Robustness and universality of surface states in Dirac materials

Citation: Shtanko, Oles and Leonid Levitov. "Robustness and Universality of Surface States in Dirac Materials." Proceedings of the National Academy of Sciences of the United States of America 115, 23 (May 2018): 5908–5913. doi:10.1073/pnas.1722663115. © 2018 The Author(s)

As Published: http://dx.doi.org/10.1073/PNAS.1722663115

Publisher: National Academy of Sciences (U.S.)

Persistent URL: http://hdl.handle.net/1721.1/121119

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Robustness and universality of surface states in Dirac materials

Oles Shtanko[a,1] and Leonid Levitov[a]

[a]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by David Vanderbilt, Rutgers, The State University of New Jersey, Piscataway, NJ, and approved April 10, 2018 (received for review December 30, 2017)

Ballistically propagating topologically protected states harbor exotic transport phenomena of wide interest. Here we describe a nontopological mechanism that produces such states at the surfaces of generic Dirac materials, giving rise to propagating surface modes with energies near the bulk band crossing. The robustness of surface states originates from the unique properties of Dirac–Bloch wavefunctions which exhibit strong coupling to generic boundaries. Surface states, described by Jackiw–Rebbi-type bound states, feature a number of interesting properties. Mode dispersion is gate tunable, exhibiting a wide variety of different regimes, including nondispersing flat bands and linear crossings within the bulk bandgap. The ballistic wavelike character of these states resembles the properties of topologically protected states; however, it requires neither topological restrictions nor additional crystal symmetries. The Dirac surface states are weakly sensitive to surface disorder and can dominate edge transport at the energies near the Dirac point.

ballistic transport | Dirac materials | surface transport

Surface states and the mechanisms allowing them to propagate along crystal boundaries—the topics of long-standing interest of the theory of solids—acquired a new dimension with the advent of topological materials (1, 2). In these materials robust surface states are made possible by nontrivial topology of the bulk bands (1, 3). Here we outline a different mechanism leading to robust surface states, realized in solids with Dirac bands that mimic relativistic particles near band crossings (2). In this scenario robust surface states originate from unusual scattering properties of Dirac particles, occurring for generic boundary conditions at the crystal boundary. As we will see, since this mechanism does not rely on band topology, it can lead to robust surface states in solids with either topological or nontopological bulk band dispersion. The surface states exist for either gapless or narrow-gapped Dirac bulk bands. Furthermore, these states are to some degree immune to surface disorder. Namely, as discussed below, surface modes can propagate coherently by diffracting around surface disorder through system bulk. This diffraction behavior suppresses backscattering and results in exceptionally long mean free paths. Since Dirac surface states require neither special topological properties of the band structure nor special symmetry, they are more generic than the topological surface states. As such, these states can shed light on recent observations of edge transport in nontopological materials.

Indeed, it is often taken for granted that an observation of edge transport signals nontrivial band topology (4–7). However, recent experiments on semiconducting structures, where tunable band inversion enables switching between topological and nontopological phases, indicate that current-carrying edge modes can appear regardless of the band topology (8–11). One piece of evidence comes from transport and scanning measurements in InAs/GaSb, which indicate that helical edge channels survive switching from a topological to a trivial band structure (8). Additionally, refs. 9 and 10 report an unexpectedly weak dependence of edge transport on the in-plane magnetic field. Namely, it is found that the edge transport is observed even when Zeeman splitting is considerably larger than the spin-orbit splitting, i.e., in the nontopological regime. A similar behavior is observed in HgTe devices (11). Furthermore, recently several groups have used Josephson interferometry to directly image long-range edge currents in graphene, a signature nontopological material (12–16). These observations point to the existence of robust nontopological surface states.

As we will see, the Dirac surface states can arise naturally due to strong coupling of electronic waves to generic boundaries. Namely, the phase shifts of waves in the bulk that scatter off the surface have a strong energy dependence near the Dirac point where the particle and "antiparticle" bands cross (or nearly cross). The energy dependence of phase shifts, as always, leads to the formation of states behaving as plane waves confined to the surface and decaying into the material bulk as evanescent waves. The formation of these states is governed by a mechanism that resembles the seminal Jackiw–Rebbi (JR) theory (17) for the states formed at the domain walls separating regions with sign-changing Dirac mass. Unlike the JR problem, however, the Dirac surface states do not have a topological character; i.e., in general they are not protected by topological invariants. Nevertheless, these states are robust and form surface modes with the energies near the Dirac crossing of bulk bands (Fig. 1*C*).

The diffraction-based mechanism that suppresses backscattering and makes the Dirac surface states insensitive to surface disorder has an interesting analogy with the properties of the high-mobility electron gas realized in GaAs/AlAs quantum wells. In these systems an exceptionally high mobility could be achieved by adjusting the well width to reduce the overlap of the carrier mode with the well boundary and, in this way, suppress carrier

Significance

We predict that generic Dirac materials host ballistically propagating surface states. Despite being nontopological, these states are robust and insensitive to surface disorder, occurring in a universal manner at generic boundaries, both atomically smooth and disordered. These properties originate from the multicomponent character of the Dirac–Bloch wavefunction, which results in strong coupling to generic boundaries. Dispersion of the surface states can be tuned by external gates to exhibit a variety of different regimes, including flat bands and linear crossings within the bulk bandgap. These surface states provide a natural explanation for anomalous long-range edge transport observed recently in a variety of Dirac materials in both topological and nontopological phases.

Author contributions: O.S. and L.L. designed research, performed research, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: oles@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1722663115/-/DCSupplemental.

Published online May 22, 2018.

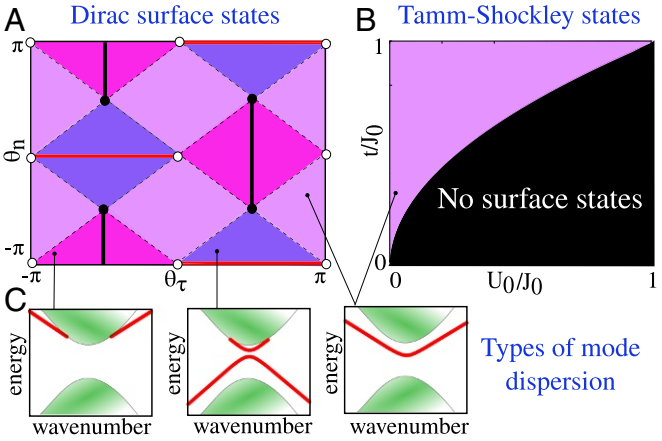

Fig. 1. (*A* and *B*) Phase diagrams for Dirac surface states (*A*) and Tamm–Shockley surface states (*B*) as a function of parameters that control boundary conditions (main text). Dirac surface states occur in the entire parameter space except for a subset of measure zero (black lines) and are therefore generic. In contrast, the Tamm–Shockley surface states appear upon fine tuning, occurring in a relatively small region of the parameter space, whereas most of phase space is incompatible with surface states. Different types of Dirac surface bands, shown in *C*, correspond to regions of different color (purple, blue, and pink) in *A* and *B*. For a detailed discussion see main text.

scattering at the surface disorder. Scattering suppression through this mechanism results in a dramatic increase of the mean free path, growing rapidly vs. the well width, $\ell \sim w^n$ with a large n (18). Recently in wide quantum wells mobilities exceeding $10^7\,\mathrm{cm}^2/\mathrm{V\,s}$ were demonstrated (19). Likewise, Dirac surface states, being nontopological, are, in principle, susceptible to disorder. However, the large width of Dirac surface states (arising due to their slow decay into the bulk) can strongly reduce their overlap with the atomic-scale disorder at the surface and make them effectively immune to surface scattering. In this regime, in direct analogy with the carriers in wide quantum wells, the surface states can propagate coherently by diffracting around surface disorder. This remarkable behavior is discussed in detail below. We will see that, while Dirac surface states may be gapped (Fig. 1) and are generally not immune to the bulk disorder, their decoupling from the surface disorder can suppress backscattering and enable large mean free paths already for moderately clean materials. This property also weakens the dependence of these states on the details of the surface structure.

At this point it is instructive to compare Dirac surface states to the well-known Tamm–Shockley states. These are nontopological states residing inside the bandgap that governs surface physics of many semiconductors. The Tamm–Shockley states form a surface band that splits off the bandgap edge upon varying the surface potential. The existence of these states depends on the details of the crystal structure near the surface, which makes them nonuniversal and less robust than the Dirac surface states (*SI Appendix, The Tamm-Shockley Surface States*). Indeed, unlike the Dirac states, they require fine tuning and are present only in a part of parameter space (Fig. 1*B*). Further, since these states are typically confined to the surface on the scale of a few lattice constants, they are sensitive to surface disorder potential and, unlike Dirac surface states, are easily localized by the disorder.

Dirac surface states arise in diverse fields, from high-energy to solid-states physics. Early work on Dirac surface states in a periodic potential dates back to the 1960s (20, 21). These studies have led to interesting developments in nuclear and particle physics such as the MIT bag model and neutrino billiards (22–25). Recently, the interest in this problem has been renewed with the advent of graphene and other Dirac materials (26, 27). However, while a number of important aspects of these states have been explored for atomically clean boundaries (28–31), the ease with which Dirac surface states emerge, as well as their ubiquitous character, has remained unnoticed. Below we discuss the mechanism underlying this behavior and address the key properties such as robustness, stability, and immunity to disorder. Our work complements recent studies of topological semimetals (32).

Surface States: General Theory

We first consider the general properties of Dirac surface states in a 3D solid and then focus on the case of a graphene monolayer. We analyze a Dirac Hamiltonian in 3D with boundary conditions of a general form

$$H = \boldsymbol{\alpha}\mathbf{p}v + \beta\Delta, \qquad M\psi_\alpha\big|_{\mathrm{B}} = \psi_\alpha\big|_{\mathrm{B}}, \qquad [1]$$

with $\boldsymbol{p} = -i(\partial_x, \partial_y, \partial_z)$ the momentum operator. Here $\{\psi_\alpha\}$ is a four-component wavefunction, and $\{\alpha_i\}$ and β are 4×4 Dirac matrices satisfying the canonical algebra $\alpha_i\alpha_{i'} + \alpha_{i'}\alpha_i = 2\delta_{ii'}$, $\alpha_i\beta + \beta\alpha_i = 0$, $\alpha_i^2 = \beta^2 = 1$. The parameters v and Δ in Eq. **1** describe the 3D Bloch band structure near the Dirac band crossing. The matrix M is a unitary Hermitian operator constrained by time-reversal symmetry and current conservation (26, 27),

$$[\mathcal{T}, M] = 0, \qquad I_{\mathrm{B}}M + MI_{\mathrm{B}} = 0, \qquad [2]$$

where $\mathcal{T}$ is the time-reversal operator, and I_{B} is the current component normal to the boundary.

The form of these boundary conditions and the constraints on M in Eq. **2** can be understood as follows. First, since the Dirac equation is first order in derivatives, the boundary condition must be stated in terms of ψ alone without invoking derivatives of ψ. The most general boundary condition can therefore be written as $(M-\hat{1})\psi\big|_{\mathrm{B}} = 0$ with M a suitably chosen 4×4 matrix with two eigenvalues equal to $+1$. Every eigenvalue equal to $+1$ yields a scalar relation between the components of ψ, providing a convenient encoding of the boundary conditions in a matrix form. A considerable simplification can be achieved, without any loss of generality, by choosing M to be a Hermitian matrix with eigenvalues $+1$, $+1$, -1, and -1 (the eigenvectors corresponding to -1 eigenvalues do not impact the boundary conditions in any way). The form of matrix M is further constrained by the requirements due to time-reversal symmetry and probability current conservation (unitarity of scattering at the boundary requires that the eigenvectors of M with $+1$ eigenvalues give current which is tangential to the boundary). These constraints are expressed by the first and second relations in Eq. **2**, respectively (for a more detailed discussion see refs. 26 and 27).

The task of finding surface states from the Dirac Hamiltonian of a general form, Eq. **1**, can be simplified by transforming it to a 1D Dirac problem as follows. Without loss of generality, we take the system boundary to be a 2D plane perpendicular to the x direction. Accounting for translation invariance along y and z, we use Fourier transform, seeking the states of the form $\psi(x)e^{ik_y y + ik_z z}$. Assuming the system to be homogeneous and isotropic in the y–z plane, we can choose a new coordinate system such that $\mathbf{k} \parallel \hat{y}$. This amounts to a unitary transformation of the spinor wavefunction and Dirac matrices, $\alpha_i' = U^{-1}\alpha_i U$, $\beta' = U^{-1}\beta U$, such that

$$\begin{aligned} \alpha_1' &= \alpha_1, & \alpha_2' &= (\alpha_2 k_y + \alpha_3 k_z)/k, \\ \beta' &= \beta, & \alpha_3' &= (\alpha_2 k_y - \alpha_3 k_z)/k, \end{aligned} \qquad [3]$$

where $k = \sqrt{k_y^2 + k_z^2}$.

To simplify the analysis, we use, without loss of generality, an asymmetric representation for the transformed matrices

$$\alpha'_{1,2}=\begin{pmatrix}\sigma_{1,2} & 0\\ 0 & \sigma_{1,2}\end{pmatrix},\ \alpha'_3=\begin{pmatrix}0 & \sigma_3\\ \sigma_3 & 0\end{pmatrix},\ \beta'=\begin{pmatrix}\sigma_3 & 0\\ 0 & -\sigma_3\end{pmatrix},\quad [4]$$

writing it in a shorthand notation as $\alpha'_1=\tau_0\sigma_1$, $\alpha'_2=\tau_0\sigma_2$, $\alpha'_3=\tau_1\sigma_3$, $\beta'=\tau_3\sigma_3$, where τ_i and σ_i are 2×2 Pauli matrices and τ_0 is a unit 2×2 matrix (from now on, we suppress it for brevity). This transforms the 3D Dirac equation into a quasi-1D problem $H\psi_k(x)=\varepsilon_k\psi_k(x)$ on a half-line $x\geq 0$:

$$H=-iv\partial_x\sigma_1+vk\sigma_2+\Delta\tau_3\sigma_3,\quad M\psi_k(0)=\psi_k(0). \quad [5]$$

Surface states in 3D correspond to the one-dimensional bound states obtained from the Hamiltonian in Eq. **5** (see Fig. 2*C*).

The advantage of this representation, in particular the choice of α'_i, is that it allows us to bring the matrix M to a tractable form. All possible situations that may occur near the surface are parameterized by different choices of the matrix M, whereas the Hamiltonian H takes a standardized form. This provides a vehicle for classifying different types of behavior, parameterized by the M manifold. The block representation in Eq. **4** greatly facilitates this analysis. In this representation the operators in Eq. **2** take the form $I_B=v\alpha'_1=v\sigma_1$ and $\mathcal{T}=\tau_2\sigma_2\mathcal{K}$, where $\mathcal{K}$ is complex conjugation. The constraints on M given in Eq. **2** can now be resolved as follows (26, 27). The relation $I_BM+MI_B=0$ implies that $M\sim\mathbf{n}\cdot\sigma$, where $\mathbf{n}$ is a vector tangential to the boundary or a linear combination of several such terms. Combining it with the first relation in Eq. **2** gives

$$M=(\nu\cdot\tau)(n\cdot\sigma),\qquad n_x=0, \quad [6]$$

where n and ν are three-component unit vectors. The Hamiltonian in Eq. **5** is invariant under unitary transformations of valley matrices τ_i preserving τ_3. Therefore, M can be fixed, without loss of generality, by specifying only two real phases θ_n and θ_τ:

$$M(\theta_\tau,\theta_n)=(\tau_3\cos\theta_\tau+\tau_2\sin\theta_\tau)(\sigma_3\cos\theta_n+\sigma_2\sin\theta_n). \quad [7]$$

The 1D problem in Eq. **5** can now be solved for M of a general form detailed in Eq. **7**, giving states that decay into the bulk as evanescent waves $\psi_k^s\sim\exp(-\mu_{k,s}x)$ (Fig. 2*C*). The energies $\varepsilon_{k,s}$ and the decay parameters $\mu_{k,s}$ obey

$$\begin{aligned}\varepsilon_{k,s}&=\Delta\cos\theta_\tau\cos\theta_n+sK\sin\theta_n\\ \mu_{k,s}&=\Delta\cos\theta_\tau\sin\theta_n-sK\cos\theta_n,\end{aligned}\quad s=\pm1, \quad [8]$$

where $K=\left(v^2k^2+\Delta^2\sin^2\theta_\tau\right)^{1/2}$ and s labels two possible dispersion branches (*SI Appendix, Dirac Mode Dispersion* and Fig. 1). Solutions confined to the surface exist only when $\mu_{k,s}>0$. The resulting modes and their evolution upon changing boundary conditions are illustrated in Fig. 2*A* and *B* (see discussion below).

The dependence of the dispersion in Eq. **8** on the angles θ_n, θ_τ indicates that the surface modes exist for generic θ_n, θ_τ values, disappearing only for a subset of measure zero. Possible dispersion types, comprising either two branches or a single branch, are shown in Fig. 1. The modes lie inside and outside the bandgap of the bulk spectrum. The two branches, when present, are separated by a minigap which closes at particular values θ_n, θ_τ. Notably, the modes are present for both $\Delta\neq 0$ and $\Delta=0$, i.e., for gapped and gapless bulk bands. In the latter case the modes lie outside the bulk Dirac continuum $|\varepsilon|>v|k|$ and have linear dispersion of the form $\varepsilon=-v\sin\theta_n|k|$. This gives propagation velocity of $v\sin\theta_n$. The reduction in velocity relative to the bulk velocity value provides a clear experimental signature of surface modes.

A Relation to the Jackiw–Rebbi Bound States

To better understand the unique properties of the bulk Bloch states which enable surface states we sketch a relation between

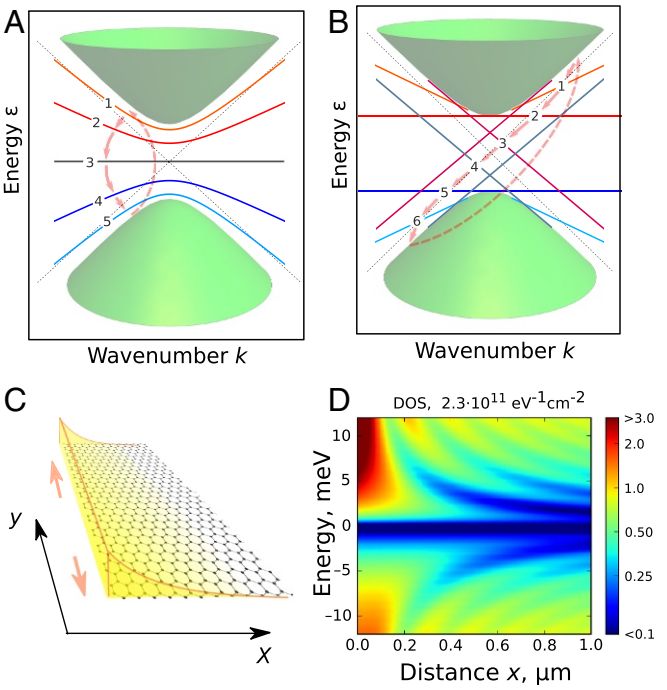

Fig. 2. (*A* and *B*) Surface states in monolayer graphene generated by surface potential for (*A*) armchair boundary conditions and (*B*) zigzag boundary conditions. Mode dispersion changes in a cyclical manner $\ldots 1\to 2\to 3\to 4\to 5\to 6\to 1\ldots$ with the increase of the effective potential strength θ_V (see Eq. **18** and accompanying discussion). (*C*) Edge modes are confined at the boundary and propagate along it in both directions, as indicated by arrows. (*D*) The density of states (DOS) as a function of energy and distance from zigzag edge, with DOS in the bulk far from the edge subtracted to enhance contrast. Shown are results for case *i* in Eq. **12** for the phase-shift value $\theta_V=-\pi/4$. The bright peak in DOS near the edge $x=0$ at positive energies is due to surface states. The surface states contribution is embedded in a family of Friedel oscillations dispersion as $x\sim\hbar v/\varepsilon$.

our problem and the seminal JR problem of the midgap states of the 1D Dirac operator with a sign-changing mass. As a first step we perform a similarity transformation that brings M to a standardized form by moving all the complexity of the problem from the boundary conditions into the transformed Hamiltonian (*SI Appendix, Transformation to the Universal Boundary Conditions*). The transformation is generated by a 4×4 unitary matrix that is position independent (but in general is k dependent), giving two decoupled 2×2 Hamiltonians

$$H=\begin{pmatrix}H_+ & 0\\ 0 & H_-\end{pmatrix},\qquad M=\begin{pmatrix}\sigma'_2 & 0\\ 0 & \sigma'_2\end{pmatrix}, \quad [9]$$

on the half-line $x\geq 0$ in the new (generally, k-dependent) basis,

$$H_s=-iv\partial_x\sigma'_1+\varepsilon_{k,s}\sigma'_2+\mu_{k,s}\sigma'_3,\quad s=\pm1, \quad [10]$$

where $\varepsilon_{k,s}$ and $\mu_{k,s}$ are defined in Eq. **8**. The matrices σ'_i are (in general, k- and s-dependent) superpositions of σ_i.

The 2×2 block structure of the transformed Hamiltonian can now be used to solve the boundary value problem. This is done by extending the problem on the half-line $x>0$ to that on a full line $-\infty<x<\infty$, described by a Hamiltonian with a mass kink:

$$H_s^{\mathrm{R}}=-iv\partial_x\sigma'_1+\varepsilon_{k,s}\sigma'_2+\mu_{k,s}\sigma'_3\mathrm{sgn}(x). \quad [11]$$

To identify the eigenstates which lie in the "physical" subspace of the doubled Hilbert space, we note that the Hamiltonian possesses reflection symmetry $[H_s^{\mathrm{R}},\mathcal{R}]=0$, where $\mathcal{R}=\sigma'_2\mathcal{I}$ and $\mathcal{I}$ is

spatial inversion $x \to -x$. The solutions of the problem in Eq. **10** are given by the R-symmetric eigenstates of H_s^{R} satisfying $R\psi(x) = \psi(x)$, projected on $x > 0$.

This representation helps us to understand the robustness of surface states. It is instructive to treat $\mu_{k,s}$ as a fixed mass and use $\varepsilon_{k,s}$ as a tuning parameter. For $\varepsilon_{k,s} = 0$ Eq. **11** is nothing but the canonical JR problem, yielding zero-mode eigenstates which at the same time are eigenstates of σ_2' (17). Upon varying $\varepsilon_{k,s}$ in Eq. **11** these states remain bound to the surface albeit with a shifted energy $\varepsilon = \varepsilon_{k,s}$. This energy, taken as a function of k, defines dispersion of surface states. For $|\varepsilon_{k,s}| \geq \mu_{k,s}$ the bound states cease to exist. The k dependence of $\varepsilon_{k,s}$ and $\mu_{k,s}$ is such that the bound state may disappear in a finite range of k but persist at large enough k (with the exception of a measure-zero subset of θ_n and θ_τ shown in Fig. 1*A*).

Edge States in Graphene

This general discussion has direct implications for graphene, the Dirac material best studied to date. In monolayer graphene, σ_i and τ_i are 2×2 matrices representing pseudospin and valley degrees of freedom, respectively. Pristine graphene is gapless with the carrier velocity $v \sim 10^6$ m/s. A gap as large as $\Delta \sim 30$ meV can be created in graphene/hexagonal boron nitride superlattices. However, as discussed above, the gap has no direct significance for the existence of surface states. The types of states depend solely on the boundary conditions, i.e., the values of the phases θ_n and θ_τ parameterizing M, which depend on the symmetries and edge structure.

Particle–hole symmetry $\mathcal{C}$, if present, generates universal values θ_n and θ_τ (27); namely, surface states are reduced to just two distinct types, isomorphic to those found for crystalline zigzag and armchair edges. The allowed values are

$$(i)\ \theta_n = 0, \pi,\ \theta_\tau = 0, \pi; \quad (ii)\ \theta_n = \pm\frac{\pi}{2},\ \theta_\tau = \pm\frac{\pi}{2}. \qquad [12]$$

Boundary conditions in these two cases are given by $M_1 = \pm\tau_3\sigma_3$ and $M_2 = \pm\tau_2\sigma_2$, respectively (33). In case *i* surface states form a flat band that touches one of the bulk bands bottom or top, $\varepsilon = \pm\Delta$. In case *ii* there are no surface states. However, as we now show, these restrictions are lifted for realistic non-particle–hole-symmetric edges, allowing the phases θ_n and θ_τ to take generic nonuniversal values.

The $\mathcal{C}$ symmetry can be lifted by an edge potential that creates Dirac band bending near the edge. The edge potential can either occur naturally due to, e.g., edge reconstruction (34) or hydrogen passivation (35, 36) or be induced externally by a side gate, as illustrated in Fig. 3 *A* and *B*. Focusing on the first case, we consider electrostatic potential localized near the edge at a lengthscale of a few atomic spacings $r_0 \sim 1$ nm:

$$H(x) = H + V(x), \qquad \lim_{x\to\infty} V(x) = 0. \qquad [13]$$

Potential $V(x)$ affects states only in the vicinity of the edge. It is therefore convenient to incorporate the effect of $V(x)$ into the boundary conditions. This can be achieved by introducing a transfer matrix $T(x_1, x_2)$ connecting the wavefunction values, separately at each k, at adjacent points x_1 and x_2,

$$\psi_k(x_1) = T(x_1, x_2)\psi_k(x_2), \qquad 0 < x_2 < x_1 < \infty, \qquad [14]$$

where $\psi_k(x)$ is obtained from the Dirac equation with the Hamiltonian in Eq. **13**. The transfer matrix can be obtained by integrating the Dirac equation over x,

$$T(x_1, x_2) = \mathrm{Xexp}\int_{x_2}^{x_1} dx \frac{i}{v}\Big(D_k - \sigma_1 V(x)\Big), \qquad [15]$$

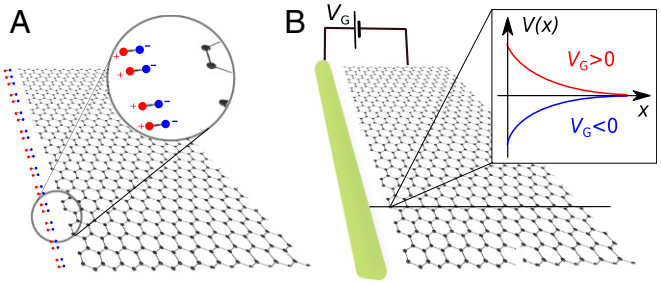

Fig. 3. (*A*) Armchair edge in graphene monolayer. Hydrogen passivation produces atomic-scale dipoles which create electrostatic potential at the edge. (*B*) Potential at the edge can be tuned by a side gate. Positive or negative potential attracts to the edge electrons or holes, respectively, modifying the dispersion of edge states as shown in Fig. 2 (Eq. **18** and accompanying discussion).

where $D_k = i\sigma_1\varepsilon + \sigma_3 vk - \tau_3\sigma_2\Delta$ and Xexp denotes an x-ordered exponential. Assuming that the term $V(x)$ gives negligible contribution for $x > r_0$, we can approximate the transfer matrix as a product of a free-particle contribution and a boundary term,

$$T(x, 0) = T(x, r_0)\,T(r_0, 0) \approx T_0(x, r_0)\,T(r_0, 0), \qquad [16]$$

where T_0 is a transfer matrix for zero potential $V(x) = 0$. The boundary contribution $T(r_0, 0)$ can be expressed in a closed form through $V(x)$ when the potential width r_0 is much smaller than electron wavelength $\lambda = \hbar v/\epsilon$. This is achieved by writing $\psi_k(x > r_0) = T_0(x, r_0)\psi_k(r_0)$, shifting the boundary position to $x = r_0$, and writing the boundary conditions as

$$\psi_k(r_0) = M_V\psi_k(r_0), \quad M_V = \Theta M\Theta^{-1}, \qquad [17]$$

where we denote $\Theta = T(r_0, 0)$. Under the conditions $kr_0 \ll 1$ and $\Delta r_0/\hbar v \ll 1$ we can ignore the first term inside Xexp in Eq. **15**. Approximating $\int_0^{r_0} V(x)dx \approx \int_0^{\infty} V(x)dx$ then gives a new matrix M that describes the boundary condition altered by $V(x)$:

$$M_V = M(\theta_\tau, \theta_n + \theta_V), \qquad \theta_V = \frac{2}{v}\int_0^{\infty} V(x)dx. \qquad [18]$$

This simple result is valid as long as the edge potential width r_0 is small compared with the electron wavelength in the bulk.

We note parenthetically that the latter condition restricts the validity of our approach to short-range edge potentials and narrow-gap Dirac band structures such that $r_0 \ll \lambda = \hbar v/\Delta$. In wide-gap band structures the interaction of carriers with the crystal surface is in general not described by a simple scalar potential model. In addition, long-range potentials can produce many bound states at the edge and thus create many surface modes.

To understand the impact of the edge potential on the edge states dispersion we consider the setup in Fig. 3*B* wherein $V(x)$ is tuned by a side gate. Through varying θ_V the edge states dispersion changes in a complex way, as illustrated in Fig. 2 *A* and *B* for the armchair and the zigzag edge. The armchair edge hosts a one-branch mode with relativistic dispersion

$$\varepsilon_k = -\sqrt{k^2v^2 + \Delta^2}\cos\theta_V. \qquad [19]$$

Interestingly, the mode in Eq. **19**, despite its relativistic appearance, has no $\mathcal{C}$-symmetric counterpart; i.e., it does not obey particle–hole symmetry. Furthermore, the dispersion acquires a flat-band character at $\theta_V = \pm\pi/2$.

The solution for the zigzag edge features a more complex behavior. For each θ_V value, the edge modes contain two distinct

branches, propagating to the right and to the left, as illustrated in Fig. 2*B*. Modes occurring at $0<\theta_V<\pi$ for zigzag edges with $\theta_n=0$ and at $-\pi<\theta_V<0$ for zigzag edges with $\theta_n=\pi$ (cases 3 and 4 in Fig. 2*B*) span both positive and negative energies. Upon variation of θ_V they sweep through the bulk bandgap. For other θ_V values, the modes also consist of two branches; however, their energies are either above or below the bulk gap (cases 1 and 6 in Fig. 2*B*). The dispersion becomes flat at $\theta_V=0,\ \pi$ (cases 2 and 5 in Fig. 2*B*). Since according to ref. 27 the zigzag boundary condition with $\theta_V=0$ describes a generic $\mathcal{C}$-symmetric lattice termination in monolayer graphene, our solution for $\theta_V\neq 0$ describes modes for a generic gated graphene edge.

The contribution of surface states to spatially resolved DOS is illustrated in Fig. 2*D* (for derivation see *SI Appendix, Spatially Resolved Density of States*). Surface states give rise to an enhanced DOS near the boundary for one type of carriers, electrons or holes, depending on the θ_V value. In Fig. 2*D* the contribution of surface states is seen as a high-DOS region at positive energies, embedded into the family of Friedel oscillations dispersing as $r\sim\hbar v/\varepsilon$.

The Role of Disorder

An interesting aspect of Dirac surface states is their weak interaction with surface disorder. Realistic crystal boundaries often feature strong disorder potential, arising due to dangling bonds and other defects, which impedes transport along the surface. Suppression of conduction by surface disorder is typically quite strong for the Tamm–Shockley states. However, Dirac surface states are to a great extent protected from surface scattering due to their small overlap with the surface disorder. This behavior is reminiscent of the carrier dynamics in GaAs/AlAs quantum wells where the mobility increases drastically with the well width due to a rapid mean free path growth $\ell\sim w^n$, with large n (18). Since the well width is much greater than the interface roughness scale, carriers can diffract around the interface disorder. In our case, a similar diffraction-enhanced conduction occurs since the width of the surface states, which defines their extent into the bulk and is on the order of the bulk wavelength $\sim\lambda_F$, is much larger than atomic surface roughness. Effectively, in this case the mode width takes on a role analogous to the width of quantum wells. This allows Dirac surface states to propagate quasi-ballistically with negligible surface scattering.

To illustrate the effect of backcattering suppression by electron-wave diffraction around surface disorder, we consider gaussian short-range correlated disorder at the graphene edge,

$$H_{\rm dis}=H+\xi(y)\delta(x),\quad \langle\xi(y)\xi(y')\rangle_{\rm dis}=\alpha\delta(y-y'), \qquad [20]$$

with α the disorder strength parameter. While the disorder spectrum is broad band, only the harmonics comparable to carrier wavelengths scatter efficiently (37) whereas the contribution of other harmonics is relatively weak (19).

In the limit of a weak disorder the mean free path can be evaluated by perturbation theory (*SI Appendix, Disorder at the Edge*). Here we discuss the results for the zigzag surface state (case *i* in Eq. **12**). In this case, the mean free path is

$$\ell=\frac{\lambda_{\rm B}^2}{\zeta},\qquad \zeta=\frac{8\pi^2\alpha}{\hbar^2v^2}\cot^2\theta_V, \qquad [21]$$

where $\lambda_{\rm B}=2\pi\hbar v\sin\theta_V/\varepsilon_F$ is the carrier wavelength in the edge mode, and ε_F is Fermi energy. The lengthscale ζ, proportional to disorder strength, can be estimated as $\zeta\sim\alpha/\hbar^2v^2\sim U_0^2a^3/\hbar^2v^2\sim 1$ nm, where $U_0\sim 1$ eV is an atomic-scale potential and $a\sim 1$ nm is surface roughness. For $\zeta\ll\lambda_{\rm B}$ Eq. **21** predicts mean free path values much greater than the carrier wavelength; the dimensionless parameter $\lambda_{\rm B}/\zeta$ describes the effect of scattering suppression by diffraction. As an illustration, for $\theta_V\approx 1$ and wavelength of order $\lambda_{\rm B}\sim 100$ nm we obtain the value $\lambda_{\rm B}/\zeta\approx 100$, giving the diffraction-enhanced mean free path as large as $L\sim 10^2\lambda_{\rm B}=10^4$ nm.

We note that localization effects may become important if the disorder is strong enough. In our case, since disorder is mainly at the surface, the behavior is expected to be quite different for electron energies inside and outside the bulk energy gap. In the first case, electron states with energies within the bulk gap reside near the surface. These states couple to surface disorder relatively strongly and can become localized. In the second case the states at the surface will hybridize with the states in the bulk, which suppresses localization due to surface disorder. In addition, as discussed above, the slow decay of electron states from the surface into the bulk gives the surface states a large width that allows electrons to diffract around surface disorder. Such diffraction also suppresses localization. For quasi-1D surface states, such as those in graphene, the 1D mean free path provides a good estimate for localization length at the energies in the bulk gap. For 2D surface states, on the other hand, the localization length is expected to be much longer than the mean free path estimated perturbatively. The latter in this case sets only a lower bound for localization length.

The Effect of Magnetic Field

Experimental detection of surface states by conventional transport techniques can be challenging since the signatures of surface states are often obscured by the continuum of bulk states (the overlap of bulk and surface states contributions to the density of states is illustrated in Fig. 3*D*). Here we consider a different approach relying on the Landau-level spectroscopy in a magnetic field applied perpendicular to the surface. The signatures of Landau levels of the states in a 3D bulk are usually softened by the momentum dispersion in the direction along the field. In contrast, the spectrum of the 2D surface states will be discrete. Therefore, while both the bulk states and the surface states produce Landau levels, the spectral features such as, e.g., the tunneling density of states measured by scanning tunneling microscopy will be dominated by the surface states.

To study the effect of magnetic field, we use a simple model of electrons confined by a 2D delta-function sheet potential of the strength proportional to θ_V (Eq. **18**):

$$H=\sigma_1vp_x+\sigma_2v\,(p_y+eBz)+\tau_2\sigma_3vp_z+\tau_3\sigma_3\Delta+v\theta_V\delta(x). \qquad [22]$$

The states confined near the $x=0$ plane can be found as evanescent solutions for $x>0$ and $x<0$, $\psi\sim\exp(ip_xx+ip_yy-\mu|x|)$ (*SI Appendix, Surface States in Magnetic Field*). For $B=0$, the spectrum of this model coincides with the spectrum of surface states for armchair boundary conditions:

$$\varepsilon_k=-{\rm sgn}(\theta_V)\cos\theta_V\sqrt{v^2k^2+\Delta^2},\qquad k^2=p_y^2+p_z^2. \qquad [23]$$

In a nonzero magnetic field B, we obtain discrete nondispersing levels resembling Landau levels of 2D Dirac particles:

$$\varepsilon_n=-{\rm sgn}(\sin\theta_V)\cos\theta_V\sqrt{2veBn+\Delta^2},\qquad n=0,1,2\ldots. \qquad [24]$$

Interestingly, similar to the $B=0$ solution Eq. **23**, the discrete levels exist only for one sign of energy, positive or negative.

The discrete character of the surface Landau levels as well as their striking lack of particle–hole symmetry provides a direct and simple diagnostic of the surface states. Further evidence can be obtained using the property of surface states to be tunable through changing the surface potential by side gates (Fig. 3 and

Eq. **18**). Due to a periodic dependence on the potential strength, the electron–hole asymmetry can be inverted by reversing the potential sign or by applying a stronger potential.

In conclusion, our key finding is that surface states are a natural attribute of a Dirac band structure, appearing in a robust manner for generic boundary conditions. The surface states feature a number of interesting and potentially useful properties. In particular, we predict that these states are insensitive to surface imperfections: By diffracting around surface disorder electron waves can propagate ballistically with abnormally long mean free path values. These states can coexist with the bulk states or appear within the bulk bandgap; their dispersion can be tuned by gate potential or by magnetic field, giving rise to a range of unique signatures amenable to a variety of experimental probes.

ACKNOWLEDGMENTS. We acknowledge support of the Center for Integrated Quantum Materials under NSF Award DMR-1231319; the MIT Center for Excitonics; the Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Basic Energy Sciences under Award de-sc0001088; and Army Research Office Grant W911NF-18-1-0116.

1. Hasan MZ, Kane CL (2010) Colloquium: Topological insulators. *Rev Mod Phys* 82:3045–3067.
2. Wehling T, Black-Schaffer A, Balatsky A (2014) Dirac materials. *Adv Phys* 63:1–76.
3. Kane CL, Mele EJ (2005) Z_2 topological order and the quantum spin Hall effect. *Phys Rev Lett* 95:146802.
4. König M, et al. (2007) Quantum spin Hall insulator state in HgTe quantum wells. *Science* 318:766–770.
5. Roth A, et al. (2009) Nonlocal transport in the quantum spin Hall state. *Science* 325:294–297.
6. Nowack KC, et al. (2013) Imaging currents in HgTe quantum wells in the quantum spin Hall regime. *Nat Mater* 12:787.
7. Pribiag VS, et al. (2015) Edge-mode superconductivity in a two-dimensional topological insulator. *Nat Nano* 10:593–597.
8. Nichele F, et al. (2016) Edge transport in the trivial phase of InAs/GaSb. *New J Phys* 18:083005.
9. Knez I, et al. (2014) Observation of edge transport in the disordered regime of topologically insulating InAs/GaSb quantum wells. *Phys Rev Lett* 112:026602.
10. Du L, Knez I, Sullivan G, Du RR (2015) Robust helical edge transport in gated InAs/GaSb bilayers. *Phys Rev Lett* 114:096802.
11. Ma EY, et al. (2015) Unexpected edge conduction in mercury telluride quantum wells under broken time-reversal symmetry. *Nat Commun* 6:7252.
12. Allen MT, et al. (2016) Spatially resolved edge currents and guided-wave electronic states in graphene. *Nat Phys* 12:128–133.
13. Allen MT, et al. (2017) Observation of electron coherence and Fabry-Perot standing waves at a graphene edge. *Nano Lett* 17:7380–7386.
14. Shalom MB, et al. (2016) Quantum oscillations of the critical current and high-field superconducting proximity in ballistic graphene. *Nat Phys* 12:318–322.
15. Calado VE, et al. (2015) Ballistic Josephson junctions in edge-contacted graphene. *Nat Nano* 10:761–764.
16. Zhu MJ, et al. (2017) Edge currents shunt the insulating bulk in gapped graphene. *Nat Comm* 8:14552.
17. Jackiw R, Rebbi C (1976) Solitons with fermion number $\frac{1}{2}$. *Phys Rev D* 13:3398–3409.
18. Gold A (1986) Metal insulator transition due to surface roughness scattering in a quantum well. *Solid State Commun* 60:531–534.
19. Kamburov D, Baldwin K, West K, Shayegan M, Pfeiffer L (2016) Interplay between quantum well width and interface roughness for electron transport mobility in GaAs quantum wells. *Appl Phys Lett* 109:232105.
20. Davison S, Steslicka M (1969) Relativistic theory of Tamm surface states. *J Phys C Solid State Phys* 2:1802.
21. Subramanian R, Bhagwat K (1972) The relativistic Tamm model. *J Phys C Solid State Phys* 5:798.
22. Bogolioubov P (1968) Sur un modèle à quarks quasi-indépendants. *Ann Inst Henri Poincare* 8:163–189.
23. Chodos A, Jaffe R, Johnson K, Thorn CB (1974) Baryon structure in the bag theory. *Phys Rev D* 10:2599.
24. Berry MV, Mondragon RJ (1987) Neutrino billiards: Time-reversal symmetry-breaking without magnetic fields. *Proc R Soc A* 412:53–74.
25. Jaffe RL, Manohar A (1989) Bound states of the Dirac equation outside a hard sphere. *Ann Phys* 192:321–330.
26. McCann E, Fal'ko VI (2004) Symmetry of boundary conditions of the Dirac equation for electrons in carbon nanotubes. *J Phys Condens Matter* 16:2371.
27. Akhmerov AR, Beenakker CWJ (2008) Boundary conditions for Dirac fermions on a terminated honeycomb lattice. *Phys Rev B* 77:085423.
28. Yao W, Yang SA, Niu Q (2009) Edge states in graphene: From gapped flat-band to gapless chiral modes. *Phys Rev Lett* 102:096801.
29. Latyshev YI, et al. (2014) Transport of massless Dirac fermions in non-topological type edge states. *Sci Rep* 4:7578.
30. Leykam D, Bliokh KY, Huang C, Chong YD, Nori F (2017) Edge modes, degeneracies, and topological numbers in non-Hermitian systems. *Phys Rev Lett* 118:040401.
31. Volkov V, Enaldiev V (2016) Surface states of a system of Dirac fermions: A minimal model. *J Exp Theor Phys* 122:608–620.
32. Kharitonov M, Mayer JB, Hankiewicz EM (2017) Universality and stability of the edge states of chiral-symmetric topological semimetals and surface states of the Luttinger semimetal. *Phys Rev Lett* 119:266402.
33. Nakada K, Fujita M, Dresselhaus G, Dresselhaus MS (1996) Edge state in graphene ribbons: Nanometer size effect and edge shape dependence. *Phys Rev B* 54:17954–17961.
34. Koskinen P, Malola S, Häkkinen H (2009) Evidence for graphene edges beyond zigzag and armchair. *Phys Rev B* 80:073401.
35. Kobayashi Y, Fukui Ki, Enoki T, Kusakabe K (2006) Edge state on hydrogen-terminated graphite edges investigated by scanning tunneling microscopy. *Phys Rev B* 73:125415.
36. Park C, et al. (2011) Formation of unconventional standing waves at graphene edges by valley mixing and pseudospin rotation. *Proc Natl Acad Sci USA* 108:18622–18625.
37. Unuma T, Yoshita M, Noda T, Sakaki H, Akiyama H (2003) Intersubband absorption linewidth in GaAs quantum wells due to scattering by interface roughness, phonons, alloy disorder, and impurities. *J Appl Phys* 93:1586–1597.