# Finite Size Effects on Helical Edge States in a Quantum Spin-Hall System

Bin Zhou,[1,2] Hai-Zhou Lu,[1] Rui-Lin Chu,[1] Shun-Qing Shen,[1] and Qian Niu[3]

[1]*Department of Physics, and Center for Theoretical and Computational Physics, The University of Hong Kong, Pokfulam Road, Hong Kong, China*

[2]*Department of Physics, Hubei University, Wuhan 430062, China*

[3]*Department of Physics, The University of Texas at Austin, Austin, Texas 78712, USA*

(Received 9 May 2008; published 10 December 2008)

The hallmark of the spin-Hall insulator is the presence of gapless edge states of different spins moving in opposite directions. Through analytical solutions in a model calculation for a strip of finite width, we find that edge states on the two sides can couple together to produce a gap in the spectrum, destroying the quantum spin-Hall effect. The application of a magnetic field can however modify and even remove the gap by shifting the momenta of the edge states relative to each other.

DOI: 10.1103/PhysRevLett.101.246807 PACS numbers: 73.43.−f, 72.25.Dc, 85.75.−d

Recent discovery of quantum spin-Hall (QSH) effect brings into the family of Hall systems a new member, a band insulator with topological properties of electrons distinct from the conventional ones [1]. First proposed for a model graphene system by Kane and Mele [2], the QSH system has an energy gap in the bulk, but has gapless helical edge states with different spins moving in opposite directions. A more realistic model was proposed by Benevig *et al.* [3] based on semiconductor HgTe/CdTe quantum wells, which was soon confirmed by experiment with the observation of ballistic edge channels [4]. Several other candidates of QSH system were also proposed, such as GaAs under shear strain [5] and a multilayer Bi thin film [6]. Intensive theoretical studies are being made to explore the exotic properties of the QSH effect and other topological insulators [7]. Very recently, it was reported that massive Dirac particles exist in the bulk of $Bi_{0.9}Sb_{0.1}$, which is a hallmark of higher dimensional QSH insulator [8].

The theory of QSH effect in HgTe/CdTe quantum well was based on an effective 4-band model that depicts the inversion crossing of electron and hole band [3]. Until the present, almost all works are based on numerical solutions of the model in a tight-binding method [2,9,10]. In this Letter, we present an analytical study of the effective 4-band model for the HgTe/CdTe quantum well, in particular, for the case of a finite strip geometry which is used in the experiment. We show that the edge states on the two sides can couple together to generate a gap in the spectrum even in the clean limit, breaking the edge channels. This is a striking difference from the quantum Hall edge states which do not couple across the width of the strip in the absence of scattering. However, the application of a magnetic field can shift or even remove the gap by changing the relative momenta of the edge channels with the gauge potential.

The effective $4 \times 4$ model for QSH effect was derived from the Kane model for semiconductors confining in a heterojunction of semiconductor Hg/CdTe [3]

$$\mathcal{H}(k_x, k_y) = \begin{pmatrix} H(k) & 0 \\ 0 & H^*(-k) \end{pmatrix}. \tag{1}$$

$H(k) = \epsilon_k \mathbf{I}_2 + d^a(k)\sigma^a$, with $\mathbf{I}_2$ being a $2 \times 2$ unit matrix, and $\sigma_a$ the Pauli matrices. For a small $k_x$ expansion, $\epsilon_k = C - D(k_x^2 + k_y^2)$, $d^1 = Ak_x$, $d^2 = Ak_y$, and $d^3 = \mathcal{M}(k) = M - B(k_x^2 + k_y^2)$. $A$, $B$, $C$, $D$, and $M$ are expansion parameters that are determined by the thickness of the quantum well and the material parameters. The upper block $H(k)$ is the for spin up ($m_s = 1/2, 3/2$), and the lower block is for the spin down ($m_s = -1/2, -3/2$). The most striking property of the model is that the mass or gap parameter $M$ changes its sign when the thickness $d$ of the quantum well is varied through a critical thickness $d_c$ ($= 6.3$ nm) associating with the transition of electronic band structure from a normal to an ''inverted'' type [11].

We will now solve this model (1) in a finite strip geometry of the width $L$ with the periodic boundary condition in the $x$ direction and an open boundary condition in the $y$ direction. In this case, $k_x$ is a good quantum number, but $k_y$ is replaced by using the Peierls substitution. $k_y = -i\partial/\partial y = -i\partial_y$. The Hamiltonian (1) is block diagonal, and the eigenvalue problem of the upper and lower blocks can be solved separately, i.e., $\hat{H}_\uparrow \Psi_\uparrow = E\Psi_\uparrow$ and $\hat{H}_\downarrow \Psi_\downarrow = E\Psi_\downarrow$. Because the lower block $H^*(-k)$ is the time reversal of the upper block $H(k)$, the solution $\Psi_\downarrow(k_x, y) = \Theta\Psi_\uparrow(k_x, y)$, where $\Theta = -i\sigma_y K$ is a ''time-reversal'' operator and $K$ stands for complex conjugation. Thus, we can only focus on the solution for the upper block of this Hamiltonian.

To have solutions for the energy spectrum and wave functions, we solve the Schrödinger equations for the upper block,

$$[M - B_+(k_x^2 - \partial_y^2)]\psi_1 + A(k_x - \partial_y)\psi_2 = E\psi_1, \tag{2}$$

$$A(k_x + \partial_y)\psi_1 - [M - B_-(k_x^2 - \partial_y^2)]\psi_2 = E\psi_2, \tag{3}$$

with $B_\pm = B \pm D$. From Eqs. (2) and (3), using the trial

 

function $\psi_{1,2} = e^{\lambda y}$, the secular equation gives four roots $\pm\lambda_1$ and $\pm\lambda_2$,

$$\lambda_{1,2}^2 = k_x^2 + F \pm \sqrt{F^2 - (M^2 - E^2)/B_+B_-}, \quad (4)$$

where $F = \frac{A^2 - 2(MB+ED)}{2B_+B_-}$. With the open boundary conditions of $\Psi_\uparrow(k_x, y = \pm L/2) = 0$, we have an analytical expression for the wave function $\Psi_\uparrow$

$$\psi_1 = \tilde{c}_+ f_+(k_x, y) + \tilde{c}_- f_-(k_x, y), \quad (5)$$

$$\psi_2 = \tilde{d}_+ f_+(k_x, y) + \tilde{d}_- f_-(k_x, y), \quad (6)$$

with

$$f_+(k_x, y) = \frac{\cosh(\lambda_1 y)}{\cosh(\lambda_1 L/2)} - \frac{\cosh(\lambda_2 y)}{\cosh(\lambda_2 L/2)}, \quad (7)$$

$$f_-(k_x, y) = \frac{\sinh(\lambda_1 y)}{\sinh(\lambda_1 L/2)} - \frac{\sinh(\lambda_2 y)}{\sinh(\lambda_2 L/2)}. \quad (8)$$

The nontrivial solution for the coefficients $\tilde{c}_\pm$ and $\tilde{d}_\pm$ in the wave functions leads to a secular equation

$$\frac{\tanh\frac{\lambda_1 L}{2}}{\tanh\frac{\lambda_2 L}{2}} + \frac{\tanh\frac{\lambda_2 L}{2}}{\tanh\frac{\lambda_1 L}{2}} = \frac{\alpha_1^2\lambda_2^2 + \alpha_2^2\lambda_1^2 - k_x^2(\alpha_1 - \alpha_2)^2}{\alpha_1\alpha_2\lambda_1\lambda_2}, \quad (9)$$

where $\alpha_{1,2}(E) = E - M + B_+k_x^2 - B_+\lambda_{1,2}^2$. Equations (4) and (9) give the energy dispersions for electrons, and the values of two characteristic quantities $\lambda_{1,2}$.

General properties of the solution for $\lambda_{1,2}$ determines the distribution of the wave functions in space. In a large $L$ limit, a purely imaginary $\lambda = ik_y$ is always the solutions of the equations, and gives two branches of spectra, $E_\pm = \epsilon_k \pm \sqrt{(M - Bk^2)^2 + A^2k^2}$. The corresponding solutions for wave function in Eqs. (5) and (6) are expressed in term of $\sin k_y y$ and $\cos k_y y$ and span in the whole space; i.e., the solutions are for the bulk states. Except for these solutions of imaginary $\lambda$, there also exist solutions of real $\lambda$s when

$$A^2/B_+B_- > 4M/B > 0. \quad (10)$$

In the large $L$ limit, we have

$$E_\pm = M - B_+\lambda_1\lambda_2 \pm B_+(\lambda_1 + \lambda_2)k_x - B_+k_x^2. \quad (11)$$

Near $k_x = 0$, we have two linear dispersions, $E_\pm(k_x) = -MD/B \pm A\sqrt{B_+B_-/B^2}k_x + O(k_x^2)$. For real roots $\lambda_{1,2}$, the function $f_\pm(k_x, y)$ are distributed dominantly near the edge ($y = \pm L/2$) in the scale of $\lambda_{1,2}^{-1}$. That's why these states are called the edge state. The linear dispersions of the edge states are characteristic of QSH effect. Our result is consistent with those by means of the tight-binding approximation [9].

One of the key features for the solution of a finite width $L$ is the gap opening for the energy dispersion of the edge state. For real $\lambda$ and finite $L$, the right hand side of Eq. (9) is always greater than 2. If $\lambda_{1,2}L \gg 1$, it is approximately $2 + 4e^{-2\lambda_2 L}$ (assuming $\lambda_1 \gg \lambda_2$). From Eq. (9), it is found that a finite energy gap $\Delta = E_+ - E_-$ opens at $k_x = 0$

$$\Delta \simeq \frac{4|AB_+B_-M|}{\sqrt{B^3(A^2B - 4B_+B_-M)}}e^{-\lambda_2 L}, \quad (12)$$

which decays in an exponential law of $L$. This is the main consequence in the present work. It is different from behaviors of the energy level discretization of the bulk states where the wave functions are confined in a finite space and energy separation decays in a power law for a large $L$.

In general cases, the energy dispersion and energy gap for a finite width $L$ are calculated numerically as shown in Fig. 1. As a concrete example, we adopt the model parameters for the "inverted" quantum well from Ref. [10] for all numerical calculations in the present Letter as listed in the caption of Fig. 1. For $L = 1000$ nm, the energy gap is very tiny, $\Delta = 1.41 \times 10^{-7}$ meV. However, for $L = 200$ nm, the gap $\Delta$ is 0.4509 meV, which becomes large enough to be measurable in experiments. We plot the energy spectra for the edge states of several sizes in Fig. 1(a). For a narrow width $L$ (e.g., $L = 200$ nm), the spectra are parabolic-like since an energy gap opens near $k_x = 0$. The size-dependence of the energy gap is plotted in Fig. 1(b).

The expressions for the wave functions of the edge states are

$$\Psi_{\uparrow+} = \tilde{c}_+ e^{ik_x x}(f_+ + \gamma_{k_x}^+ f_-, \eta_1^+ f_- + \gamma_{k_x}^+\eta_2^+ f_+)^T, \quad (13)$$

$$\Psi_{\uparrow-} = \tilde{c}_- e^{ik_x x}(f_- + \gamma_{k_x}^- f_+, \eta_2^- f_+ + \gamma_{k_x}^-\eta_1^- f_-)^T, \quad (14)$$

where

$$\eta_1^\pm = \frac{B_+(\lambda_1^2 - \lambda_2^2)/A}{\lambda_1 \coth\frac{\lambda_1 L}{2} - \lambda_2 \coth\frac{\lambda_2 L}{2}},$$

$$\eta_2^\pm = \frac{B_+(\lambda_1^2 - \lambda_2^2)/A}{\lambda_1 \tanh\frac{\lambda_1 L}{2} - \lambda_2 \tanh\frac{\lambda_2 L}{2}},$$

$$\gamma_{k_x}^\pm = \frac{B_+(\lambda_1^2 - \lambda_2^2)k_x(\eta_1/\eta_2)^\pm}{\beta\lambda_1 \tanh^\pm\frac{\lambda_1 L}{2} - \alpha\lambda_2 \tanh^\pm\frac{\lambda_2 L}{2}},$$

for $E = E_\pm$ and $\tilde{c}_\pm$ are normalization constants. The solutions can be simplified in a large $L$ limit as $\eta_1^\pm = \eta_2^\pm = \eta = B_+(\lambda_1 + \lambda_2)/A$ and $\gamma_{k_x}^+ = -\gamma_{k_x}^- = -\mathrm{sgn}(k_x)$. Other two solutions for the lower block can be produced by means of the time-reversal operation, $\Psi_{\downarrow\pm} = \Theta\Psi_{\uparrow\pm}$, and the spectra are degenerate with those of the upper block as a result of time-reversal invariance.

The density distribution at the two edges are mainly determined by the larger one of the length scales $\lambda_{1,2}^{-1}$, according to the present analytic solutions of wave functions. In the example in Fig. 1, we notice that $\lambda_2^{-1} \gg \lambda_1^{-1}$. For a larger size of the sample, the density of the wave function increases in an exponential law in the shorter scale of $\lambda_1^{-1}$ and then decays exponentially in a longer scale of $\lambda_2^{-1}$ near the edge, which is consistent with the work by König *et al.* [10]. The wave function almost vanishes far

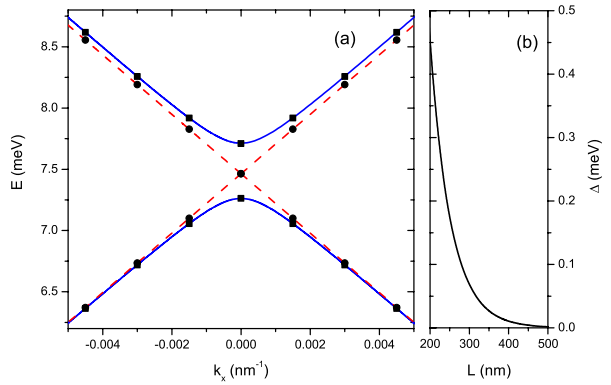


FIG. 1 (color online). (a) Energy spectra of edge states for $L = 200$ nm (solid lines) and $L = 1000$ nm (dashed lines). The parameters are adopted from Ref. [10], $A = 364.5$ meV nm, $B = -686$ meV nm$^2$, $M = -10$ meV, $D = -512$ meV nm$^2$. As a comparison, numerical results of tight-binding approximation are also plotted as black squares for $L = 200$ nm and black dots for $L = 1000$ nm. (b) The width dependence of the energy gap $\Delta$ of the edge states.

away from the edges if the width of the sample is much larger than the longer scale $\lambda_2^{-1}$. As an example, the density distributions of $\Psi_{\uparrow\pm}(k_x, y)$ for $L = 200$ nm are plotted for demonstration in Fig. 2 where $\lambda_2^{-1}$ at $k_x = 0$ are 55.9 and 51.8 nm, respectively. The states of $\Psi_{\uparrow+}(k_x, y)$ and $\Psi_{\uparrow-}(-k_x, y)$ $(k_x > 0)$ have the same spin $[\propto (1, -\eta)^T$ in the large $L$ limit] and the positive velocity, $v_x > 0$ when $k_x$ deviates away from $k_x = 0$ and the density distribution is located at one side while the states of $\Psi_{\uparrow+}(-k_x, y)$ and $\Psi_{\uparrow-}(+k_x, y)$ $(k_x > 0)$ have spin $[\propto (1, \eta)^T$ in the large $L$ limit] and a negative velocity, $v_x < 0$, and are distributed on the other side. However, near $k_x = 0$, from the solution we found that $\Psi_{\uparrow\pm}(k_x, y)$ and $\Psi_{\uparrow\pm}(-k_x, y)$ couple together due to the finite-size effect. Consequently, the densities of the wave functions $\Psi_{\uparrow\pm}(k_x = 0, y)$ are symmetrically distributed at the two sides. This fact is consistent with the opening of an energy gap in the spectra at $k_x = 0$.

As a finite-size effect, the charge conductance of edge states in a strip geometry will be modified. The charge conductance for a QSH phase was predicted theoretically to be $2e^2/h$ because of the presence of two 1D spin-resolved conducting channels at the edges of the strip, which was measured experimentally [4]. When the energy gap of edge states opens, the conductance will be modified. Following the Landauer-Büttiker formula [12], the charge conductance has the form,

$$G(\Delta) = \frac{2e^2}{h}\left[\frac{1}{e^{[(\Delta/2)-\mu]/k_BT}+1} - \frac{1}{e^{[-(\Delta/2)-\mu]/k_BT}+1} + 1\right].$$

At low temperatures, $G(\Delta) \to 2e^2/h$ only when the Fermi level locates out of the gap $\Delta(L)$. Below the characteristic temperature of $k_BT^* = \Delta$, a dip will be obviously exhibited in the curve of conductance via the gate voltage. We plot the temperature dependence of the conductance in Fig. 3. In the experiment by König *et al.* [4], the smallest size of the sample is $L = 0.5$ $\mu$m, and the measurement was performed at temperature of 30 mK. The calculated energy gap is $1.6 \times 10^{-3}$ meV (~19 mK), which is already comparable with the experiment temperature. For a smaller sample, the energy gap $\Delta$ of $L = 0.2$ $\mu$m is about 0.451 meV, and the temperature $T^*$ is enhanced to be 5.22 K. This effect should be measured according to the current technique. By the way, it is also worth noting that the gap is also highly sensitive to the thickness of quantum well, as all parameters in Eq. (1) are functions of the thickness.

Our solutions show that the QSH edge states are quite different from the edge states of a conventional QH strip. For a QH strip with translational symmetry along the strip, the states are classified with the momentum $k_x$. The edge states at the two sides have different $k_x$ and do not mix

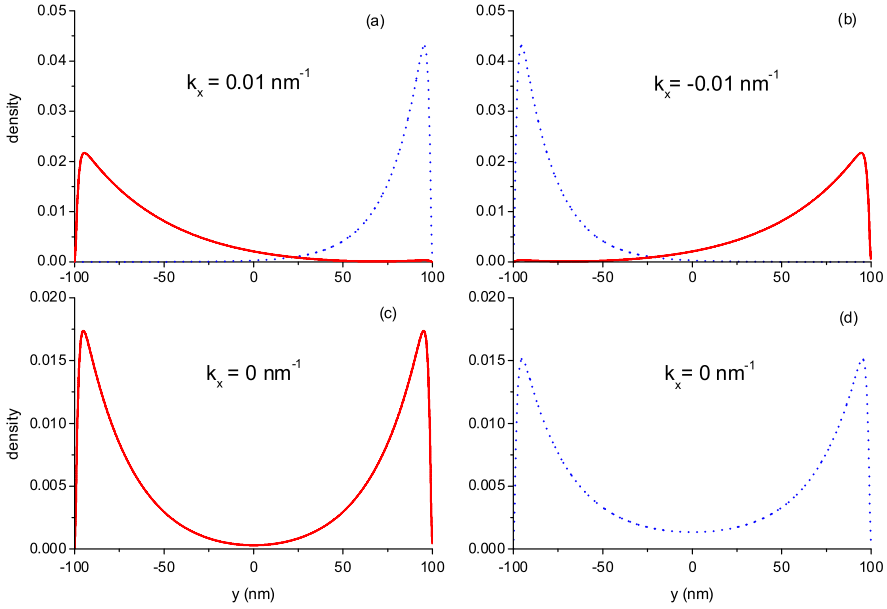


FIG. 2 (color online). The density distribution of the two edge states $\Psi_{\uparrow\pm}(k_x, y)$ for $L = 200$ nm. (a) The solid line corresponds to $|\Psi_{\uparrow+}(k_x, y)|^2$, and the dotted line to $|\Psi_{\uparrow+}(-k_x, y)|^2$ at $k_x = 0.01$ nm$^{-1}$; (b) The solid line corresponds to $|\Psi_{\uparrow-}(k_x, y)|^2$, and the dotted line to $|\Psi_{\uparrow-}(-k_x, y)|^2$ at $k_x = -0.01$ nm$^{-1}$; (c) and (d) for $k_x = 0$ nm$^{-1}$.

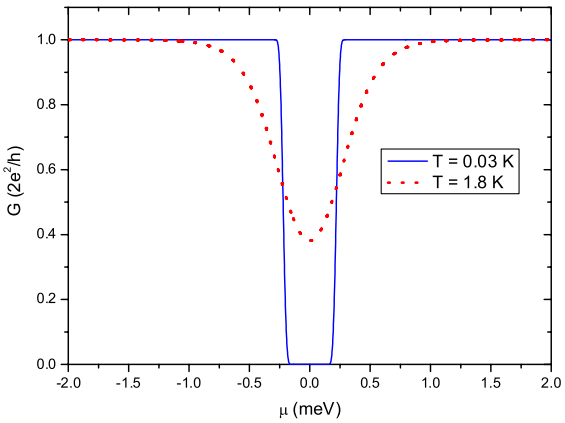


FIG. 3 (color online). The variations of conductance $G$ via the chemical potential $\mu$ inside the bulk insulting gap for $L = 200$ nm and at temperatures 30 mK and 1.8 K, respectively. Note that the energy zero point is shifted to the center of the gap $\Delta$ (=0.4509 meV).

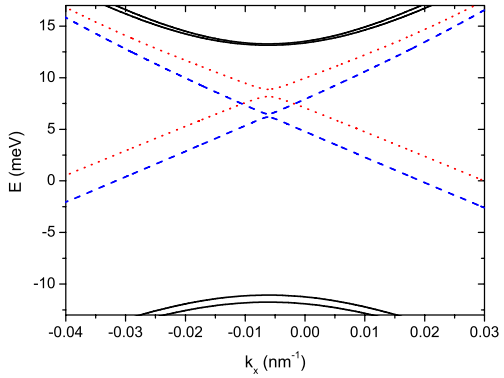


FIG. 4 (color online). The energy dispersion of the edge states in a weak field of $B = 0.04$ T. The dashed lines are for the upper block, and dotted lines are for the lower block. The solid lines are for the bulk spectra. The width of sample is $L = 200$ nm.

together even when the two states overlap in space [13]. In the case of a QSH strip, the two states have the nearly equal energy and momentum near the anticrossing points. So they can couple together to open an energy gap when their energies becomes closer and the wave functions have overlaps in a finite space. The magnetic field dependence of the QSH edge states also reflects this peculiar property. Consider the sample is subjected to a weak perpendicular magnetic field $\mathbf{B}_z$. Using the Peierls substitution, $k_x \rightarrow k_x - eA_x/\hbar$ in Eq. (1) by taking the gauge, $A_x = -B_z y$ (for $|y| < L/2$) and $A_y = 0$ in order to keep $k_x$ as a good quantum number. As the wave functions of the edge states decay exponentially, the expectation values of $A_x$ in the two edge states is proportional to $L$ approximately for $L \gg \lambda_{1,2}^{-1}$. The energy will be shifted by $\Delta E = +g\mu_B B_z$ $\{g \approx m_e v_x[L - \lambda_1^{-1} - \lambda_2^{-2} - 2(\lambda_1 + \lambda_2)^{-1}]/\hbar\}$ near $k_x = 0$. Thus, the energy spectra of the two edge states of the upper block will shift downward or upward $E(k_x) \approx \pm v_x \hbar k_x + g\mu_B B_z$ for a large $L$. By increasing the magnetic field $B_z$ the anticrossing point of energy spectra is eventually moved out of the bulk insulating gap, and the spectra will not cross in momentum between the gap. However, other edge states from the lower block have opposite spin, and their spectra will move towards an opposite direction. The two sets of the spectra may cross in momentum inside the insulating gap. Numerical results of detailed calculations are plotted in Fig. 4. The energy shift is very sensitive to the magnetic field and the width of the sample. As for the conductance, the value of $2e^2/h$ will recover near the crossing points at low temperatures. Thus the magnetoresistance is very sensitive to a tiny field, which might have potential applications for a sensitive detection.

Finally, we point out that other factors such as disorder or interaction will further affect the QSH effect in a finite-size system. In this case, since the edge states at two sides are not well separated, electron scattering between the edge states at the two side becomes possible if the weak disorder or interaction scattering are taken into account. As a result, the single particle elastic backscattering of the edge states by disorder will no longer be forbidden, and the edge states will not be protected completely by the time-reversal symmetry as in a large $L$ limit [2]. On the other hand, it is known that the disorder will drive electrons to be localized in two dimension. The bulk electron localization will also affect the stability of the edge states if the disorder is strong enough.

S. Q. S. thanks S. C. Zhang for helpful discussions. This work was supported by the Research Grant Council of Hong Kong under Grant No. HKU 7042/06P and the CRCG of the University of Hong Kong.

[1] C. Day, Phys. Today **61**, No. 1, 19 (2008).
[2] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005); **95**, 226801 (2005).
[3] B. A. Bernevig *et al.*, Science **314**, 1757 (2006).
[4] M. König *et al.*, Science **318**, 766 (2007).
[5] B. A. Bernevig and S. C. Zhang, Phys. Rev. Lett. **96**, 106802 (2006).
[6] S. Murakami, Phys. Rev. Lett. **97**, 236805 (2006).
[7] D. N. Sheng *et al.*, Phys. Rev. Lett. **97**, 036808 (2006); C. Wu, B. A. Bernevig, and S. C. Zhang, *ibid.* **96**, 106401 (2006); M. Onoda *et al.*, *ibid.* **98**, 076802 (2007); Z. H. Qiao *et al.*, *ibid.* **101**, 016804 (2008); C. Xu and J. E. Moore, Phys. Rev. B **73**, 045322 (2006); L. Fu and C. L. Kane, *ibid.* **74**, 195312 (2006); T. Fukui and Y. Hatsugai, *ibid.* **75**, 121403 (2007); H. Obuse *et al.*, *ibid.* **76**, 075301 (2007); S. Murakami *et al.*, *ibid.* **76**, 205304 (2007); X. Dai *et al.*, *ibid.* **77**, 125319 (2008).
[8] D. Hsieh *et al.*, Nature (London) **452**, 970 (2008).
[9] X. L. Qi *et al.*, Phys. Rev. B **74**, 085308 (2006).
[10] M. König *et al.*, J. Phys. Soc. Jpn. **77**, 031007 (2008).
[11] A. Pfeuffer-Jeschke, Ph.D. thesis, University of Würzburg, 2000.
[12] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, 1995).
[13] B. I. Halperin, Phys. Rev. B **25**, 2185 (1982); A. H. MacDonald and P. Steda, Phys. Rev. B **29**, 1616 (1984).