat{e}_{\theta_k}/\partial\theta_k=-\hat{e}_k$, it is straightforward to obtain

$$F_\pm(k)=\frac{1}{2k}\frac{\partial}{\partial k}P_\pm(k)$$

with

$$P_\pm(k)=2\mathrm{i}\langle\psi_\pm|\frac{\partial}{\partial\theta_k}|\psi_\pm\rangle,$$

which is dependent on k not θ_k. Substituting these relations into Eq. (9), we obtain for the isospin-Chern numbers

$$C_\pm=\frac{1}{2}\left[P_\pm(\infty)-P_\pm(0)\right]. \tag{10}$$

The evaluation of the Chern numbers now amounts to the calculation of the two limits of $P_\pm(k)$ at $k=0$ and $k\to\infty$. It is easy to obtain $P_\pm(0)=\mp\mathrm{sgn}(M_0)-1$ and $P_\pm(\infty)=\pm\mathrm{sgn}(B)-1$, so the isospin-Chern numbers are

$$C_\pm=\pm\frac{1}{2}\left[\mathrm{sgn}(M_0)+\mathrm{sgn}(B)\right]. \tag{11}$$

This result indicates that when M_0 and B have opposite signs, $C_\pm=0$ and the system is an ordinary insulator. When M_0 and B have the same sign and $|U|<U_0$, $C_\pm=\pm1$ and the system is in a QSH state. Equation (11) does not apply to $|U|>U_0$, where the isospin spectrum gap always closes, as mentioned above. By continuity analysis, we find that the system is an ordinary insulator for $|U|>U_0$.

2.3. Comparison with Z_2 index

Based upon the above result, we can establish a relation between the spin (isospin)-Chern numbers and the widely used Z_2 index. By defining a function $P(k)=[P_+(k)-P_-(k)]/2$, one can find that the Z_2 index essentially counts the number of zeros of $P(k)$ for $k\in[0,\infty)$.[11,41] If $P(0)P(\infty)>0$, $P(k)$ will have an even number of zeros, which is considered to be topologically trivial. If $P(0)P(\infty)<0$, $P(k)$ has an odd number of zeros, corresponding to a topologically nontrivial state. On the other hand, the spin-Chern number classification evaluates the difference between $P(\infty)$ and $P(0)$. Clearly, whenever the Z_2 index is nontrivial, the spin-Chern number $C_+-C_-=P(\infty)-P(0)$ must be nonzero. In general, the inverse proposition is not true. Therefore, the Z_2 nontrivial systems represent a subset of the spin-Chern insulators. The spin-Chern numbers are so far found to yield an equivalent description to the Z_2 index for TR-invariant systems. We notice that the Z_2 index does not apply to systems without the TR symmetry, as its definition relies on the assumption of the presence of the TR symmetry. In contrast, the spin-Chern numbers remain valid when the TR symmetry is broken.

3. Time-reversal-symmetry-broken QSH effect

The QSH effect was considered to be closely related to the TR symmetry which provides a protection for the edge states and the Z_2 invariant. An open question is whether we can have a QSH-like phase in a system where the TR symmetry is broken. We have carried out an investigation[30] of the Kane and Mele model by including an exchange field, which breaks the TR symmetry. We calculated the spin-Chern numbers C_s analytically, and used the integer invariants to distinguish different topological phases in the model. We found a TR symmetry-broken QSH phase with $C_\pm=\pm1$, indicating that the nontrivial topological properties of the bulk energy bands remain intact, regardless of the broken TR symmetry. As the exchange field is beyond a critical value, the bulk band gap closes and reopens, and the system enters a quantum anomalous Hall (QAH) phase[6] with $C_\pm=1$ (or -1). By further inclusion of an alternating sublattice potential, we showed that the transition from the TR-symmetry-broken QSH phase to the ordinary insulator state is generally accompanied by closing of the band gap. Our conclusion extended the conditions under which the topological QSH state of matter can happen, and opened the door to the magnetic manipulation of the QSH effect. Let us now introduce the theory[30] in detail.

3.1. Model and phase diagram

We begin with the Kane and Mele model defined on a 2D honeycomb lattice with the Hamiltonian

$$\begin{aligned}H\;=\;&-t\sum_{\langle ij\rangle}c_i^\dagger c_j+\frac{\mathrm{i}V_\mathrm{so}}{\sqrt{3}}\sum_{\langle\langle ij\rangle\rangle}c_i^\dagger\boldsymbol{\sigma}\cdot(\boldsymbol{d}_{kj}\times\boldsymbol{d}_{ik})c_j\\&+\mathrm{i}V_\mathrm{R}\sum_{\langle ij\rangle}c_i^\dagger\hat{\mathrm{e}}_z\cdot(\boldsymbol{\sigma}\times\boldsymbol{d}_{ij})c_j+g\sum_i c_i^\dagger\sigma_z c_i.\end{aligned} \tag{12}$$

Here, the first term is the usual nearest neighbor hopping term with $c_i^\dagger=(c_{i\uparrow}^\dagger,c_{i\downarrow}^\dagger)$ being the electron creation operator on site i and angular bracket $\langle i,j\rangle$ standing for the nearest-neighboring sites. The second term is the intrinsic spin–orbit coupling with coupling strength V_so, where $\boldsymbol{\sigma}$ is the Pauli matrix, i and j

are two next nearest neighbor sites, k is their unique common nearest neighbor, and vector $\boldsymbol{d}_{ik}$ points from k to i. The third term stands for the Rashba spin–orbit coupling with coupling strength $V_{\rm R}$, and the last term represents a uniform exchange field of strength g. For convenience, we set $\hbar$, t, and the distance between next nearest neighbor sites all to be unity.

Hamiltonian (12) is expanded in the long-wavelength limit at Dirac points K and K' to the linear order in the relative wave vector $k = \sqrt{k_x^2 + k_y^2}$. The eigenenergies for the two valence bands near Dirac point K are given by

$$E_{1,2} = \frac{\sqrt{3}}{2}\left(-\sqrt{k^2 + (V_{\rm R} \pm V_{\rm so})^2} \pm V_{\rm R}\right)$$

with subscripts 1 and 2 representing two bands originating from the sublattice degrees of freedom. The wavefunctions will be denoted as $\varphi_{1,2}(\boldsymbol{k})$. The energy gap Δ_E between the conduction and valence bands is plotted in Fig. 1(a) as a function of $|g|/V_{\rm so}$ for some different values of $V_{\rm R}/V_{\rm so}$. It is found that, for $V_{\rm R} < V_{\rm so}$ with increasing $|g|$ from 0, the gap first decreases; as $|g|$ reaches a critical value $g_{\rm c}$, the gap closes at the $k = 0$ point; and as $|g|$ further increases, the gap reopens. The critical exchange energy $g_{\rm c}$ is determined by the condition of touching the conduction and valence bands. For $V_{\rm R} < V_{\rm so}$, we have

$$\frac{g_{\rm c}}{V_{\rm so}} = \frac{\sqrt{3}}{2}\left[1 - \left(\frac{V_{\rm R}}{V_{\rm so}}\right)^2\right]. \tag{13}$$

It indicates that $g_{\rm c}$ decreases with increasing $V_{\rm R}/V_{\rm so}$. For $V_{\rm R} \geq V_{\rm so}$, we have $g_{\rm c} = 0$, and the band gap always exists for finite g. As will be shown below, the insulating state for $|g| < g_{\rm c}$ corresponds to the QSH phase, while that for $|g| > g_{\rm c}$ is also topologically nontrivial with gapless chiral edge states, exhibiting a quantized charge Hall conductance.

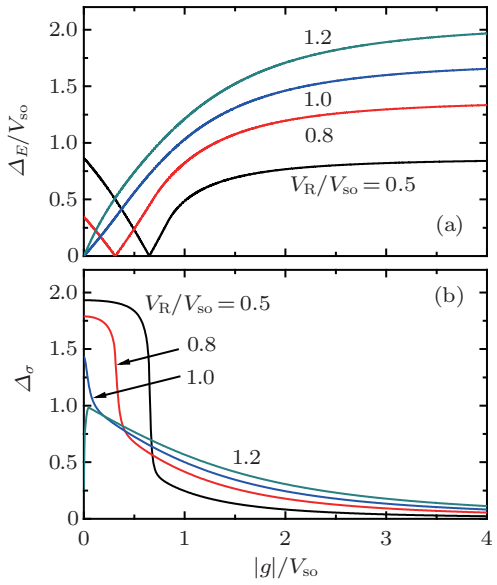


Fig. 1. (a) Normalized energy band gap $\Delta_E/V_{\rm so}$ and (b) spectrum gap of $\hat{\sigma}_z$ as functions of $|g|/V_{\rm so}$ for some different values of $V_{\rm R}/V_{\rm so}$.[30]

To define the spin-Chern number $C_{\rm s}$, a smooth decomposition of the occupied valence band into two sectors is carried out through the diagonalization of the electron spin operator $\hat{\sigma}_z$ in the valence band. Since $\hat{\sigma}_z$ commutes with momentum, the decomposition can be done for each $\boldsymbol{k}$ separately. The minimal spectrum gap Δ_σ between the eigenvalues of $\hat{\sigma}_z$ as a function $|g|/V_{\rm so}$ for different values of $V_{\rm R}/V_{\rm so}$ is plotted in Fig. 1(b). The spectrum gap is always nonzero, so we can unambiguously calculate the corresponding spin-Chern numbers. It is found that for $|g| < g_{\rm c}$, $C_\pm = \pm 1$. Out of this region, we obtain $C_\pm = 1$ for $g > g_{\rm c}$, and $C_\pm = -1$ for $g < -g_{\rm c}$. Figure 2 shows a phase diagram determined by the spin-Chern numbers in the $g/V_{\rm so}$ versus $V_{\rm R}/V_{\rm so}$ plane. There are three topologically distinct phases characterized by $C_\pm = \pm 1$, $C_+ = C_- = 1$, and $C_+ = C_- = -1$, respectively. From our calculation, the boundary between the different topological phases is just the condition of closing the band gap.

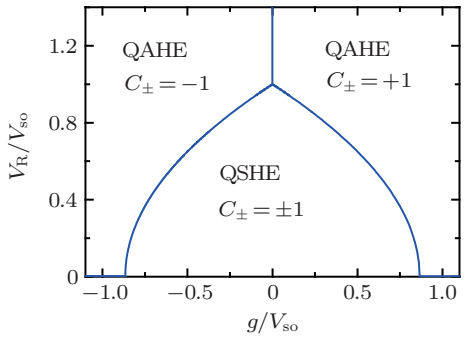


Fig. 2. Phase diagram determined by the Chern numbers in the $V_{\rm R}/V_{\rm so}$ versus $g/V_{\rm so}$ plane. The phase diagram in the half plane of $V_{\rm R}/V_{\rm so} < 0$ is mirror symmetric to that of $V_{\rm R}/V_{\rm so} > 0$, and hence not plotted.[30]

3.2. Properties of edge states

To study the edge states in each region, we calculate the energy spectrum of a long ribbon with zigzag edges and 240 zigzag chains across the ribbon. For $|g| < g_{\rm c}$, corresponding to the $C_\pm = \pm 1$ region in the phase diagram, the energy spectrum is shown in Fig. 3(a). One can easily distinguish the edge states from the bulk states. There is a small energy gap in the edge modes due to the absence of TR and inversion symmetries, as can be seen from the inset of Fig. 3(a). At a given Fermi level in the band gap, there exist four different edge states labeled as A, B, C, and D. Through the analysis of the spatial distributions of the wave functions, one can find that states A and B localize near one boundary of the ribbon, while C and D localize near the other boundary. Take states A and B on one boundary for example. From the slopes of dispersion curves at points A and B, it is easy to determine that the two edge states are counterpropagating. We also examine the spin polarization of the wave functions, state A being almost fully spin-up polarized, and state B spin-down polarized. Therefore, in the $C_\pm = \pm 1$ region there exist two counterpropagating edge states with opposite spin polarizations on the sample

edge, which give rise to no net charge transfer but contribute to the transport of spin.

The characteristic of the edge states in the $C_{\pm} = \pm 1$ region with $g \neq 0$ discussed above is very similar to that of the QSH phase at $g = 0$ protected by the TR symmetry. In particular, they have the same spin-Chern numbers $C_{\pm} = \pm 1$, indicating that they belong to the same topological class. As a result, we call it the TR symmetry-broken QSH phase. For the QSH phase protected by the TR symmetry, nonmagnetic impurities do not cause backscattering on each boundary, and the spin transport in the edge states is dissipationless at zero temperature. In the TR symmetry-broken QSH phase, there is usually a weak scattering between forward and backward movers, as evidenced by the small energy gap in the edge state spectrum, leading to a low-dissipation spin transport.

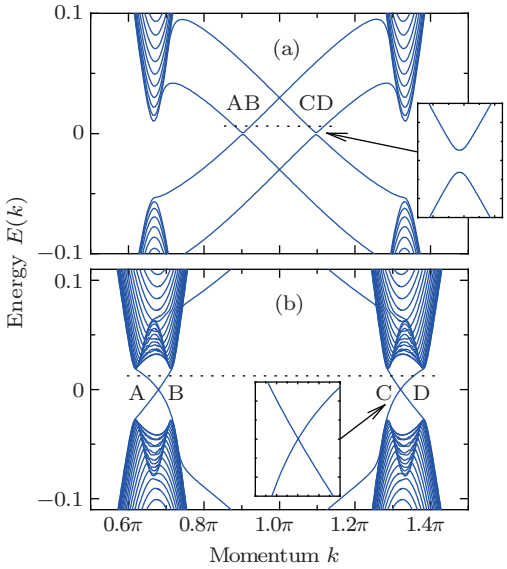


Fig. 3. Energy spectrum of a zigzag-edged graphene ribbon. The parameters are chosen to be $V_{so} = 0.1$, $V_R = 0.05$, and (a) $g = 0.03$ or (b) $g = 0.15$. At a given Fermi level in the band gap, there exist four different edge states, which are labeled as A, B, C, and D.[30]

A similar analysis can be applied to the $C_{\pm} = 1$ region, where the Chern numbers of the filled bands sum up to $C = 2$, corresponding to a QAH phase with Hall conductivity $\sigma_{xy} = 2e^2/h$. The related edge state spectrum is shown in Fig. 3(b). It is found that states A and C localize at one boundary and propagate along the same $-x$ direction, while states B and D localize at the other boundary and propagate along the same $+x$ direction. As a result, in the QAH phase, two edge states at each boundary lead to spin-up and spin-down currents propagating along the same direction, yielding a quantized charge conductance. The symmetry-broken QSH and QAH phases are topologically distinct. The topological phase transition between them can occur at $|g| = g_c$ where the band gap just closes. We mention here that there is great revived interest[31–35] in the QAH phase in TIs, and the QAH effect has been realized experimentally very recently.[36]

To further investigate the transition from the TR-symmetry-broken QSH phase to an ordinary insulator state, we include an alternating sublattice potential $M\sum_i \tau c_i^{\dagger} c_i$ into Hamiltonian (12) with $\tau = \pm 1$ for i on sublattices A and B, respectively. For $g \neq 0$ and $M \neq 0$, since both the TR and the two-fold rotation symmetries are lifted, the two Dirac cones at K and K' become asymmetric, leading to an indirect minimal band gap Δ_E between the two Dirac points. We find that for a system initially in the QSH phase of $|g| < g_c$ with g_c given by Eq. (13), as $|M|$ is increased, the indirect band gap closes at a smaller critical value $|M| = M_-$ and reopens at a greater critical value $|M| = M_+$ with $M_{\mp} = g_c \mp |g|$. The conduction and valence bands overlap in between, i.e., $\Delta_E = 0$ for $M_- \leq |M| \leq M_+$. The spin-Chern numbers are calculated from the lattice model by projecting the two valence bands into two spin sectors in a similar manner to that shown above for the continuum model. For the phase diagram shown in Fig. 2, numerical calculations based on the lattice model have been performed, the obtained result is found to agree with that from the continuum model. The spin-Chern numbers are well defined only in the regions $|M| < M_-$ and $|M| > M_+$, where $\Delta_E > 0$. We find $C_{\pm} = \pm 1$ for $|M| < M_-$ and $C_{\pm} = 0$ for $|M| > M_+$. The phase diagram in g versus M plane obtained is plotted in Fig. 4(a). The calculated $C_{\pm}$ and the minimal band gap Δ_E varying along the arrowed dashed line in Fig. 4(a) are shown in Fig. 4(b) as functions of M/V_{so}. A general feature of the phase diagram is that the transition between the QSH phase with $C_{\pm} = \pm 1$ and the ordinary insulator state with $C_{\pm} = 0$ is always accompanied by closing of the band gap, which serves as another signature that the TR-symmetry-broken QSH phase is topologically nontrivial and distinct from the ordinary insulator.

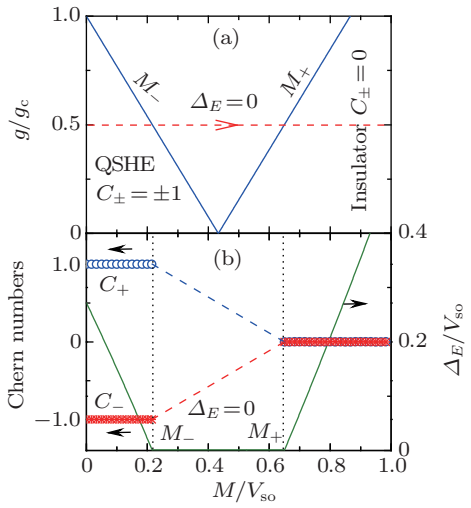


Fig. 4. (a) Phase diagram on the g/g_c versus M/V_{so} plane for $g_c = 0.5 \times (\sqrt{3}/2)V_{so}$, and (b) calculated Chern numbers and band gap as functions of M/V_{so} for $g = 0.5g_c$ and $V_{so} = 0.1$.[30]

It is known that topological invariants must be carried by extended states. It becomes fundamentally important to show

that extended states in the QSH systems survive TR symmetry breaking. By using the well-established scaling theory for Anderson localization, we have shown that indeed there exist extended states in the QSH systems when the TR symmetry is broken, as long as the spin-Chern numbers are quantized to nonzero values.[37,38]

4. General relation between edge states and bulk spin-Chern numbers

Now we have shown that the nontrivial topological properties of the QSH state can survive weak perturbations which break the TR symmetry. We then turn to the properties of the edge states in general conditions without any fictitious symmetry constraints. In order to reveal the general characteristics of the edge states and their connection to the bulk topological invariant in a QSH system, we have presented a topological argument[39] similar to the Laughlin's gedanken experiment[28] without considering any symmetries. We showed that as required by the nontrivial spin-Chern numbers and the gauge invariance, edge states must appear near the sample boundary, for which either the energy gap or the spin spectrum gap needs to close. These two different scenarios lead to gapless or gapped edge modes, respectively. In particular, it was demonstrated that gapless edge states can appear in a TR symmetry-broken QSH system by tuning the confining potential at the boundaries. They are associated with the bulk topological invariants, and are robust against relatively smooth impurity scattering potentials. Our result offered an interesting example for counterpropagating gapless edge states that are not protected by symmetries, which shed some light on the underlying mechanism of the QSH effect in a broad sense. Let us go into more details of the work.[39]

4.1. A general gauge argument

We first look back on a looped ribbon of the QH system with a magnetic flux ϕ (in units of flux quantum hc/e) threading the ring adiabatically. The Fermi energy $E_{\rm F}$ is assumed to lie in an energy gap. In the spirit of the Laughlin argument,[28] increasing ϕ from 0 to 1 effectively pumps one occupied state from one edge to the other, giving rise to the transfer of one charge between the edges, essentially because there is a nonzero Chern number (Hall conductivity) in the bulk. On the other hand, the system Hamiltonian is gauge-invariant under integer flux changes, i.e., if ϕ is increased from 0 to 1, the system will reproduce the same eigenstates as at $\phi = 0$. To assure this gauge invariance, there must be gapless edge modes on the edges (when the perimeter of the ring is large), so that the spectral flow can form a closed loop, as illustrated in Fig. 5(a), along which the electron states can continuously move with changing ϕ.

We then propose a topological understanding of the QSH system in terms of the same looped ribbon geometry. The occupied valence band can be decomposed into two spin sectors by using the projected spin operator. The unoccupied conduction band can be divided similarly. The two spin sectors are separated by a nonzero spin spectrum gap in the bulk. They carry opposite spin-Chern numbers, so that increasing ϕ pumps a state of the spin-up sector in the occupied band from one edge to the other, and pumps another state of the spin-down sector in the opposite direction. In order for the system to recover the initial eigenstates as ϕ changes from 0 to 1, the spectral flow needs to form closed loops, similarly to the QHE system. However, for the QSH system, if not enforced by any symmetry, two different scenarios can occur on the edges. One is that gapless edge modes appear on the edges, so that states can move between the conduction and valence bands with changing ϕ to form closed loops in the spin-up and spin-down sectors separately, as shown in Fig. 5(b). In this case, the states in the two loops cannot evolve into each other due to the nonvanishing spin spectrum gap both in the bulk and on the edges. The other scenario is that a closed loop of spectral flow is formed between the two spin sectors within the valence (or conduction) band, as shown in Fig. 5(c). In this case, the spin spectrum gap must vanish on the edges, but the energy gap may remain open in the edge state spectrum.

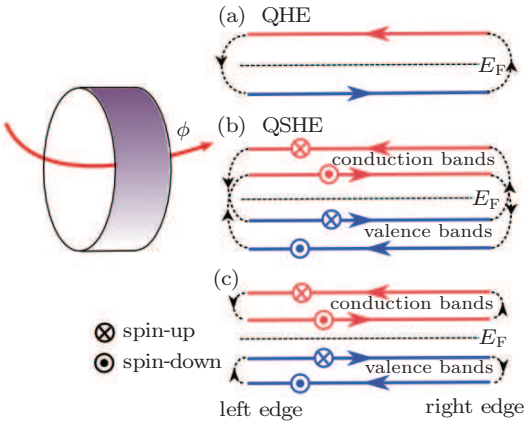


Fig. 5. (a) Schematic diagram of the flow of electron states in a looped ribbon of the QHE system, with adiabatically increasing the magnetic flux ϕ that threads the ring. The bulk electron states below Fermi energy $E_{\rm F}$ drift from left to right, gapless edge modes ascend through $E_{\rm F}$ on the right edge, states above $E_{\rm F}$ drift from right to left, and gapless edge modes descend through $E_{\rm F}$ on the left edge, forming a closed loop. (b) In the first scenario for the QSH system, gapless edge modes appear on the edges of the ring, so that electron states in each spin sector behave like that in the QHE system, but those in two spin sectors move in opposite directions. (c) In the second scenario for the QSH system, the edge states are gapped, whereas the spin spectrum gap closes on the edges. In the bulk, electron states in the two spin sectors drift in the same way as in that in panel (b), but on the edges they join together within the valence (conduction) band.[39]

4.2. Gapped and gapless edge states

To demonstrate the above two different scenarios, we consider a sample of strip geometry with the Hamiltonian given

in Eq. (12). Besides, we also include an alternating sublattice potential with strength H_C representing a sublattice staggered confining potential, which is given by $H_C = \sum_i V_i c_i^\dagger c_i$ with

$$V_i = \pm V_0 \left(e^{-x_i/\xi} + e^{-(x_{N_x}-x_i)/\xi} \right), \tag{14}$$

where $\pm$ is taken to be positive for sites on sublattice A (solid dots) and negative on sublattice B (hollow dots), as shown in Fig. 6(a). In Eq. (14), V_i is strongly x-dependent across the ribbon, and equals to $\pm V_0$ at the edges ($x_i = 0$ and $x_i = x_{N_x}$). It decays exponentially away from the edges with a characteristic length ξ, as shown in Fig. 6(b). When the ribbon width is much greater than ξ, V_i essentially vanishes in the middle region of the ribbon. Here, we note the case of a uniform staggered potential $V_i = \pm V_0$ ($|V_i|$ being independent of x_i). As has been shown above, with increasing $|V_i|$, there is a transition from the TRS-broken QSH phase to an ordinary insulator state, where the middle band gap closes and then reopens. Therefore, for the confining potential V_i given by Eq. (14) with a large V_0, the ribbon in Fig. 6 can be regarded as a TRS-broken QSH ribbon sandwiched in between two trivial band insulators.

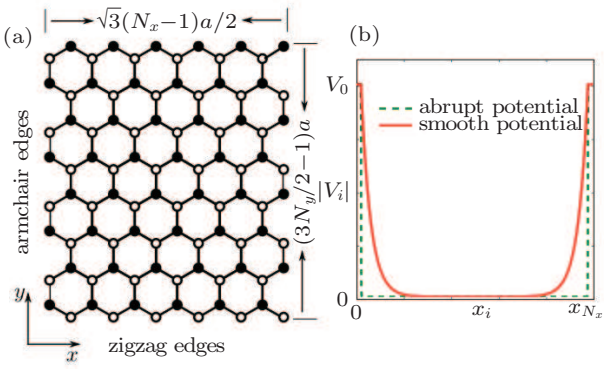


Fig. 6. (a) Schematic diagram of an armchair honeycomb lattice ribbon with atom sites in two sublattices separately labeled by solid and hollow dots, where a is the distance between nearest-neighbor sites. (b) Profiles of $|V_i|$ as functions of x_i for an abrupt confining potential (dashed line) and a relatively smooth confining potential (solid line).[39]

In order to assure the system in the TRS-broken QSH state, we set the parameters $V_{so} = 0.1t$, $V_R = 0.1t$, and $g = 0.1t$. The length of the armchair ribbon N_y is taken to be infinite. The energy spectrum of the ribbon, together with the corresponding eigenfunctions $\varphi_m(k_y)$, can be numerically obtained by diagonalizing the Hamiltonian for each momentum k_y in the y direction. The calculated energy spectrum of the armchair ribbon with width $N_x = 240$, for the confining potential with $V_0 = 12t$ fixed and two different decay lengths, is plotted in Figs. 7(a) and 7(b). One can easily see that the edge states appear as thin lines in the middle bulk band gap of the energy spectrum. In Figs. 7(a) and 7(b), the spin polarization of the edge states is labeled with $\uparrow$ and $\downarrow$, indicating that two spin-filtered channels on each edge flow along opposite directions. For the nearly hard-wall confining potential of $\xi = 0.1$, the sublattice potential V_i is nonzero only on the outermost armchair lines. In this case, two small energy gaps are observed in the edge state spectrum shown in Fig. 7(a) as a consequence of the broken TRS, in agreement with the previous observation. With increasing the decay length ξ of the confining potential, the energy gaps of the edge states become smaller and smaller. As ξ is large enough, e.g., $\xi = 4$ in Fig. 7(b), interestingly, the edge states become gapless.

We calculate the k_y-dependent spectrum of the projected spin operator $P\hat{\sigma}_z P$, whose matrix elements are given by $\langle \varphi_m(k_y) | \hat{\sigma}_z | \varphi_n(k_y) \rangle$ with m and n running over all the occupied states. By diagonalizing this matrix, the spectrum of the projected spin $P\hat{\sigma}_z P$ can be obtained. For the Kane–Mele model given by Eq. (12), if $V_R = 0$, $\hat{\sigma}_z$ commutes with Hamiltonian H. It follows that the spectrum of $P\hat{\sigma}_z P$ consists of just two values ± 1, which are highly degenerate. When the Rashba term is turned on, $\hat{\sigma}_z$ and H no longer commute, and the degeneracy is lifted. In this case, the spectra of $P\hat{\sigma}_z P$ between $+1$ and -1 spread towards the origin, but a gap remains in a bulk sample if the amplitude of the Rashba term dose not exceed a threshold.

For the ribbon geometry, the situation is more complicated due to the existence of the edges, and numerical calculations are performed to obtain the spin spectrum. The calculated spectrum of $P\hat{\sigma}_z P$ for the same parameters as those used in Figs. 7(a) and 7(b) is shown in Figs. 7(c) and 7(d), which exhibits a very interesting behavior. For the hard-wall confining potential with $\xi = 0.1$, while the energy spectrum of edge states is slightly gapped, the spectrum of $P\hat{\sigma}_z P$ continuously changes between $+1$ and -1 with increasing k_y without showing any gap, corresponding to the second scenario shown in Fig. 5(c). On the other hand, for a relatively smooth confining potential with $\xi = 4.0$, the energy spectrum of the edge states is gapless, but the spectrum of $P\hat{\sigma}_z P$ displays a big gap, and sudden changes happen at the cross points in the energy spectrum of the edge states, corresponding to the first scenario shown in Fig. 5(b). From Fig. 7, it follows that as long as the system is in the QSH state, a gapless characteristic always appears either in the energy spectrum of edge states or in the spectrum of $P\hat{\sigma}_z P$, leading to the two types of closed loops for the continuous flow of electron states illustrated in Figs. 5(b) and 5(c).

The result shown in Figs. 7(b) and 7(d), for the relatively smooth confining potential, is of particular interest. It indicates that gapless edge states can exist in the TRS-broken QSH system, accompanied with a gapped spectrum of $P\hat{\sigma}_z P$. Such an interesting behavior can be further understood by the following argument. As long as the bulk energy gap does not close, the projected spin operator $P\hat{\sigma}_z P$ is exponentially localized in the real space with a characteristic length about $\lambda \sim \hbar v_F / \Delta_E$, where v_F is the Fermi velocity and Δ_E is the

magnitude of the energy gap. For the parameter set used in Fig. 7, λ is estimated to be between 1 and 2 lattice constants. When the confining potential V_i is varying relatively slowly in space, i.e., $\xi \gg \lambda$, one can find that $P\hat{\sigma}_z P$ roughly commutes with the confining potential. In this case, the confining potential is of no influence to the spectrum of $P\hat{\sigma}_z P$. Since the spin spectrum has a gap in the bulk, this gap remains open on the smooth edges, as seen in Fig. 7(d). As a result, the energy gap has to close due to the topological requirement, resulting in gapless edge modes, as observed in Fig. 7(b). The gapless nature of the edge states is found to remain robust in the presence of a weak long-range correlated disorder.

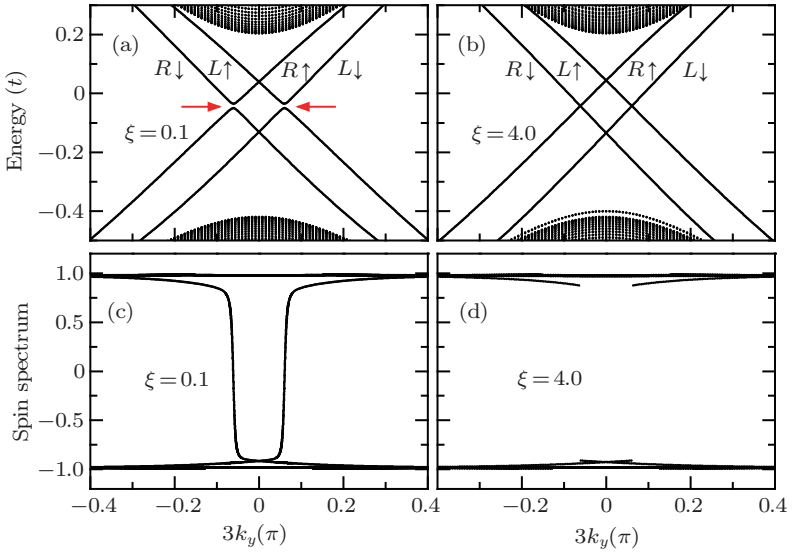


Fig. 7. Energy spectra (panels (a) and (b)) and spectra of the projected spin $P\hat{\sigma}_z P$ (panels (c) and (d)) of an armchair ribbon for $\xi = 0.1$ (panels (a) and (c)) and $\xi = 4$ (panels (b) and (d)), in which L (R) stands for states on the left (right) edge, and $\uparrow$ ($\downarrow$) for the up (down) spin polarization. The horizontal arrows in panel (a) point to the small energy gaps in the energy spectrum. At $\xi = 0.1$, while the energy spectrum in panel (a) is gapped, the spin spectrum in panel (c) is gapless. At $\xi = 4$, the energy spectrum in panel (b) is gapless, however the spin spectrum in panel (d) is gapped.[39]

5. Stabilization of the QSH effect

As pointed out in the introduction, the QSH effect protected the TR symmetry is unstable in realistic environments where perturbations violating the TR symmetry are often unavoidable. It is therefore highly desirable to search for QSH effect that is protected by topology alone, rather than by any symmetries. The study in the preceding sections has confirmed the possibility to realize the QSH effect, characterized by gapless edge states, in the absence of the TR symmetry. However, the proposed method to tune the confining potential at the sample boundary is difficult in practice. We have shown above that the nontrivial bulk band topology of the QSH systems remains intact, even when the TR symmetry is broken, implying that the instability of the QSH effect is solely due to the properties of the edge states.

In a recent work,[40] we proposed a more practical way to realize the QSH effect without the TR symmetry, which is expected to be stable to symmetry-breaking perturbations. We showed that creating a narrow ferromagnetic (FM) region near an edge of a QSH sample can push one of the counterpropagating edge states to the inner boundary of the FM region, and leave the other on the outer boundary. The moving directions and the spin orientations of the individual edge states are unchanged, so that the QSH effect persists. Importantly, the edge states are spatially separated, so that the QSH effect becomes robust against general perturbations without fictitious symmetry constraints. We presented both qualitative discussion and quantitative calculation to demonstrate the physical picture and the practical feasibility of this proposal. Our work paved a road to realize a robust QSH effect via magnetic manipulation.

5.1. Model Hamiltonian

We start from the effective Hamiltonian for an HgTe quantum well[5] with an exchange field given by $H = H_0 + H_1$ with

$$H_0 = v_{\rm F}(\hat{\tau}_z k_x \hat{\sigma}_x + k_y \hat{\sigma}_y) + Dk^2 + (M_0 - Bk^2)\hat{\sigma}_z \,. \qquad (15)$$

Here, $v_{\rm F}$, D, M_0, and B are the parameters of the model, $\hat{\tau}$ stands for Pauli matrices for two spin states, and $\hat{\sigma}$ for the electron and hole bands. An exchange field can be created in the HgTe layer by doping magnetic atoms, such as Mn.[33] Within the mean-field approximation, the exchange field can be described by[33] $H_1 = (g_0\hat{\sigma}_z + g_1)\hat{\tau}_z$, where $g_0 = \frac{1}{2}(G_{\rm H} - G_{\rm E})$ and $g_1 = \frac{1}{2}(G_{\rm H} + G_{\rm E})$ with $2G_{\rm E}$ ($2G_{\rm H}$) being the exchange splitting of the electron (hole) bands. For convenience, we set the reduced Planck constant $\hbar$ to be unity. And $B^2 > D^2$ is assumed to ensure the valence bands to be inverted.[11,41,42]

It is worth mentioning that the model Hamiltonian (15) can describe not only the HgTe quantum wells, but also a class of QSH systems.[43] For example, setting $g_1 = 0$ and making a unitary transformation $\mathscr{H} = U^\dagger H U$ with

$$U = \frac{1}{2}[(1+\hat{\tau}_z)+(1-\hat{\tau}_z)\hat{\sigma}_y]\exp\left(-\mathrm{i}\frac{\pi}{4}\hat{\sigma}_z\right),$$

we obtain

$$\mathscr{H} = v_\mathrm{F}(k_y\hat{\sigma}_x - k_x\hat{\sigma}_y) + \left(M_0 - Bk^2\right)\hat{\tau}_z\hat{\sigma}_z + Dk^2 + g_0\hat{\sigma}_z.$$

One can find that this Hamiltonian is identical to that of a thin film of 3D TI Bi_2Se_3 with an exchange field,[11,41,42] for which $\hat{\boldsymbol{\sigma}}$ stands for spin up and spin down, and $\hat{\boldsymbol{\tau}}$ for bonding and antibonding of the surface states on two surfaces.

Since $\hat{\tau}_z$ is a conserved quantity, one can easily diagonalize Eq. (15), and obtain two conduction bands and two valence bands. Under the condition of $|g_1| < \max(|M_0|, |g_0|)$, a nonzero middle band gap exists, except at $g_0 = \pm M_0$ where the conduction and valence bands touch; otherwise the conduction and valence bands overlap and the system is a metal. As has been discussed in Ref. [33], for HgTe quantum wells doped with Mn, G_E and G_H have opposite signs, so the above condition is satisfied. Given $|g_1| < \max(|M_0|, |g_0|)$, the spin-Chern numbers for $\tau_z = \pm 1$ can be derived to be

$$C_\pm = \pm\frac{1}{2}[\mathrm{sgn}(B) + \mathrm{sgn}(M_0 \pm g_0)]\,. \tag{16}$$

At $g_0 = 0$, $C_\pm = \pm\mathrm{sgn}(B)$ if $BM_0 > 0$, corresponding to a QSH phase, and $C_\pm = 0$ if $BM_0 < 0$, corresponding to an ordinary insulator. We focus on systems with $BM_0 > 0$, and for the sake of definiteness, we will confine ourselves to the parameter region of $B < 0$ and $M_0 < 0$, which is the case of the HgTe quantum wells exhibiting the QSH effect. (All the conclusions reached in this work do also apply to $B > 0$ and $M_0 > 0$.) In this case, $C_\pm = (-1, 1)$ at $g_0 = 0$. With increasing g_0 to $g_0 = |M_0|$, $C_\pm$ undergo a transition from $(-1, 1)$ to $(0, 1)$.

5.2. Properties of edge states

Next we consider a QSH sample with a strip geometry, as shown in Fig. 8. Since the spin-Chern numbers in Eq. (16) do not depend on D and g_1 for $|g_1| < \max(|M_0|, |g_0|)$, without loss of generality, we set $D = g_1 = 0$ for now to make a physical discussion, and the effect of finite D and g_1 will be taken into account in numerical calculation later. The exchange field g_0 is taken to be nonzero in narrow regions of width l near the two edges of the strip, namely, region I of $0 < y < l$ and region II of $(L-l) < y < L$, and vanishes in region III of $l < y < (L-l)$. The system as a whole has a bulk energy gap around energy 0 for $g_0 \neq \pm M_0$. The edge states in the bulk energy gap can be solved analytically by replacing k_y with $-\mathrm{i}\partial_y$ in the system Hamiltonian. For $\tau_z = 1$, an edge mode with energy $E_+(k_x) = v_\mathrm{F}k_x$ is found on one side of the strip, whose wavefunction is given by

$$\varphi_+(k_x, y) = |1, 1\rangle\phi_+(k_x, y). \tag{17}$$

Here, $|\tau_z, \sigma_x\rangle$ with $\tau_z = \pm 1$ and $\sigma_x = \pm 1$ is used to represent the common eigenstate of $\hat{\tau}_z$ and $\hat{\sigma}_x$. The spatial wavefunction is $\phi_+(k_x, y) = C[\mathrm{e}^{-y/\xi_1(g_0)} - \mathrm{e}^{-y/\xi_2(g_0)}]$ for $y < l$, and $\phi_+(k_x, y) = D_1\,\mathrm{e}^{-(y-l)/\xi_1(0)} - D_2\,\mathrm{e}^{-(y-l)/\xi_2(0)}$ for $y \geq l$, with C, D_1, and D_2 to be determined from the conditions of continuity and normalization of $\phi_+(k_x, y)$. The two characteristic length functions are defined as

$$\xi_{1,2}(\varepsilon) = \frac{2|B|}{v_\mathrm{F} \pm \sqrt{v_\mathrm{F}^2 - 4B(M_0 - Bk_x^2 + \varepsilon)}}\,. \tag{18}$$

For $\tau_z = -1$, we find another edge mode with energy $E_-(k_x) = -v_\mathrm{F}k_x$ and wavefunction

$$\varphi_-(k_x, y) = |-1, 1\rangle\phi_-(k_x, y), \tag{19}$$

where $\phi_-(k_x, y) = E[\mathrm{e}^{-y/\xi_1(-g_0)} - \mathrm{e}^{-y/\xi_2(-g_0)}]$ for $y < l$, and $\phi_-(k_x, y) = F_1\,\mathrm{e}^{-(y-l)/\xi_1(0)} - F_2\,\mathrm{e}^{-(y-l)/\xi_2(0)}$ for $y \geq l$. Owing to the two-fold rotation symmetry, the edge modes on the other side of the strip have dispersion relations $E_\pm(k_x) = \mp v_\mathrm{F}k_x$. Their wavefunctions can be obtained through replacements $y \to L - y$, $\sigma_x \to -\sigma_x$, and $k_x \to -k_x$ in Eqs. (17) and (19). The energy spectrum of the edge states appears to have a mirror symmetry between k_x and $-k_x$. However, this is an accidental event, and the system does not possess the mirror symmetry if $g_0 \neq 0$.

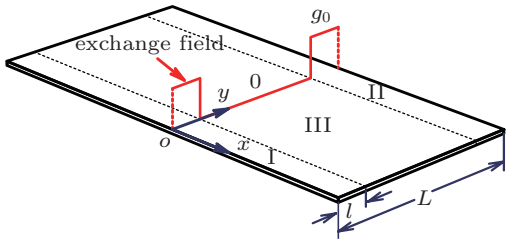


Fig. 8. Schematic diagram of the QSH sample with a long strip geometry. The profile of the y-dependent exchange field is shown by the thick (red) line.[40]

In Eqs. (17) and (19), the $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\sigma}}$ parts of wavefunctions do not change with varying g_0. In Fig. 9(a), the modulus squared of spatial wavefunctions $\phi_+(0, y)$ and $\phi_-(0, y)$ are plotted as functions of y for several values of g_0. Here, the momentum is taken to be dimensionless by properly choosing the units for v_F and B, and v_F is used as the unit of energy. At $g_0 = 0$, we have $|\phi_+(0, y)|^2 = |\phi_-(0, y)|^2$, and the two lines coincide with each other, which is required by the TR symmetry at $g_0 = 0$, as mentioned above. With increasing g_0, the peak of $|\phi_-(0, y)|^2$ becomes sharper and closer to $y = 0$. On the contrary, the shape of $|\phi_+(0, y)|^2$ widens with increasing g_0,

and spreads across region I at $g_0 = |M_0| = 0.2$. With further increasing g_0, $|\phi_+(0,y)|^2$ becomes localized near $y = l = 100$, i.e., the inner boundary of region I. This transition is directly caused by the sign change of $\xi_2(g_0)$ in $\phi_+(0,y)$ at $g_0 = |M_0|$, from positive to negative. The other characteristic lengths in $\phi_+(0,y)$ and $\phi_-(0,y)$ remain positive during this process.

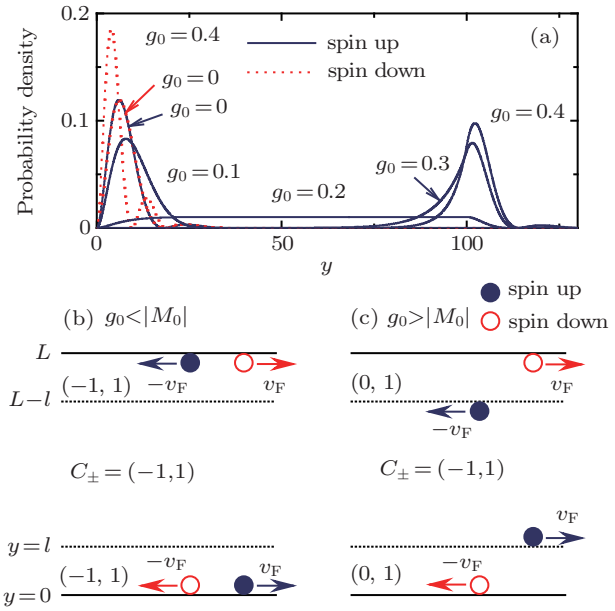


Fig. 9. (a) $|\phi_+(k_x,y)|^2$ and $|\phi_-(k_x,y)|^2$ at $k_x = 0$ as functions of y for different g_0, where $v_F = 1$, $B = -5$, $M_0 = -0.2$, and $l = 100$ are taken. Here, $|\phi_-(k_x,y)|^2$ is plotted only for two values of g_0, as its change with g_0 is relatively small. Spatial distributions of spin-Chern numbers $C_\pm$ and edge states for (b) $g_0 < |M_0|$ and (c) $g_0 > |M_0|$.[40]

The evolution of the edge states with varying g_0 is further illustrated in Figs. 9(b) and 9(c), and can be understood in terms of calculated spin-Chern numbers. For $g_0 < |M_0|$, the spin-Chern numbers $C_\pm$ in the three regions take the same value $(-1,1)$. This indicates that the three regions are topologically equivalent, and can be regarded as a QSH system as a whole. As a result, the edge states of both up spin ($\tau_z = 1$) and down spin ($\tau_z = -1$) appear near the sample boundaries $y = 0$ and $y = L$, as shown in Fig. 9(b). For $g_0 > |M_0|$, the situation is quite different, because $C_\pm$ undergo a transition at $g_0 = |M_0|$ from $(-1,1)$ to $(0,1)$ in regions I and II. For the spin-down electrons, the three regions have the same Chern number $C_- = 1$, and as a whole are equivalent to a QH system. Therefore, the spin-down edge states remain localized near $y = 0$ and $y = L$. For the spin-up electrons, region III with $C_+ = -1$ is a QH system, sandwiched between two insulators in regions I and II with $C_+ = 0$. The spin-up edge states thus shift to their interfaces, namely, $y = l$ and $y = L - l$, as shown in Fig. 9(c). Comparing Fig. 9(c) for $g_0 > |M_0|$ with Fig. 9(b) for $g_0 < |M_0|$, one finds that both systems have very similar edge states: the lower edge contains a right mover with up spin and a left mover with down spin, and conversely for the upper edge, so as to exhibit the same QSH effect. An important difference is that for $g_0 > |M_0|$, the counterflows of electrons at the lower (upper) edge are spatially separated into two different "lanes" located at $y = 0$ ($y = L$) and $y = l$ ($y = L - l$), which provides an essential protection for the edge states against backscattering from symmetry-breaking random potentials. As a result, the QSH state for $g_0 > |M_0|$ should be more stable than that for $g_0 < |M_0|$. To show this point, we introduce a specific perturbation violating the TR symmetry. Since the two edge states given by Eqs. (17) and (19) have opposite τ_z, we include an in-plane Zeeman field throughout the sample with potential

$$V = \gamma \hat{\tau}_y \,, \tag{20}$$

which represents a "maximal" coupling between the two edge states near the degenerate point. A tight-binding model on a square lattice is constructed for the strip geometry, which recovers the form of Eq. (15) in the continuum limit. The energy spectrum is calculated numerically from the tight-binding model.

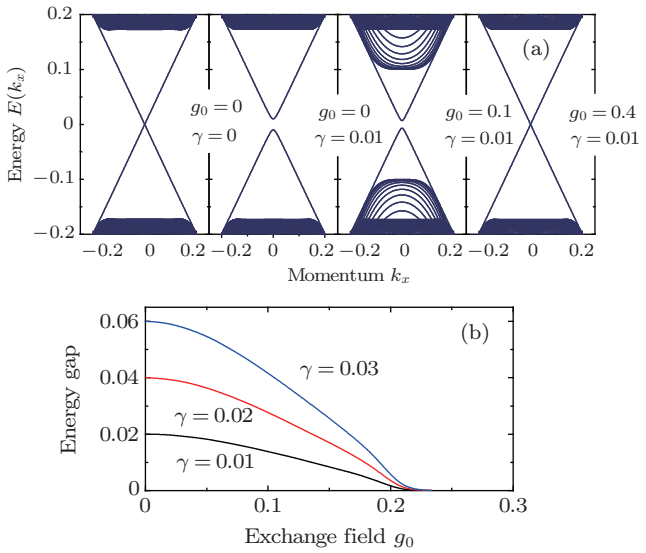


Fig. 10. (a) Calculated energy spectra for four sets of g_0 and γ. (b) Magnitude of the energy gap of edge states as a function of g_0 for three values of γ. The other parameters are taken to be $v_F = 1$, $B = -5$, $M_0 = -0.2$, $L = 1000$, and $l = 200$.[40]

The calculated energy spectra for four sets of g_0 and γ are plotted in Fig. 10. Since the dispersion relation for the edge states is given by $E_\pm(k_x) = \pm v_F k_x$ on one side of the strip and $E_\pm(k_x) = \mp v_F k_x$ on the other side, the energy spectra of the edge states on the two sides collapse, and only two branches of edge states are seen in each figure. For $g_0 = \gamma = 0$, the edge states are gapless, as the system is TR-invariant, and $\hat{\tau}_z$ is conserved as well. Upon turning on $\gamma = 0.01$, an energy gap appears in the edge state spectrum even at $g_0 = 0$, signaling the onset of backscattering. With increasing g_0 to 0.1, the energy gap decreases but remains finite. With further increasing g_0 to 0.4, greater than $M_0 = 0.2$, the energy gap vanishes, an indication of quenching of backscattering. The magnitude of the energy gap of the edge states is plotted as a function of g_0 in the

lower panel of Fig. 10 for three different values of γ. While it increases almost proportionally to γ, the energy gap decreases with increasing g_0. When g_0 slightly exceeds the critical point $M_0 = 0.2$, the energy gap essentially vanishes for all the three values of γ. Furthermore, we find that for $g_0 < M_0$, the spin polarizations of the edge states $\langle \hat{\tau}_z \rangle$ deviate from ± 1 appreciably near the energy gap, as nonzero γ destroys the conservation of $\hat{\tau}_z$, however nearly full spin polarizations are restored for $g_0 > M_0$. These numerical results confirm our argument that the QSH effect in the $g_0 > M_0$ region is robust against general perturbations without symmetry constraints.

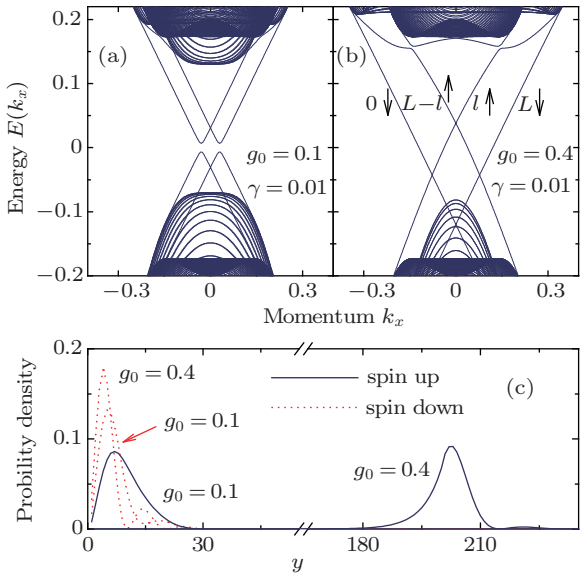


Fig. 11. Calculated energy spectra with $\gamma = 0.01$ and $g_1 = 0.3g_0$ for (a) $g_0 = 0.1$ and (b) $g_0 = 0.4$, and (c) the corresponding spatial probability distributions of the two edge states near the lower edge at energy $E = -0.01$. The other parameters are the same as those in Fig. 10. Here, spin up and spin down refer to the edge states with positive and negative spin polarizations $\langle \hat{\tau}_z \rangle$, as $\hat{\tau}_z$ is not conserved.[40]

Finally, we investigate the effect of nonzero g_1 in the Hamiltonian on the QSH effect. Nonzero g_1 together with g_0 is present in FM regions I and II of Fig. 8. The calculated energy spectra are plotted in Fig. 11 for $g_0 = 0.1$ and 0.4, with $g_1 = 0.3g_0$ set in proportion to g_0. The energy spectra of the edge states on the two sides of the strip are no longer degenerate in the presence of nonzero g_1, the positions and spin orientations of the split edge states are labeled in Fig. 11(b). The inclusion of nonzero g_1 also breaks the particle–hole symmetry. It is found that the edge states are gapped for $g_0 = 0.1$ ($< M_0$) and become gapless for $g_0 = 0.4$ ($> M_0$), as shown respectively in Figs. 11(a) and 11(b). The spatial probability distributions of the two edge states at $E = -0.01$ on one side of the strip are plotted in Fig. 11(c). While the spin-down polarized edge state stays near $y = 0$, the spin-up polarized edge state moves from near $y = 0$ to $y = l = 200$ with tuning g_0 from 0.1 to 0.4, exhibiting the same transition as shown in Fig. 9(a). This result indicates that the presence of nonzero g_1 does not change the physical picture qualitatively, and the robust QSH effect can be achieved when the exchange field is sufficiently strong.

5.3. A more realistic model: effect of disorder

We have shown how the QSH effect can be strengthened by doping magnetic atoms near the edges of a QSH system. Now we consider a more realistic model for HgTe quantum wells doped with Mn atoms ($Hg_{1-c}Mn_cTe$), for which Hamiltonian H_0 is still given by Eq. (15). However, we take into account the fact that the doped-Mn atoms are spatially randomly distributed, and their spins may not be fully aligned. Instead of the assumption of a uniform exchange field, H_1 is taken to be

$$H_1 = -\frac{1}{\pi\lambda^2} \sum_{\alpha=0}^{N_{\rm Mn}-1} (j_0\hat{\sigma}_z + j_1)\hat{\boldsymbol{\tau}} \cdot \boldsymbol{s}_\alpha \times \exp\left(-|\boldsymbol{r} - \boldsymbol{R}_\alpha|^2/\lambda^2\right). \quad (21)$$

Here, factor $(j_0\hat{\sigma}_z + j_1)$ accounts for different electron–spin interaction strengths in the electron and hole bands, $\boldsymbol{r}$ is the electron coordinate operator, $\boldsymbol{R}_\alpha$ is the position of the α-th Mn atom, and $N_{\rm Mn}$ is the total number of the Mn atoms. The local spins ($S = 5/2$) of the Mn atoms are treated as classical vectors, and $\boldsymbol{s}_\alpha$ is a unit vector in the direction of the local spin of the α-th Mn atom. The distribution of the orientations of the local spins is assumed to be Boltzmann–Maxwell-like[44] $f(\boldsymbol{s}_\alpha) \propto \mathrm{e}^{-\eta\cos\theta_\alpha}$, where θ_α is the polar angle of $\boldsymbol{s}_\alpha$, and η is a parameter which can be related to the ratio of magnetization M to saturated magnetization $M_{\rm s}$, $M/M_{\rm s} = -\langle\cos\theta_\alpha\rangle = \coth(\eta) - 1/\eta$. Therefore, for a given ratio $M/M_{\rm s}$, the distribution is fully determined. We will set $M = M_{\rm s}/3$, for which the local spins are randomly oriented to a large degree. As a result, H_1 given by Eq. (21) not only provides an exchange field, but also acts as a scattering potential of magnetic impurities. If one makes the mean-field approximation, by replacing $\boldsymbol{s}_\alpha$ with its average $\langle \boldsymbol{s}_\alpha \rangle = -(M/M_{\rm s})\boldsymbol{e}_z$ and averaging Eq. (21) over a random distribution of $\boldsymbol{R}_\alpha$, equation (21) recovers the form $H_1 = (g_0\hat{\sigma}_z + g_1)\hat{\tau}_z$, where $g_0 = j_0 cM/M_{\rm s}a_0^2$ and $g_1 = j_1 cM/M_{\rm s}a_0^2$ with a_0 being the lattice constant. By using the known expressions for mean-field parameters g_0 and g_1,[33] we get $j_0 = 464$ meV·nm^2 and $j_1 = 286$ meV·nm^2, which are independent of $M/M_{\rm s}$ and doping concentration c, as g_0 and g_1 are proportional to $cM/M_{\rm s}$. The coupling range λ is set to be 10 nm.[45]

The other parameters of the model are taken from Ref. [46]: $v_{\rm F} = 364.5$ meV·nm, $B = -686$ meV·nm^2, $D = -512$ meV·nm^2, and $M_0 = -10$ meV, corresponding to a HgTe quantum well of thickness 7.0 nm. We consider a sample having the strip geometry shown in Fig. 8, with linear sizes $L_x = 80$ nm and $L = 560$ nm in the x and y directions. The doped-Mn atoms are randomly distributed in regions of width $l = 80$ nm near the two edges, with the potential described by Eq. (21). We employ the supercell algorithm,[47] in which

the $L_x \times L$ sample (supercell) is duplicated along the x direction to form a superlattice. A tight-binding model on square meshes is constructed, which recovers the form of Eq. (15) in the continuum limit. The eigenenergies of the superlattice as functions of the longitudinal momentum q_x are calculated by exact diagonalization, and the results for four different doping concentrations are plotted in Figs. 12(a)–12(d). The mesh size is set to be 4 nm, and a good convergence is verified with smaller mesh sizes. (We note that the longitudinal momentum in the supercell model is equivalent to the magnetic flux in the Laughlin's gedanken experiment.[28] Given this equivalence, the two systems have an identical energy spectrum.) For $c = 0.005$, apparent energy gaps exist in the edge state spectrum, indicating the occurrence of backscattering. With increasing c to 0.01, the energy gaps decrease but remain finite. With further increasing c to 0.02 and 0.03, the energy gaps essentially vanish, an indication of quenching of backscattering. In this case, the electron transport through the edge channels is nondissipative, and the presence of a weak disorder has no influence on it.

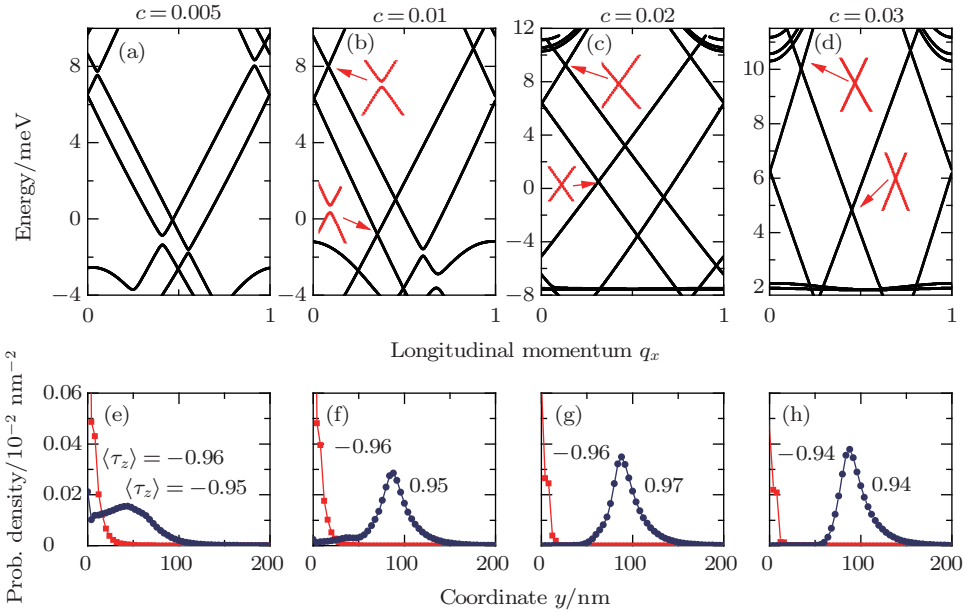


Fig. 12. Calculated eigenenergies for four different doping concentrations, i.e., (a) $c = 0.005$, (b) $c = 0.01$, (c) $c = 0.02$, (d) $c = 0.03$, as functions of momentum q_x (in units of $2\pi/L_x$). Panels (e)–(h) shown the probability density distributions of the edge states at $E_F = 4$ meV, which is normalized in $L_x \times L$ sample. The spin polarizations $\langle\hat{\tau}_z\rangle$ of the edge states are indicated in panels (e)–(h).[40]

At a given Fermi energy $E_F = 4$ meV, the typical spatial probability distributions of the edge states on an arbitrarily chosen cross section of the sample for different doping concentrations are plotted in Figs. 12(e)–12(h), in which only the profile on one side of the sample is shown. For $c = 0.005$, the spin-up and spin-down polarized edge states both locate at $y = 0$ and have a large spatial overlap. For $c = 0.01$, while the spin-down edge state remains near $y = 0$, the spin-up edge state moves away from $y = 0$ toward $y = l = 80$ nm. However, an appreciable overlap still exists between the two edge states, which is the origin of the finite energy gaps for the edge state spectra in Figs. 12(a) and 12(b). With increasing c to 0.02 and 0.03, the spin-up edge state is peaked at $y = l$, and there is no longer any overlap between the spin-up and spin-down edge states, as shown in Figs. 12(g) and 12(h). This accounts for the vanishing energy gaps of the edge states shown in Figs. 12(c) and 12(d). These numerical results are in good agreement with the physical picture established in the above discussion, and further support the feasibility of the present proposal for a robust QSH effect.

6. Topological phase transitions in 3D TIs

So far, we have focused on the 2D QSH systems. Recently, we took a step forward, and applied the spin-Chern number approach to the study of the topological properties of 3D TIs.[48] For a 3D TI with an exchange field of strength g, we considered a momentum component, e.g., k_z, as a parameter, and analytically calculated spin-Chern numbers $C_\pm(k_z)$ for the effective 2D system. The obtained phase diagram for $C_\pm(k_z)$ describes the systematic evolution of the bulk band topology. For small g, an unconventional topological phase transition was discovered, which controls the basic properties of the surface states. The $C_\pm(k_z)$ undergo a transition from ± 1 to 0 with increasing $|k_z|$ to a critical value k_z^C. Correspondingly, on a sample surface parallel to the z axis, helical surface states were found to exist in the region $|k_z| < k_z^C$, and disappear for $|k_z| \geq k_z^C$. At $|k_z| = k_z^C$, the bulk spin spectrum gap closes, but the energy gap remains open. When $|g|$ is greater than a critical value g_C, a transition of the 3D TI into a Weyl semi-metal occurs. Two Weyl points appear as the critical points separating a QH phase of the effective 2D system for

$|k_z| < k_z^{\rm W}$ from an ordinary insulator for $|k_z| > k_z^{\rm W}$, indicating that their appearance is topological rather than accidental. Chiral surface states existing in the region $|k_z| < k_z^{\rm W}$ give rise to the Fermi arcs.

6.1. Model and method

Let us start from the effective Hamiltonian proposed in Ref. [15], $H = A_2\tau_x(k_x\sigma_x + k_y\sigma_y) + M(\boldsymbol{k})\tau_z + A_1 k_z\tau_x\sigma_z + g\sigma_z$, which was used to describe the strong TI of Bi_2Se_3. Here, σ_m and τ_m ($m = x, y, z$) denote the Pauli matrices in spin and orbital spaces, and $M(\boldsymbol{k}) = M_0 - B_1 k_z^2 - B_2 k_\perp^2$ with $k_\perp^2 = k_x^2 + k_y^2$ is the mass term expanded to the second order. In the last term, we include an exchange field of strength g to study the TR-symmetry-broken effect on the topological properties of the TI. For convenience, the momentum is set to be dimensionless by properly choosing the units of parameters in the model, namely, M_0, A_1, A_2, B_1, and B_2.

Making a unitary transformation $\mathscr{H} = U^\dagger H U$ with

$$U = \frac{1}{2}(1+\tau_x) + \frac{1}{2}(1-\tau_x)\sigma_z,$$

we obtain

$$\mathscr{H} = A_2(k_x\sigma_x + k_y\sigma_y) + [M(\boldsymbol{k})\tau_z + A_1 k_z\tau_x + g]\sigma_z. \quad (22)$$

The eigenstates of Hamiltonian (22) can be easily solved by first diagonalizing the operator in the square bracket. The four eigenenergies are obtained as

$$E_\pm^{\rm v(c)}(\boldsymbol{k}) = -(+)\sqrt{A_2^2 k_\perp^2 + [g \pm \lambda(\boldsymbol{k})]^2}\,, \quad (23)$$

where $\lambda(\boldsymbol{k}) = \sqrt{M^2(\boldsymbol{k}) + A_1^2 k_z^2}$, and subscripts $\pm$ indicate two valence (conduction) bands with superscript v (c). The electron wavefunctions in the valence bands are given by

$$\varphi_\pm(\boldsymbol{k}) = \phi_\pm(\boldsymbol{k}) \otimes \chi_\pm(\boldsymbol{k})\,. \quad (24)$$

Here, $\phi_+(\boldsymbol{k}) = [\mathrm{sgn}(A_1 k_z)\cos\alpha_k, \sin\alpha_k]^{\rm T}$ and

$$\phi_-(\boldsymbol{k}) = [\mathrm{sgn}(A_1 k_z)\sin\alpha_k, -\cos\alpha_k]^{\rm T}$$

are wavefunctions in the τ_z space, and

$$\chi_\pm = \left(\frac{k_x - \mathrm{i}k_y}{k_\perp}\sin\theta_k^\pm, -\cos\theta_k^\pm\right)^{\rm T}$$

are wavefunctions in the σ_z space, with

$$2\alpha_k = \mathrm{ctg}^{-1}[M(\boldsymbol{k})/|A_1 k_z|]$$

and $2\theta_k^\pm = \mathrm{ctg}^{-1}[(g \pm \lambda(\boldsymbol{k}))/|A_2 k_\perp|]$.

The basic idea of our theoretical calculation is explained as follows. We consider one of the momentum components, e.g., k_z, as a parameter. For a given k_z, equation (22) is equivalent to a 2D system, for which spin-Chern numbers $C_\pm(k_z)$ can be defined. For a semi-infinite sample of the 3D TI with its surface parallel to the z axis, k_z remains to be a good quantum number. Correspondingly, nonzero $C_\pm(k_z)$ indicate that edge states with the given k_z must appear on the edge of the effective 2D system. The edge states at various k_z essentially form the surface states of the 3D sample. Therefore, the characteristics of the surface states can be determined from the calculation of the k_z-dependent spin-Chern numbers $C_\pm(k_z)$.

6.2. Phase diagram

The spin-Chern numbers for the effective 2D system are calculated in the standard way, which has been described in details in the previous sections. By studying a special case of $k_z = 0$, we find that the topological properties of Eq. (22) can be described by the spin-Chern numbers $C_\pm(k_z)$ associated with τ_z. Here, τ_z corresponds to $U\tau_z U^\dagger = \tau_z\sigma_z$ in the original Hamiltonian H, and so can be considered as a spin operator measuring the difference of spin polarization between the two orbits. The eigenstates of the projected spin operator $P\tau_z P$ need to be calculated first, where P is the projection operator into the valence bands. Since $P\tau_z P$ commutes with the momentum operator, its eigenstates can be obtained at each momentum $\boldsymbol{k}$ separately. The eigenvalues of $P\tau_z P$ are given by

$$\xi^\pm(\boldsymbol{k}) = \pm\sqrt{\cos^2 2\alpha_k + \cos^2(\theta_k^+ - \theta_k^-)\sin^2 2\alpha_k}. \quad (25)$$

The corresponding eigenfunctions are denoted by $\Psi^\pm(\boldsymbol{k})$, whose expressions are lengthy and will not be written out here. The spin-Chern numbers are just the Chern numbers of the two spin sectors formed by $\Psi^\pm(\boldsymbol{k})$, i.e.,

$$C_\pm(k_z) = \frac{\mathrm{i}}{2\pi}\int \mathrm{d}k_x\mathrm{d}k_y\hat{\boldsymbol{e}}_z\cdot[\nabla_2 \times \langle\Psi^\pm(\boldsymbol{k})|\nabla_2|\Psi^\pm(\boldsymbol{k})\rangle],$$

where ∇_2 is the 2D gradient operator acting on (k_x, k_y). By some algebra, $C_\pm(k_z)$ are derived to be

$$C_\pm(k_z) = \pm\frac{1}{2}\left[\mathrm{sgn}(B_2) + \mathrm{sgn}\Big(Q(k_z)\mathrm{sgn}[P(k_z)] \pm g\Big)\right], \quad (26)$$

with $P(k_z) = M_0 - B_1 k_z^2$ and $Q(k_z) = \sqrt{P^2(k_z) + A_1^2 k_z^2}$.

Equations (23), (25), and (26) are the main analytical results of this work. For $g = 0$, the spin-Chern numbers at $k_z = 0$ reduce to $C_\pm(0) = \pm\frac{1}{2}[\mathrm{sgn}(B_2) + \mathrm{sgn}(M_0)]$. So $C_\pm(0)$ are nonzero when B_2 and M_0 have the same sign, and vanish otherwise. The $C_\pm(0)$ play a role similar to the Z_2 index. Nonzero $C_\pm(0)$ ensure that surface states exist in the vicinity of $k_z = 0$ on a surface parallel to the z axis. Without loss of generality, we will focus on the parameter region of $B_2 > 0$ and $M_0 > 0$, in which Bi_2Se_3 belongs. We wish to emphasize here that when k_z is considered as a parameter, the effective 2D Hamiltonian (22) breaks the TR symmetry for any $k_z \neq 0$, even if $g = 0$, as its TR counterpart is at $-k_z$. Therefore, while the k_z-dependent spin-Chern numbers given by Eq. (26) remain valid at any k_z, a Z_2 index cannot be defined for any $k_z \neq 0$.

Equation (26) allows us to extract more information about the basic characteristics of the surface states.

A typical phase diagram for the spin-Chern numbers in k_z versus g plane, as determined by Eq. (26), is plotted in Fig. 13. For simplicity, A_2 is taken to be the unit of energy. For small $|g|$, $C_{\pm}(k_z) = \pm 1$ at small $|k_z|$, corresponding to a QSH phase of the effective 2D system, and drop to 0 with increasing $|k_z|$ to a critical value $k_z^{\rm C} \equiv \sqrt{M_0/B_1} = 0.2$, as indicated by the dotted lines. The system becomes an ordinary insulator for $|k_z| > k_z^{\rm C}$. When $|g|$ is greater than a critical value $g_{\rm C}$, the effective 2D system enters a QH phase with a nonzero total (first) Chern number $C(k_z) \equiv C_+(k_z) + C_-(k_z) = 1$ if $g > 0$, and -1 if $g < 0$. The boundary enclosing this phase is determined by equation $g = \pm Q(k_z)$, as indicated by the solid curves. The critical exchange field is given by $g_{\rm C} = \min[Q(k_z)]$, and $g_{\rm C} \simeq 0.1$ in Fig. 13.

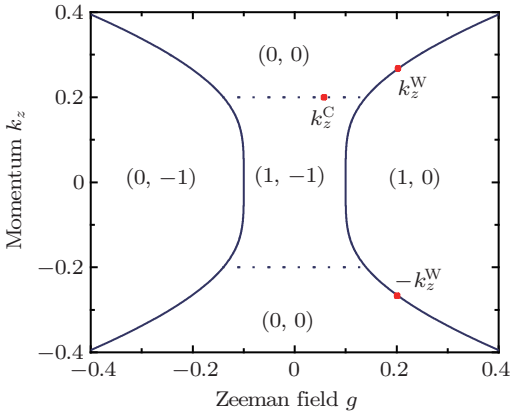


Fig. 13. Phase diagram determined from spin-Chern numbers in the k_z–g plane. The parameters are chosen to be $M_0 = 0.1$, $A_1 = 0.7$, $A_2 = 1$, $B_1 = 2.5$, and $B_2 = 14$.[48]

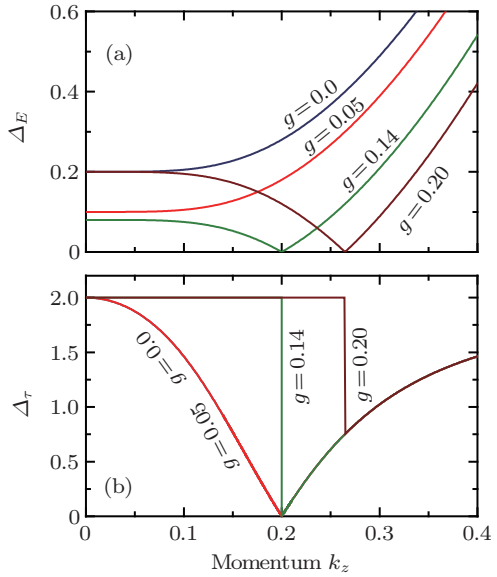


Fig. 14. (a) Energy gap Δ_E and (b) spin spectrum gap Δ_τ as functions of momentum k_z for different values of g. The other parameters are the same as those in Fig. 13.[48]

It is interesting to see how the energy gap

$$\Delta_E(k_z) = \min[E_{\pm}^c(\boldsymbol{k}) - E_{\pm}^v(\boldsymbol{k})]|_{k_z}$$

and the spin spectrum gap

$$\Delta_\tau(k_z) = \min[\xi^+(\boldsymbol{k}) - \xi^-(\boldsymbol{k})]|_{k_z}$$

behave on the boundary (solid and dashed lines) between different phases. From Eqs. (23) and (25), we find that on the dotted lines in Fig. 13, the spin spectrum gap closes at $k_x = k_y = 0$, but the energy gap remains open. On the contrary, on the solid boundary lines, the energy gap closes at $k_x = k_y = 0$, but the spin spectrum gap remains open. The Δ_E and Δ_τ as functions of k_z for several values of g are plotted in Figs. 14(a) and 14(b), respectively. We notice that at $g = 0.14$ and $k_z = k_z^{\rm C} = 0.2$, the energy gap and the spin spectrum gap vanish simultaneously. This is because the dotted and the solid boundary lines in Fig. 13 intersect at that point.

6.3. Properties of surface states

To study the surface states directly, we construct a tight binding model on a cubic lattice with two spins and two orbits on each site, which recovers the Hamiltonian of Eq. (22) in the continuum limit. A semi-infinite sample with its surface parallel to the y–z plane is considered, where k_y and k_z remain to be good quantum numbers.

The calculated energy spectrum for $g = 0.05$ is plotted as a function of k_y for four different values of k_z in Fig. 15. Although $g \neq 0$ breaks the TR symmetry, the surface states remain gapless at $k_z = 0$, because τ_z in Eq. (22) is conserved at $k_z = 0$. From Fig. 15, it is found that for $k_z = 0$, $k_z^{\rm C}/3$, and $2k_z^{\rm C}/3$, the surface states always exist in the bulk energy gap, but no surface states appear at $k_z = k_z^{\rm C}$. To see the evolution of the surface states with k_z more clearly, we define a maximum level spacing δE between the surface and the bulk states, as illustrated in Fig. 15. In Fig. 16(a), δE is plotted as a function of k_z. One can see that δE decreases with increasing k_z, and drops to nearly 0 at $k_z \geq k_z^{\rm C}$. Therefore, we conclude that the surface states exist only in the region $|k_z| < k_z^{\rm C}$, and vanish for $|k_z| \geq k_z^{\rm C}$, which is well consistent with the phase diagram in Fig. 13.

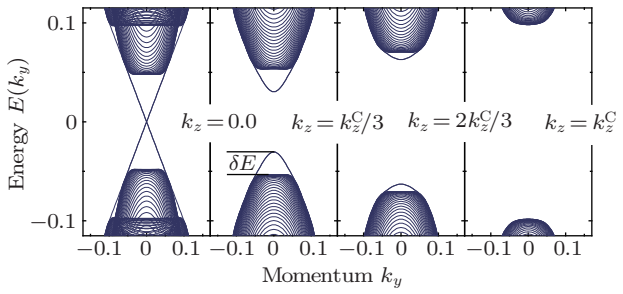


Fig. 15. Energy spectra for a semi-finite sample of the 3D TI for four different k_z at $g = 0.05$. The other parameters are the same as those in Fig. 13.[48]

The critical momentum $k_z^{\rm C}$ marks a topological phase transition characterized by the change in the spin-Chern num-

bers, with the disappearance of the surface states as an observable consequence. This topological phase transition is unconventional, which is accompanied by closing the spin spectrum gap rather than the energy gap. While the closing of the spin spectrum gap may not be observed directly, we find that the average of operator τ_z in H, namely,

$$\sum_{\beta=\pm} \langle \varphi_\beta(\boldsymbol{k})|\, U^\dagger \tau_z U |\varphi_\beta(\boldsymbol{k})\rangle \propto \cos(2\alpha_k),$$

changes its sign at $\boldsymbol{k} = (0,0,\pm k_z^{\rm C})$, which indicates a reversal of the orbital polarization, leading to a bulk physical observable at the transition. For the strong TI under consideration, the nontrivial bulk band topology can be examined in any direction. For example, by considering k_y as a parameter, we can calculate the spin-Chern numbers in the effective 2D space of (k_x, k_z), and obtain a critical $k_y^{\rm C} = \sqrt{(M_0+|g|)/B_2}$. Combining $k_y^{\rm C}$ and $k_z^{\rm C}$ together, we can depict the region in the k_y-k_z plane, where the topological surface states exist on a surface parallel to the y–z plane, as shown in Fig. 16(b).

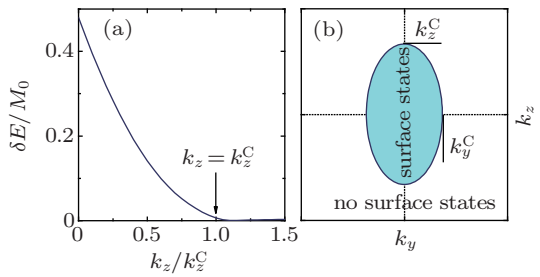


Fig. 16. (a) Maximum level spacing δE between the edge and the bulk states as a function of $k_z/k_z^{\rm C}$. (b) Schematic diagram of the distribution of surface states in 2D momentum space.[48]

For $|g| > g_{\rm C}$, the 3D system enters another topological phase characterized by a nonzero total Chern number $C(k_z)$ for small $|k_z|$, which is essentially a Weyl semi-metal phase.[49–52] The quantum phase transition from the 3D TI to the Weyl semi-metal with tunning g can be understood as a topological transition, at which the energy gap closing causes one of the spin-Chern numbers $C_\pm(k_z)$ to vanish, while the other remains to be quantized.

At a given g, from equation $g = \pm Q(k_z)$ for the solid boundary lines in Fig. 13, one can obtain two critical values of k_z, namely, $\pm k_z^{\rm W}$, as indicated in Fig. 13. The $\boldsymbol{k} = (0,0,\pm k_z^{\rm W})$ are a pair of Weyl points, or 3D Dirac points, at which the conduction and valence bands touch. The appearance of the Weyl points is usually attributed to an accidental degeneracy or symmetry.[49–52] The obtained phase diagram in Fig. 13 sheds more light on its topological origin. At a given g, the Weyl points appear as two critical points separating a QH state of the effective 2D system for $|k_z| < k_z^{\rm W}$ from an ordinary insulator state for $|k_z| > k_z^{\rm W}$, at which the energy gap must vanish. Changing parameters cannot open energy gaps at the two Weyl points, unless they come together so that the QH state in between is destroyed.

At $g = 0.2$, the calculated energy spectrum as a function of k_y for four different k_z is plotted in Fig. 17, for a semi-infinite sample with its surface parallel to the y–z plane. It is found that for $k_z < k_z^{\rm W}$, chiral surface states appear in the energy gap; the conduction and valence bands touch at $k_z = k_z^{\rm W}$; and for $k_z > k_z^{\rm W}$, the energy gap reopens, but the surface states no longer exist. These results are in good agreement with the phase diagram in Fig. 13. The chiral surface states appearing in the region of $|k_z| < k_z^{\rm W}$ give rise to the Fermi arcs.[49–52] The Weyl semi-metals with Fermi arcs have been paid much attention recently.

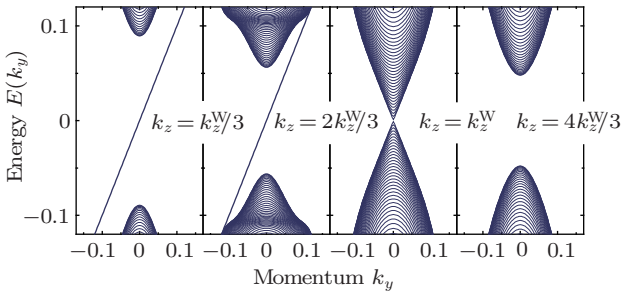


Fig. 17. Energy spectra of a semi-finite sample for four different k_z at $g = 0.2$. The other parameters are the same as those in Fig. 13.[48]

7. Conclusions and perspectives

In contradiction to the general belief that the QSH effect must be protected by the TR symmetry or other symmetries, the nontrivial topology of the bulk energy bands of the QSH systems characterized by nonzero spin-Chern numbers is found to remain intact when the TR symmetry is broken by small perturbations. Without considering any symmetries, it is shown by using a Laughlin-like gauge argument that nonzero spin-Chern numbers require edge states to appear at the sample boundary. The edge states may be gapped or gapless, depending on the local environment at the sample boundary. Creating a narrow FM region near the boundary of the QSH sample can cause the counterpropagating edge states to separate spatially, which provides an essential protection for the edge states against backscattering from disorders. This result points to a new direction for the search of a robust QSH effect in realistic environments. Some future research directions include: i) engineering the edge states of the QSH systems to achieve special device functions, ii) the effect of TR symmetry breaking in 3D TIs, and iii) the effect of electron interaction in QSH systems and thin films of 3D TIs.

References

[1] Kane C L and Mele E J 2005 *Phys. Rev. Lett.* **95** 226801
[2] Bernevig B A and Zhang S C 2006 *Phys. Rev. Lett.* **96** 106802
[3] König M, Wiedmann S, Brüne C, Roth A, Buhmann H, Molenkamp L W, Qi X L and Zhang S C 2007 *Science* **318** 766

[4] Bernevig B A, Hughes T L and Zhang S C 2006 *Science* **314** 1757
[5] Haldane F D M 1988 *Phys. Rev. Lett.* **61** 2015
[6] Sheng L, Sheng D N, Ting C S and Haldane F D M 2005 *Phys. Rev. Lett.* **95** 136602
[7] Wu C, Bernevig B A and Zhang S C 2006 *Phys. Rev. Lett.* **96** 106401
[8] Kane C L and Mele E J 2005 *Phys. Rev. Lett.* **95** 146802
[9] Sheng D N, Weng Z Y, Sheng L and Haldane F D M 2006 *Phys. Rev. Lett.* **97** 036808
[10] Prodan E 2009 *Phys. Rev. B* **80** 125327
[11] Li H C, Sheng L, Sheng D N and Xing D Y 2010 *Phys. Rev. B* **82** 165104
[12] Moore J E and Balents L 2007 *Phys. Rev. B* **75** 121306
[13] Fu L and Kane C L 2007 *Phys. Rev. B* **76** 045302
[14] Fu L, Kane C L and Mele E J 2007 *Phys. Rev. Lett.* **98** 106803
[15] Zhang H J, Liu C X, Qi X L, Dai X, Fang Z and Zhang S C 2009 *Nat. Phys.* **5** 438
[16] Hasan M Z and Kane C L 2010 *Rev. Mod. Phys.* **82** 3045
[17] Qi X L and Zhang S C 2010 *Phys. Today* **63** 33
[18] Fu L and Kane C L 2008 *Phys. Rev. Lett.* **100** 096407
[19] Qi X L, Hughes T and Zhang S C 2008 *Phys. Rev. B* **78** 195424
[20] Essin A M, Moore J E and Vanderbilt D 2009 *Phys. Rev. Lett.* **102** 146805
[21] Hsieh D, Qian D, Wray L, Xia Y, Hor Y S, Cava R J and Hasan M Z 2008 *Nature* **452** 970
[22] Hsieh D, Xia Y, Wray L, Qian D, Pal A, Dil J H, Osterwalder J, Meier F, Bihlmayer G, Kane C L, Hor Y S, Cava R J and Hasan M Z 2009 *Science* **323** 919
[23] Xia Y, Qian D, Hsieh D, Wray L, Pal A, Lin H, Bansil A, Grauer D, Hor Y S, Cava R J and Hasan M Z 2009 *Nat. Phys.* **5** 398
[24] Chen Y L, Analytis J G, Chu J H, Liu Z K, Mo S K, Qi X L, Zhang H J, Lu D H, Dai X, Fang Z, Zhang S C, Fisher I R, Hussain Z and Shen Z X 2009 *Science* **325** 178
[25] Roushan P, Seo J, Parker C V, Hor Y S, Hsieh D, Qian D, Richardella A, Hasan M Z, Cava R J and Yazdani A 2009 *Nature* **460** 1106
[26] Zhang T, Cheng P, Chen X, Jia J F, Ma X C, He K, Wang L L, Zhang H J, Dai X, Fang Z, Xie X C and Xue Q K 2009 *Phys. Rev. Lett.* **103** 266803
[27] Jiang H, Cheng S, Sun Q F and Xie X C 2009 *Phys. Rev. Lett.* **103** 036803
[28] Laughlin R B 1981 *Phys. Rev. B* **23** 5632
[29] Fu L and Kane C L 2006 *Phys. Rev. B* **74** 195312
[30] Yang Y, Xu Z, Sheng L, Wang B G, Xing D Y and Sheng D N 2011 *Phys. Rev. Lett.* **107** 066602
[31] Qi X L, Wu Y S and Zhang S C 2006 *Phys. Rev. B* **74** 085308
[32] Qi X L, Hughes T L and Zhang S C 2008 *Phys. Rev. B* **78** 195424
[33] Liu C X, Qi X L, Dai X, Fang Z and Zhang S C 2008 *Phys. Rev. Lett.* **101** 146802
[34] Yu R, Zhang W, Zhang H J, Zhang S C, Dai X and Fang Z 2010 *Science* **329** 61
[35] Qiao Z H, Yang S Y A, Feng W X, Tse W K, Ding J, Yao Y G, Wang J and Niu Q 2010 *Phys. Rev. B* **82** 161414
[36] Chang C Z, *et al*. 2013 *Science* **399** 1582
[37] Xu Z, Sheng L, Xing D Y, Prodan E and Sheng D N 2012 *Phys. Rev. B* **85** 075115
[38] Xu Z, Sheng L, Shen R, Wang B and Xing D Y 2013 *J. Phys.: Condens. Matter* **25** 065501
[39] Li H C, Sheng L and Xing D Y 2012 *Phys. Rev. Lett.* **108** 196806
[40] Li H C, Sheng L, Shen R, Shao L B, Wang B G, Sheng D N and Xing D Y 2013 cond-mat/1301.6460
[41] Shan W Y, Lu H Z and Shen S Q 2010 *New J. Phys.* **12** 043048
[42] Lu H Z, Shan W Y, Yao W, Niu Q and Shen S Q 2010 *Phys. Rev. B* **81** 115407
[43] Brüne C, Liu C X, Novik E G, Hankiewicz E M, Buhmann H, Chen Y L, Qi X L, Shen Z X, Zhang S C and Molenkamp L W 2011 *Phys. Rev. Lett.* **106** 126803
[44] Sheng L, Xing D Y, Sheng D N and Ting C S 1997 *Phys. Rev. Lett.* **79** 1710
[45] Liu Q, Liu C X, Xu C K, Qi X L and Zhang S C 2009 *Phys. Rev. Lett.* **102** 156603
[46] König M, Buhmann H, Molenkamp L W, Hughes T, Liu C X, Qi X L and Zhang S C 2008 *J. Phys. Soc. Jpn.* **77** 031007
[47] Smith D L and Mailhiot C 1990 *Rev. Mod. Phys.* **62** 173
[48] Yang Y, Li H, Sheng L, Shen R, Sheng D N and Xing D Y 2013 cond-mat/1301.1618
[49] Wan X, Turner A M, Vishwanath A and Savrasov S Y 2011 *Phys. Rev. B* **83** 205101
[50] Burkov A A and Balents L 2011 *Phys. Rev. Lett.* **107** 127205
[51] Sun K, Liu W V, Hemmerich A and Sarma S D 2011 *Nat. Phys.* **8** 67
[52] Fang C, Gilbert M J, Dai X and Bernevig B A 2012 *Phys. Rev. Lett.* **108** 266802