Hexagonal Warping Effects in the Surface States of the Topological Insulator Bi$_2$Te$_3$

Liang Fu

Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA

(Received 12 August 2009; revised manuscript received 21 October 2009; published 21 December 2009)

A single two-dimensional Dirac fermion state has been recently observed on the surface of the topological insulator Bi$_2$Te$_3$ by angle-resolved photoemission spectroscopy. We study the surface band structure using $k \cdot p$ theory and find an unconventional hexagonal warping term, which is the counterpart of cubic Dresselhaus spin-orbit coupling in rhombohedral structures. We show that this hexagonal warping term naturally explains the observed hexagonal snowflake Fermi surface. The strength of hexagonal warping is characterized by a single parameter, which is extracted from the size of the Fermi surface. We predict a number of testable signatures of hexagonal warping in spectroscopy experiments on Bi$_2$Te$_3$. We also explore the possibility of a spin-density wave due to strong nesting of the Fermi surface.

DOI: 10.1103/PhysRevLett.103.266801

PACS numbers: 73.20.–r, 73.43.Cd, 75.10.–b

Recently a new state of matter called a topological insulator has been observed in a number of materials [1–5]. A topological insulator has a time-reversal-invariant band structure with nontrivial topological order, which gives rise to gapless surface states bound to the sample boundary [6–8]. The two-dimensional surface band has a unique Fermi surface that encloses an odd number of Dirac points in the surface Brillouin zone [6], which is prohibited in conventional materials by fermion doubling theorem [9]. Soon after the theoretical prediction [10], the semiconducting alloy Bi$_x$Sb$_{1-x}$ was found to be a topological insulator having a Dirac surface band as well as other electron and hole pockets [1]. Subsequently, a family of materials Bi$_2$X$_3$ (X = Se and Te) was found to be topological insulators with a single Dirac-fermion surface state [3–5]. The observation of an undoubled Dirac fermion is not only of great conceptual interest but also paves the way for studying unusual electromagnetic properties [10,11] and realizing topological quantum computation [12]. Therefore surface states of Bi$_2$X$_3$ are being intensively studied in transport and spectroscopy experiments [13,14].

In this work, we study the electronic properties of surface states in Bi$_2$Te$_3$ using $k \cdot p$ theory. Our motivation is to understand the shape of Fermi surface observed in recent angle-resolved photoemission spectroscopy (ARPES) experiments [4,5], reproduced in Fig. 1. By considering the crystal symmetry of Bi$_2$Te$_3$, we find an unconventional hexagonal warping term in the surface band structure, which is the counterpart of cubic Dresselhaus spin-orbit coupling in rhombohedral structures. This hexagonal warping term naturally explains the snowflake shape of the Fermi surface, and its magnitude is extracted from the size of the Fermi surface. We predict that hexagonal warping of the Fermi surface should have important effects in several spectroscopy experiments. Finally, we observe that the Fermi surface of Bi$_2$Te$_3$ is nearly a hexagon with strong nesting for an appropriate range of surface charge density. This motivates us to explore theoretically a possible spin-density wave (SDW) phase. We discuss various types of SDW order in a Landau-Ginzburg theory.

Bi$_2$Te$_3$ has a rhombohedral crystal structure with space group R$3\bar{m}$. In the presence of a [111] surface, the symmetry of the crystal is reduced to $C_{3v}$, which consists of a threefold rotation $C_3$ around the trigonal $z$ axis and a mirror operation $M : x \rightarrow -x$ where $x$ is in $\Gamma K$ direction. Two surface bands are observed to touch at the origin of the surface Brillouin zone $\Gamma$. The degeneracy is protected by time-reversal symmetry and the doublet $|\psi_{1\downarrow}\rangle$ forms a Kramers pair. We choose a natural basis for the doublet according to total angular momentum $J = L + S = \pm 1/2$ so that $C_3$ is represented as $e^{-i\sigma_z \pi/3}$. Since $M^2 = -1$ for spin $1/2$ electron and $MC_3M^{-1} = C_3^{-1}$, the mirror operation can be represented as $M = i\sigma_z$ by defining the phase of $|\psi_{1\downarrow}\rangle$ appropriately. The antiunitary time-reversal operation $\Theta$ is represented by $i\sigma_z K$ ($K$ is complex conjugation) and commutes with both $M$ and $C_3$. Here the pseudospin $\sigma_z$ is proportional to the electron’s spin: $\langle \sigma_z \rangle \propto \langle s_z \rangle \propto \langle \sigma_x \gamma \rangle$.

The Kramers doublet is split away from $\Gamma$ by spin-orbit interaction. We study the surface band structure near $\Gamma$ using $k \cdot p$ theory. To lowest order in $k$, the $2 \times 2$ effective

![FIG. 1 (color online). (i) Snowflakelike Fermi surface of the surface states on 0.67% Sn-doped Bi$_2$Te$_3$ observed in ARPES. (ii) A set of constant energy contours at different energies. From Y.L. Chen et al., Science 325, 178 (2009). Reprinted with permission from AAAS.](image-url)
Hamiltonian reads \( H_0 = v(k_x \sigma_y - k_y \sigma_x) \), which describes an isotropic 2D Dirac fermion. The form of \( H_0 \) is strictly fixed by symmetry. In particular, the Fermi velocity \( v \) in \( x \) and \( y \) directions are equal because of the \( C_3 \) symmetry. The Fermi surface of \( H_0 \) at any Fermi energy is a circle. However, the Fermi surface observed in ARPES, reproduced in Fig. 1(i), is noncircular but snowflakelike: it has relatively sharp tips extending along six \( \Gamma M \) directions and curves inward in between. Moreover, as shown in Fig. 1(ii) (we refer the reader to the original work \[4\] for better resolution), the shape of constant energy contour is energy-dependent, evolving from a snowflake at \( E = 0.25 \) eV to a hexagon and then to a circle near the Dirac point. Throughout this Letter, energy is measured with respect to the Dirac point.

The observed anisotropic Fermi surface can only be explained by higher order terms in the \( k \cdot p \) Hamiltonian \( H(\tilde{k}) \) that breaks the emerging \( U(1) \) rotational symmetry of \( H_0 \). The form of \( H(\tilde{k}) \) is highly constrained by crystal and time-reversal symmetry. Under the operation of \( C_3 \) and \( M \), momentum and spin transform as follows:

\[
\begin{align*}
C_3 &: k_{\pm} \rightarrow e^{\pm i 2 \pi/3} k_{\pm}, \quad \sigma_{\pm} \rightarrow e^{\pm i 2 \pi/3} \sigma_{\pm}, \quad \sigma_z \rightarrow \sigma_z \\
M &: k_{+} \rightarrow -k_{-}, \quad \sigma_x \rightarrow \sigma_x, \quad \sigma_{yz} \rightarrow -\sigma_{yz},
\end{align*}
\]  

(1)

where \( k_{\pm} = k_x \pm ik_y \) and \( \sigma_{\pm} = \sigma_x \pm i \sigma_y \). \( H(\tilde{k}) \) must be invariant under (1). In addition, time-reversal symmetry gives the constraint

\[
H(\tilde{k}) = \Theta H(-\tilde{k}) \Theta^{-1} = \sigma^y H^*(\tilde{k}) \sigma^y. \tag{2}
\]

We then find that \( H(\tilde{k}) \) must take the following form up to third order in \( \tilde{k} \):

\[
H(\tilde{k}) = E_0(\tilde{k}) + v_1(k_x \sigma_y - k_y \sigma_x) + \frac{\lambda}{2} (k_x^3 + k_y^3) \sigma_z,
\]

(3)

where \( E_0(\tilde{k}) = k^2/(2m^*) \) generates particle-hole asymmetry and the Dirac velocity \( v_1 = v(1 + ak^2) \) contains a second-order correction. The last term in (3), which we call \( H_w \), is most important. Unlike the other terms, \( H_w \) is only invariant under threefold rotation (as the Bi\(_2\)Te\(_3\) crystal structure does) and therefore is solely responsible for the hexagonal distortion of the otherwise circular Fermi surface. We note that \( H_w(\tilde{k}) \) vanishes in mirror-symmetric direction \( \Gamma M \) because \( \sigma^z \) is odd under mirror, and \( H_w(\tilde{k}) \) reaches maximum along \( \Gamma K \). The surface band dispersion of \( H(\tilde{k}) \) is

\[
E_{\pm}(\tilde{k}) = E_0(\tilde{k}) \pm \sqrt{v_1^2 k^2 + \lambda^2 k^6 \cos^2(3\theta)}. \tag{4}
\]

Here \( E_{\pm} \) denote the energy of upper and lower band, and \( \theta \) is the azimuth angle of momentum \( \tilde{k} \) with respect to the \( x \) axis (\( \Gamma K \)). Although the Hamiltonian \( H \) is threefold invariant, the band structure is sixfold symmetric under \( \theta \rightarrow \theta + 2\pi/6 \) because of time-reversal symmetry.

The hexagonal warping term \( H_w \) describes cubic spin-orbit coupling at the surface of rhombohedral crystal systems, and, to the best of our knowledge, it has not been reported before. It is instructive to compare \( H_w \) with the well-studied trigonal warping in graphene \[15\]. Although graphene’s band structure also has Dirac points and its \( k \cdot p \) Hamiltonian \( H_K(\tilde{k}) \) has \( C_{3v} \) symmetry, the warping term in graphene is of a completely different form. This is because time-reversal operation \( \Theta \) acts differently for spin \( 1/2 \) (\( \Theta = i\sigma_z K \)) and spinless fermions (\( \Theta = K \)). In graphene time-reversal symmetry takes the latter form, and, together with inversion symmetry, leads to \( H_K(\tilde{k}) = \tau_H \tilde{k}^3 \tau_z (\tau_z = \pm 1 \) denote two sublattices), as opposed to its partner Eq. (2) in Bi\(_2\)Te\(_3\). As a result, a different trigonal warping term \((k_x^3 \tau_+ + k_y^3 \tau_-)\) is symmetry-allowed in graphene.

We now show that \( H_w \) naturally explains the observed energy-dependent shape of the Fermi surface in Bi\(_2\)Te\(_3\). Using (4) we plot a set of constant energy contours of \( H(\tilde{k}) \) for \( 0 < E < 2E^* \) in Fig. 2, where \( E^* = v/a \) and \( a = \sqrt[3]{\sqrt{\lambda}/v} \) are the characteristic energy and length scale introduced by hexagonal warping. For simplicity, we have discarded \( E_0 \) and the quadratic correction to velocity, since they do not change the shape of the Fermi surface significantly. By plotting the \( k_x \) and \( k_y \) axis in the unit of \( \sqrt{v/a} \), Fig. 2 is obtained with no free parameter. As shown in the figure, the Fermi surface starts to deviate considerably from a circle and becomes more hexagonlike around \( E = 0.55E^* \). When \( E > E_c = \sqrt[3]{6/4}E^* = 0.69E^* \), the edge of the hexagon curves inward so that Fermi surface ceases to be purely convex. As \( E \) further increases, rounded tips start to develop at the vertices of the hexagon, which eventually become sharper, making the Fermi surface snowflakelike. The evolution of the Fermi surface with respect to energy matches well with the ARPES result shown in Fig. 1. Moreover, it follows from (4) that the vertices of the hexagon—where the Fermi surface extends outmost—always lie along \( \Gamma M \) independent of the sign of \( \lambda \), in agreement with ARPES data.

FIG. 2 (color online). (a) Constant energy contour of \( H(\tilde{k}) \). \( k_x \) and \( k_y \) axis are in the unit of \( \sqrt{v/a} \). (b) Constant energy contour at \( E = 1.2E^* \) is superimposed on the Fermi surface of Bi\(_2\)Te\(_3\).
Comparing the set of Fermi surfaces in Fig. 2(a) with the real Fermi surface in 0.67% Sn-doped Bi$_2$Te$_3$ (Fig. 1), we find the Fermi surface at $E_F = 1.2E^*$ is almost identical to the one measured in ARPES, as shown by superimposing the two in Fig. 2(b). By fitting the theoretical value of Fermi momentum along $\Gamma M (1.2/a)$ to the experimental one (0.11 Å$^{-1}$), we find $a = 10.9$ Å. Using the measured Fermi velocity $v = 2.55$ eV ⋅ Å, we obtain the magnitude of the hexagonal warping term: $\lambda = 250$ eV ⋅ Å$^3$. From that we find $E^* = 0.23$ eV and $E_p = 1.2E^*$ = 0.28 eV which agrees fairly well with the measured Fermi energy 0.25 eV [shown in Fig. 1(ii)]. The quantitative agreement between theory and experiment suggests that the Hamiltonian (3) describes the surface band structure of Bi$_2$Te$_3$ quite well in a wide energy window at least up to 0.25 eV. As an independent check of the theory, we consider the nonlinear correction to surface band dispersion near $\Gamma$. Equation (4) predicts that the leading order correction due to $H_w$ starts at fifth order in $k$ and is angle-dependent:

$$\delta E(k, \theta) = na^4k^2\cos^2(3\theta)/2.$$  

Since the surface band dispersions along $\Gamma K$ and $\Gamma M$ directions have been measured in ARPES [4,5], Eq. (5) can be tested by fitting to $E_{\Gamma M}(k) - E_{\Gamma K}(k)$, which also gives an independent way of obtaining $\lambda$. The two other parameters $m^*$ and $\alpha$ in $H(\vec{k})$ can also be extracted by a careful fitting to the band dispersion.

From now on, we predict a variety of important effects of hexagonal warping in Bi$_2$Te$_3$. First, because $H_w$ couples to $\sigma_z$, the spin polarization of surface states should have an out-of-plane component $s_z \propto (\sigma_z)$. Since spin polarization along $\Gamma M$ has been found to be almost 100% polarized in a very recent ARPES experiment [5], we conclude that the doublets $|\psi_{1\pm}\rangle$ at $\Gamma$ are almost pure spin eigenstates, i.e., $s_z \approx (\sigma_z)$, which agrees with a theoretical band structure calculation [16]. $s_z$ is then calculated from (3): $s_z = \cos(3\theta)/\sqrt{\cos^2(3\theta) + 1/(ka)^2}$. The out-of-plane spin polarization is momentum-dependent and can reach as high as 60% of the full polarization along $\Gamma K$ for the Fermi surface in Fig. 1. We hope this pattern of out-of-plane spin polarization can be tested in future spin-resolved ARPES.

Second, hexagonal warping gives a novel mechanism for opening up an energy gap at the Dirac point. Consider an in-plane magnetic field $B_{\parallel}$, which only couples to the spin $H_{\text{Zeeman}}^\parallel = g_{\parallel}B_{\parallel} \cdot \vec{\sigma}$. From (3) we find the Dirac point is shifted away from $\Gamma$ to $\vec{k} - g_{\parallel}B_{\parallel}/v$. In addition to that, a mass term is generated at $\vec{k}^*$: $M\sigma_z = (g_{\parallel}B_{\parallel})^3 \times \sin(3\varphi)/E^{2*}$ ($\varphi$ is the angle between $\vec{B}_{\parallel}$ and $\Gamma K$), which opens up an energy gap. When the Fermi energy is tuned, e.g., by doping [4,5], to lie within the gap, the insulating state at the surface realizes quantum Hall effect without Landau levels [17].

Third, hexagonal warping of the Fermi surface has drastic effects on the Friedel oscillation of local density of states (LDOS) around a nonmagnetic point defect in STM. The LDOS oscillation at a fixed energy decays algebraically as a function of distance away from the defect. In a normal 2D metal, the leading order $(1/x)$ decay of LDOS in a given direction $\hat{x}$ comes from scattering between states at “stationary points” on the Fermi surface, where the Fermi velocity is parallel to $\hat{x}$ [18]. For a convex constant energy contour below $E_c$ as shown in Fig. 3(a), only a single pair of stationary points exists at $\vec{k}$ and $-\vec{k}$. However, since the two states at $\vec{k}$ and $-\vec{k}$ here carry opposite spins, scattering between them is forbidden by time-reversal symmetry—a fundamental property of surface states on a topological insulator. The LDOS oscillation then vanishes at leading order. Now consider a nonconvex constant energy contour above $E_c$. As shown in Fig. 3(b), multiple pairs of stationary points exist. Since interpair scattering is still allowed, the LDOS oscillation will be restored at leading order. Therefore, according to the convexity of constant energy contour, two types of Friedel oscillation patterns should appear at different ranges of bias voltage.

In the last part of this work, we explore the possibility of a SDW phase on the surface of Bi$_2$Te$_3$. We note that the Fermi surface is nearly a hexagon for $0.55E^* < E < 0.9E^*$. The almost flat pieces on the edges of the hexagon lead to strong nesting at wave vectors $Q_i = 2k_Fe_i$, $i = 1, \ldots, 3$, where $k_Fe_i$ is the Fermi momentum in three equivalent $\Gamma K$ directions. A density-wave ordered phase may then exist at a finite interaction strength. Since the surface states at $\vec{k}$ and $-\vec{k}$ have opposite spins, a charge-density wave cannot connect them and is thus disfavored. We are therefore motivated to consider possible SDW phases.

We now discuss the phase diagram of SDW in a Landau-Ginzburg theory based on general symmetry considerations. We define the order parameters of the SDW as

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig3.png}
\caption{(color online). Illustration of scattering processes due to a point defect that causes the oscillation of LDOS. In a given direction $\hat{x}$ along $\Gamma M$, the oscillation is dominated by scattering between stationary points marked by dots, where the Fermi velocity is parallel to $\hat{x}$. (a) A convex constant energy contour has a single pair of stationary points at $\vec{k}$ and $-\vec{k}$. (b) A nonconvex constant energy contour has three pairs of stationary points. Intrapair scatterings in (a) and (b) are forbidden by time-reversal symmetry. But interpair scatterings in (b), for example, those between $k_1$ and $k_3$, are allowed. Therefore, the LDOS oscillation at leading order is absent in (a) but exists in (b).}
\end{figure}
follows,
\[
\begin{align*}
\phi_\parallel &= \sum_k \langle c_{k+Q_i}^\dagger \mathbf{e}_i \cdot \mathbf{s} c_k \rangle, \\
\phi_{i\perp} &= i \sum_k \langle c_{k+Q_i} (\hat{z} \times \mathbf{e}_i) \cdot \mathbf{s} c_k \rangle, \\
\phi_{i\parallel} &= i \sum_k \langle c_{k+Q_i} \sigma^\parallel c_k \rangle,
\end{align*}
\]
(6)
where \(c_{k+Q_i}^\dagger = (c_{k+Q_i}^\dagger, c_{k+Q_i}^\dagger)\) are electron creation operators. For each \(Q_i\), we have chosen a local frame for in-plane spin components labeled by \(\parallel\) and \(\perp\) which are parallel and perpendicular to \(Q_i\), respectively. The order parameters thus defined transform nicely under the operations of rotation, mirror, time reversal, and translation:
\[
\begin{align*}
C_3: \phi_{i\mu} &\to \phi_{i+1,\mu}, \\
\Theta: \phi_{i\mu} &\to -\phi_{i\mu}, \\
M_z: \phi_{i\mu} &\to \phi_{i\mu}^*, \; \phi_{2\mu} &\to \phi_{3\mu}^*, \\
T_d: \phi_{i\mu} &\to e^{i\Theta} d^{\mu} \phi_{i\mu}, \quad \mu = ||, \perp, z.
\end{align*}
\]
(7)
We remark that because spin and momentum are locked by spin-orbit coupling, there is no \(SU(2)\) symmetry for spin alone. Thanks to the appropriate choice of order parameters, symmetry operations (7) only act in the space of ordering wave vectors labeled by \(i\) index.

The Landau free energy \(F\) must be invariant under these symmetry operations. Only terms with even powers of \(\phi_{i\mu}\) can exist because of time-reversal symmetry. At second order, we have
\[
F_2 = \frac{1}{2} \chi_{\mu\nu} \sum_{i=1}^3 \phi_{i\mu}^* \phi_{i\nu},
\]
(8)
where the susceptibility matrix \(\chi_{\mu\nu}\) is real and symmetric because of mirror symmetry. \(\chi_{\mu\nu}\) is positive definite in the normal state. When the temperature is lowered below \(T_c\), one of the eigenvalues of \(\chi_{\mu\nu}\) first becomes negative, and the surface undergoes a transition to a SDW. The spin configuration is then determined by the corresponding eigenvector \(\nu_{\mu}\). For example, for a stripe SDW along the \(x\) direction, \(\hat{S}(x, y) = (v_{||} \cos(Qx), v_{\perp} \sin(Qx), v_z \sin(Qx))\) with an appropriate choice of origin.

The free energy (8) to second-order has an emerging \(U(3)\) symmetry \(\phi_{i\mu} \to U_{ij} \phi_{j\mu}\). So single- and multiple-Q SDWs are degenerate. We now show that higher order terms in \(F\) break the \(U(3)\) symmetry and pick out a particular spatial ordering pattern. For that purpose, it is convenient to write \(\phi_{i\mu} = \xi_i v_{\mu} \sum_j |\xi_j|^2 = 1 \) and use \(\xi_i\) as a new set of order parameters, which also transforms according to (7). At fourth order, we find an anisotropy term \(F_4 = u \sum_{i=1}^3 |\xi_i|^4\). The sign of \(u\) determines the relative weight of \(\phi_i\) in the ordered phase. For \(u < 0\), only one of \(\phi_i\), say, \(\xi_1\), is nonzero. The resulting SDW forms a one-dimensional stripe, which breaks \(C_3\) but is invariant under mirror symmetry. For \(u > 0\), \(|\xi_1| = |\xi_2| = |\xi_3|\) in the ordered phase, so that SDW forms a two-dimensional lattice. Each individual phase of \(\xi_i\) depends on the choice of origin. Only the global phase of \(\xi_1\xi_2\xi_3\) is gauge invariant and is fixed by the sixth-order term of the form \(C(\phi_1^2 \phi_2 \phi_3^2)^2 + C'(\phi_1^2 \phi_2^2 \phi_3^2)^2\) in \(F\).

This work was initiated at University of Pennsylvania. We thank Charlie Kane for inspiring discussions. We are especially indebted to Bertrand Halperin for many insightful discussions and helpful comments on the manuscript. We thank Zahid Hasan and David Hsieh for numerous discussions on ARPES, as well as Aharon Kapitulnik, Anton Akhmerov, and especially Cenke Xu for useful conversations. This work was supported by the Harvard Society of Fellows and NSF Grant No. DMR-0605066.

Note added.—During the final stage of this work, we learned that Alpichshev et al. [19] imaged with STM the standing wave of surface states on \(Bi_2Te_3\) near a line defect (instead of a point defect considered in this work). The LDOS oscillation was found to exist in the energy range with snowflakelike constant energy contour, but strongly suppressed in the range with circular constant energy contour. This supports our explanation of the correlation between LDOS oscillation and convexity of constant energy contour.

\*liangfu@physics.harvard.edu