Revisiting the physical origin and nature of surface states in inverted-band semiconductors

Alexander Khaetskii,[1,*] Vitaly Golovach [ID],[2,3,4,5] and Arnold Kiefer [ID][1]

[1]*Air Force Research Laboratory, Wright-Patterson AFB, Ohio 45433, USA*
[2]*Centro de Física de Materiales (CFM-MPC), Centro Mixto CSIC-UPV/EHU, 20018 Donostia-San Sebastián, Spain*
[3]*Departamento de Polímeros y Materiales Avanzados: Física, Química y Tecnología, Facultad de Química,*
University of the Basque Country UPV/EHU, 20080 Donostia-San Sebastián, Spain
[4]*Donostia International Physics Center (DIPC), 20018 Donostia-San Sebastián, Spain*
[5]*IKERBASQUE, Basque Foundation for Science, 48013 Bilbao, Spain*

We revisit the problem of surface states in semiconductors with inverted-band structures, such as α-Sn and HgTe. We unravel the confusion that arose over the past decade regarding the origin of the surface states, their topological nature, and the role of strain. Within a single minimalistic description, we reconcile different solutions found in the 1980s with the results obtained from modern-day numerical simulations, allowing us to unambiguously identify all branches of surface states around the Γ point of the Brillouin zone in different regimes. We also show that strain is a smooth deformation to the surface states, following the usual continuity principle of physics, and not leading to any drastic change of the physical properties in these materials, in contrast to what has recently been advanced in the literature. We consider biaxial in-plane strain that is either tensile or compressive, leading to different branches of surface states for topological insulators and Dirac semimetals, respectively. Our model can help in interpreting numerous experiments on topological surface states originating from inverted-band semiconductors.

I. INTRODUCTION

Electronic surface states of solids have long been a topic of interest because of their importance in practical physical devices. The origin and properties of surface states for many semiconductor solids have been explained in great detail over several decades, beginning with Tamm and Shockley [1,2]. More recently, the prediction and discovery of topological insulators and their surface states have revived interest in this topic owing to their special properties [3,4]. Present-day researchers often generate models of surface bands with sophisticated computational software, but they do not usually provide insight into the physical origin of these surface bands, often leading to a misguided treatment. Furthermore, recent works seem unaware of the foundational analytical work conducted decades ago that provide similar physical predictions about surface states though at that time not recognizing them as having a specific topological character [5–9].

Topological insulators (TIs) serve as an excellent bridge between traditional electronic band structure theory and a more modern approach including topological aspects. TIs emerge from a class of narrow-gap materials whose strong spin-orbit coupling leads to an inverted-band structure and the formation of helical (i.e., spin-momentum locked) states on their surfaces [5–12] that are protected from nonmagnetic perturbations. Growing attention to topological insulators is

fueled by a fundamental interest in solid-state spin physics and the prospect of designing novel electronic devices [13]. Topological surface states have been observed in a number of three-dimensional (3D) compounds, such as Bi_2Se_3, Bi_2Te_3, etc. [14,15]. Examples of 2D TIs with 1D helical channels at the sample edges include HgTe/CdHgTe [16,17] and InAs/GaSb [18,19] quantum wells of certain thicknesses. Experimentally, surface states in 3D topological insulators are usually revealed by the angle- and spin-resolved photoemission spectroscopy [15,20] and magnetotransport measurements [21–27].

Strain is frequently present in real films and plays a role in revealing topological states by breaking crystallographic symmetries. Several attempts to reveal experimentally the surface states in strained HgTe films (3D topological insulators) were made in magnetotransport experiments [21,22]. It is interesting that the theoretical interpretation of the experimental data in these two works are based on very different physical pictures of the surface states in HgTe material. In particular, the surface states in Ref. [22] are treated as linear in k-vector Dirac states that exist in the gap between conduction s band and light-hole p band when the presence of the heavy-hole band is totally ignored. In other words, the topological surface states and bulk heavy-hole states are considered as completely independent. In an opposing treatment offered in Ref. [21], strong hybridization between surface states and bulk heavy-hole states makes the physical picture totally different, where only the surface states lying within the strain-created gap between the light-hole and heavy-hole bands have physical meaning. Such a difference in interpretation requires further
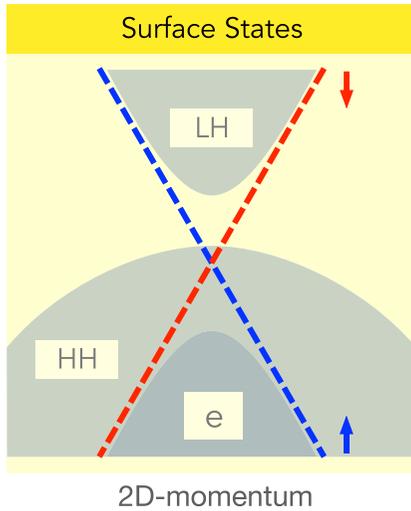
---

*khaetskii@gmail.com

035305-1

FIG. 1. Incorrect picture of surface states obtained after neglecting the coupling to the heavy-hole (HH) band. The surface states are formed in a two-band model consisting of electrons (e) and light holes (LH) in an inverted-band semiconductor.

FIG. 2. Generic band structure variation across an interface between an inverted-band-gap semiconductor (SC) (left: $E_g < 0$) and a SC with a regular band order (right: $E_g > 0$). A tensile in-plane strain applied to the inverted-band-gap SC induces a splitting between the light-hole and heavy-hole states and opens a gap Δ at the Γ point. At certain energies (not shown), the interface supports surface states with a probability density which exponentially decays into both regions, thus being localized at the interface (bottom).

examination. As we will explain in this work, we rather support the physical description of the surface states used in Ref. [21].

Another material attracting much attention now is α-Sn, which has a band structure similar to HgTe. Recently published research concerning surface states in α-Sn (see, for example, Refs. [28–34]) clearly demonstrate the confusion that exists in the community about the role of band structure and strain. Many authors who treat the surface states by various numerical methods often ignore the presence of the heavy-hole band and solve the problem with only the light-hole and electron s-type bands. As we show below, this leads to an incorrect physical picture of the surface states, see Fig. 1. Moreover, application of even a small strain in their solutions causes a sudden appearence of the Dirac surface states deeply below the degeneracy point (Γ_8) of the light and heavy holes, i.e., within the $\Gamma_8 - \Gamma_6$ gap. This manifests the unphysical character of the obtained solutions. (We mean here a tensile in-plane strain applied to the inverted-band-gap semiconductor that opens a gap at the Γ point between the light- and heavy-hole bands.) The troubles outlined above lead, in turn, to difficulties interpreting the experimental data obtained with the surface states in the materials such as HgTe and α-Sn.

The goal of this work is to clarify the roles of band structure, strain, and spin in constructing a physical picture of surface states consistent with conventional solid-state theory. The authors of many results obtained solely by numerical calculation seem completely unaware of previous analytical works that treated surface states in similar materials (such as HgTe) decades ago [5–9], making them difficult to relate. We seek to reconcile these differences and show below that those older works contain all the essential physics and treat the band structure of the materials correctly while still describing the surface states completely. In particular, we describe the intimate relation of the surface states in HgTe gapless
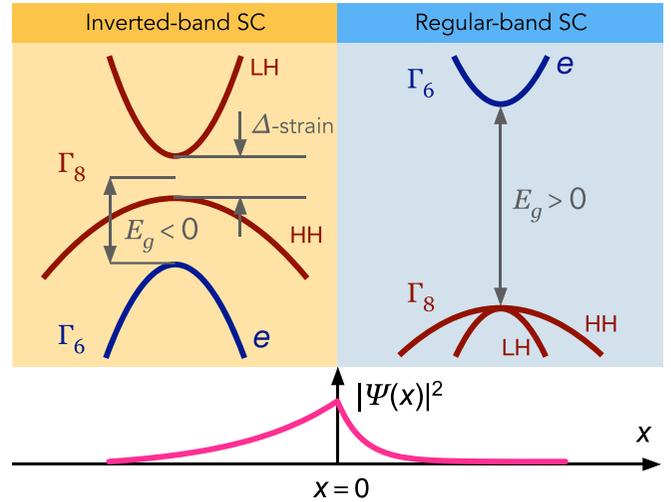
semiconductor first predicted within the Luttinger model in Ref. [5] and surface states obtained within the more general Kane model that also takes the s band into account. We demonstrate that these Dyakonov-Khaetskii (DK) states [5] possess all the essential features of the topological states in the presence of strain, which induces a gap between the light- and heavy-hole bands (topological insulator regime). As such, they should play the most important role in the experiments that probe the topological properties of the sample when the Fermi level is located within this strain-induced gap. Moreover, we show that strain does not drastically change the properties of surface states that form the Dirac cone within the $\Gamma_8 - \Gamma_6$ gap, which is again in strong contrast to the results advanced recently in the literature.

We take a pedagogical approach, beginning with a simple, idealized model that captures the essential physics. As such we consider a variant of the Kane k-P model with infinite mass of heavy holes in the case of a single abrupt interface between an inverted material such as HgTe and a direct material such as CdTe or vacuum. This model takes into account three bands: an s-type conduction band and p-type light- and heavy-hole bands, see Fig. 2. The split-off band is relatively far away in energy (strong spin-orbit coupling) and does not play a significant role in forming the surface states. By considering the flat heavy-hole band we deal with the case of a real gap between hole bands (Γ_8) and electron s band (Γ_6) when the density of states within this gap is zero. These features of the model make the physics of the surface states located in the $\Gamma_8 - \Gamma_6$ gap very simple and reveals their essential features. We then address the realistic case of finite heavy-hole mass and show the surface states located within the $\Gamma_8 - \Gamma_6$ gap overlap the bulk heavy-hole states in energy and are strongly

hybridized with them, making those surface states marginal in observable phenomena.

For a detailed description of the surface states that exist within the projected gap between the light- and heavy-hole bands (DK states) we use the Luttinger model, which is applicable for the case of arbitrary masses of light and heavy holes. We were able to find exact analytical solutions for the surface states within this model for both signs of the strain energy Δ, covering the TI regime ($\Delta > 0$) and the Dirac semimetal regime ($\Delta < 0$). In the latter case, which is achieved by applying a compressive in-plane strain (hence tensile in the normal direction), an intersection of the light-hole and heavy-hole bands occurs forming two Dirac points with linear bulk dispersion nearby. This Dirac semimetal phase has been attracting a lot of attention recently both theoretically and experimentally, see Refs. [33,34]. We show that the dispersion of surface states, which arise around the conic point of the bulk projected states crucially depends on the applied strain, in contrast to the TI regime.

Our simple models allow us to trace the origins of the incorrect physical picture of the surface states obtained in many recent publications, namely, neglecting the coupling to the heavy-hole bands that leads directly to the simple picture of linear in k-vector Dirac surface states crossing the whole gap between the conduction and light-hole bands, see Fig. 1. We stress again that this latter picture does not reflect the true nature of the surface states.

II. 3 × 3 KANE MODEL: AN ABRUPT INTERFACE OF TWO MATERIALS WITH INVERTED AND DIRECT BANDS

We consider here the simplest possible problem: the interface between two materials, one of them having an inverted band structure and the other a direct one. Our description is quite general and is applicable for many materials, but as a concrete example one can consider the (HgTe/CdTe) case. Additionally, we assume a biaxial tensile in-plane strain applied to the inverted band material as is often accomplished in practice by epitaxial growth of a thin film on a much thicker substrate, see Fig. 2. The band structures are described by the 6 × 6 Kane model that takes into account three bands, electrons (e) (s-symmetry band), and two p bands: light (lh) and heavy holes (hh). To simplify the treatment, we can from the very beginning use the time-reversal character of the problem and reduce the model to 3 × 3 by choosing the proper coordinate system [35]. We choose the angular momentum quantization axis z in the plane of the interface and the direction of the carrier motion along y; x denotes the direction normal to the interface. Then the two groups of states (e 1/2, lh −1/2, hh 3/2) and (e −1/2, lh 1/2, hh −3/2) do not mix. The Hamiltonian matrix for the first group of states is

$$\hat{H} = \begin{bmatrix} \epsilon_{c1,2} & Pk_-/\sqrt{6} & Pk_+/\sqrt{2} \\ Pk_+/\sqrt{6} & \epsilon_{v1,2} - \Delta/4 & \sqrt{3}\Delta/4 \\ Pk_-/\sqrt{2} & \sqrt{3}\Delta/4 & \epsilon_{v1,2} + \Delta/4 \end{bmatrix}. \quad (1)$$

[The Hamiltonian matrix that refers to the second group of states with the opposite sign of the angular momentum projection on the z axis can be obtained from Eq. (1) by replacing k_y by $-k_y$.] In Eq. (1) P is the Kane matrix element,

which we consider coordinate independent, $\hat{k}_\pm = \hat{k}_x \pm ik_y$, $\hat{k}_x = -id/dx$, $\epsilon_{c1,2}, \epsilon_{v1,2}$ are conduction- and valence-band center energies in inverted-band semiconductor I and in semiconductor II with direct bands, respectively. We assume that HgTe layer is strained due to the lattice mismatch between CdTe and HgTe and this tensile in-plane strain opens the gap $\Delta > 0$ between the light and heavy-hole bands. If the CdTe is oriented along (001), then the strain tensor components are given by $\epsilon_{yy} = \epsilon_{zz} = (a_{CdTe} - a_{HgTe})/a_{HgTe} > 0$ (tensile in-plane strain), $\epsilon_{xx} < 0$; the nondiagonal components are equal to zero. Here a_{CdTe} and a_{HgTe} are the lattice constants of the corresponding materials. The effects of the strain tensor are incorporated in the Kane model through the Bir-Pikus Hamiltonian [36], which is easily obtained from the Kane Hamiltonian with the substitution $k_ik_j \rightarrow \epsilon_{ij}$. The value of the strain gap in our case is given by $\Delta = 2b(\epsilon_{xx} - \epsilon_{yy}) > 0$, where b is the uniaxial deformation potential ($b = -1.5$ eV for HgTe).

With the proper unitary transformation one can bring the Hamiltonian Eq. (1) to the following form:

$$\hat{H} = \begin{bmatrix} \epsilon_{c1,2} & \frac{(2\hat{k}_x + ik_y)P}{\sqrt{6}} & iPk_y/\sqrt{2} \\ \frac{(2\hat{k}_x - ik_y)P}{\sqrt{6}} & \epsilon_{v1,2} + \Delta/2 & 0 \\ -iPk_y/\sqrt{2} & 0 & \epsilon_{v1,2} - \Delta/2 \end{bmatrix}. \quad (2)$$

A. No coupling to the heavy-hole band

Before proceeding further we can make here the following important note, which explains the origin of the wrong physical picture of the surface states obtained in Refs. [28–31]. If one neglects altogether the interaction of the Γ_6 band with the heavy-hole band, given by elements H_{13} and H_{31} in Eq. (2), then one obtains the anisotropic Dirac model, which describes the interaction of the two bands, electrons and light holes, as described by the 2 × 2 block in the top left corner of the matrix in Eq. (2). Let us find the energy dispersion of the surface states for this reduced case, considering semi-infinite sample of HgTe located at $x < 0$ and CdTe sample located at $x > 0$. The wave functions decay in the left-hand region as $\exp(\kappa_1 x)$ and in the right-hand region as $\exp(-\kappa_2 x)$ (i.e., $k_{x1} = -i\kappa_1$, $k_{x2} = i\kappa_2$, with positive $\kappa_{1,2}$); the dependence along the y coordinate is given by $\exp(ik_y y)$. For simplicity we consider here the symmetric case $\Delta = 0$, $\epsilon_{v1} - \epsilon_{c1} = \epsilon_{c2} - \epsilon_{v2} = \epsilon_g > 0$. Then from the bulk energy dispersion $\epsilon_b^2 = \epsilon_g^2/4 + (P^2/6)(k_y^2 + 4k_x^2)$ we easily find $\kappa_1 = \kappa_2 = \kappa = (1/2)\sqrt{k_y^2 + (6/P^2)(\epsilon_g^2/4 - \epsilon^2)}$, where ϵ is the energy of the surface state obtained from the boundary conditions. As the boundary conditions we use the continuity of both components of the two-component spinors found in regions I and II. For our symmetric case it gives the following equation: $-4\kappa\epsilon = k_y\epsilon_g$. Using the expression for the κ, we finally obtain the equation for the surface states energy

$$\epsilon\sqrt{k_y^2 + \frac{6}{P^2}\left(\frac{\epsilon_g^2}{4} - \epsilon^2\right)} = -\frac{k_y}{2}\epsilon_g. \quad (3)$$

This equation has the exact solution $\epsilon = -Pk_y/\sqrt{6}$. (The second branch of surface states is obtained by changing the sign of k_y.) Note that the surface curve $\epsilon = Pk_y/\sqrt{6}$ asymptotically

approaches the bulk light-hole dispersion curve at large k_y and $k_x = 0$. Thus we see that the Dirac solution for the surface states is obtained by total neglect of the interaction with the heavy-hole band.

B. Full treatment

Let us now find the correct energy spectrum of the surface states for the same problem with heavy-hole interactions included. As it follows from Hamiltonian (2), an interaction of the Γ_6 band with the heavy-hole band is given by the $ik_y P/\sqrt{2}$ element, which does not depend on d/dx. For that reason we can exclude the heavy-hole component of the wave function and still work with the two-band model, i.e., electrons and light holes. It is easy to work out from Eq. (2) that this procedure leads to the system of two equations in each region: $\hat{H}_{\text{eff}} \, \chi = 0$, $\chi_{tr} = (u, v)$ with the effective Hamiltonian (see also Ref. [9]):

$$\hat{H}_{\text{eff}} = \begin{bmatrix} \epsilon_{c1,2} - \epsilon - \frac{P^2 k_y^2}{2(\epsilon_{v1,2} - \epsilon - \Delta/2)} & \frac{(2\hat{k}_x + ik_y)P}{\sqrt{6}} \\ \frac{(2\hat{k}_x - ik_y)P}{\sqrt{6}} & \epsilon_{v1,2} - \epsilon + \Delta/2 \end{bmatrix}. \quad (4)$$

We assume that strain is zero in region II and Δ is positive in region I, i.e., a gap is open between the light- and heavy-hole bands (see Fig. 2). From Eq. (4) we find the equation for the energy dispersions of the bulk materials:

$$\left[\epsilon_{v1,2} - \epsilon - \frac{\Delta}{2} \right] \cdot \left[(\epsilon_{c1,2} - \epsilon)\left(\epsilon_{v1,2} - \epsilon + \frac{\Delta}{2} \right) \right.$$
$$\left. - \frac{2}{3} P^2 \left(k_x^2 + k_y^2 \right) \right] = P^2 k_y^2 \frac{\Delta}{2}. \quad (5)$$

We see that in the absence of strain the spectrum of heavy holes with infinite mass is simple: $\epsilon_h = \epsilon_{v1,2}$. Strain mixes the light and heavy holes, the heavy holes acquire finite mass in the y direction (within the plane), and the spectrum becomes complicated. In the x direction, however, when $k_y = 0$, the spectra of light and heavy particles remain simple:

$$(\epsilon_{c1,2} - \epsilon(k_x))\left(\epsilon_{v1,2} - \epsilon(k_x) + \frac{\Delta}{2} \right) = \frac{2}{3} P^2 k_x^2 \quad (6)$$

$$\epsilon_{h1,2}(k_x) = \epsilon_{v1,2} - \frac{\Delta}{2}. \quad (7)$$

Equation (6) gives the bulk spectrum of the light particles—electrons and light holes with dispersions $\epsilon_e(k_x)$ and $\epsilon_l(k_x)$, respectively.

We seek the surface state that is localized at the interface ($x = 0$) and decays in the left-hand region as $\exp(\kappa_1 x)$ and in the right-hand region as $\exp(-\kappa_2 x)(\kappa_{1,2} > 0)$. The corresponding expressions for the $\kappa_{1,2}$ can be found from Eq. (5):

$$\kappa_1^2 = k_y^2 + \frac{3(\epsilon - \epsilon_{c1})\left(\epsilon_{v1} - \epsilon + \frac{\Delta}{2} \right)}{2P^2} + \frac{3\Delta k_y^2}{4\left(\epsilon_{v1} - \epsilon - \frac{\Delta}{2} \right)}; \quad (8)$$

$$\kappa_2^2 = k_y^2 + \frac{3(\epsilon_{c2} - \epsilon)(\epsilon - \epsilon_{v2})}{2P^2}. \quad (9)$$

The spinors in both regions (I) and (II) can be easily found with the help of Eq. (4); then from the continuity conditions of both components of the wave functions at $x = 0$ we find:

$$\left(\epsilon_{v1} - \epsilon + \frac{\Delta}{2} \right)(2\kappa_2 - k_y) = (\epsilon - \epsilon_{v2})(2\kappa_1 + k_y). \quad (10)$$

The surface states spectra are determined by the solution of the system of equations (8), (9), (10).

1. Symmetric case, $\Delta = 0$, $|\epsilon_{g1}| = \epsilon_{g2} = \epsilon_g$

The discussion of the solution for the surface states within the full treatment starts with the symmetric case [Fig. 3(a)] $\Delta = 0$, $\epsilon_{v1} = -\epsilon_{c1} = \epsilon_{c2} = -\epsilon_{v2} = \epsilon_g/2 > 0$ so we may compare it with the results obtained in Sec. II A. In this case we have $\kappa_1 = \kappa_2 = \kappa = \sqrt{k_y^2 + (3/2P^2)(\epsilon_g^2/4 - \epsilon^2)}$, and from Eq. (10) one has the same equation as before, $-4\kappa\epsilon = k_y\epsilon_g$, therefore finally we obtain

$$\epsilon\sqrt{4k_y^2 + \frac{6}{P^2}\left(\frac{\epsilon_g^2}{4} - \epsilon^2 \right)} = -\frac{k_y}{2}\epsilon_g. \quad (11)$$

[The time-reversed solutions can be obtained by changing the sign of k_y in Eq. (11).] Despite the fact that Eq. (3) and Eq. (11) are nearly the same (except for the coefficient of the first term under the radical), they have completely different solutions, see Fig. 3(b). For $|k_y| \ll \epsilon_g/P$ Eq. (11) has the same behavior: $\epsilon = -Pk_y/\sqrt{6}$; however, at large $|k_y| \gg \epsilon_g/P$ the surface branch saturates at $+\epsilon_g/4$ (at negative k_y) and at $-\epsilon_g/4$ (at positive k_y). Similar behavior was found in Refs. [7–9]. This behavior is due to an interaction with the heavy-hole band, which has the character of repulsion.

Besides the surface states that lie within the fundamental gap, Eq. (11) has the solutions which start from the heavy-hole bands ($\epsilon = \epsilon_g/2$, $\epsilon = -\epsilon_g/2$), see Fig. 3(b). For the first of those additional solutions, for example, we obtain at small $|k_y|P \ll \epsilon_g$

$$\epsilon_{s1} = \frac{\epsilon_g}{2} + \frac{P^2 k_y^2}{2\epsilon_g}. \quad (12)$$

This solution follows from Eq. (11) at $k_y < 0$. [For $k_y > 0$ the solution with this energy does not exist since the left-hand side of Eq. (11) is strictly positive.] The same energy dispersion for $k_y > 0$ follows from the time-reversed Hamiltonian and corresponds to the opposite spin direction. It means that there is a momentum-spin locking for these solutions, and back scattering is strictly forbidden. We will discuss this issue in more detail below. The effective mass of this surface state, Eq. (12), is by the factor of 4/3 larger than the effective mass of the light hole m_l. [The latter is found from the bulk dispersion equation for the light holes: $\epsilon_l(k_y) = \sqrt{\epsilon_g^2/4 + (2/3)P^2 k_y^2}$.] This new surface branch was predicted theoretically for HgTe for the first time in Ref. [5]. The effective mass of this surface state depends on the ratio β of the bulk effective masses of light and heavy holes and approaches the value $(4/3)m_l$ (see Ref. [5]) in the limit $\beta \to 0$ considered here. As it follows from Eq. (11) (and from the time-reversed one), at large $|k_y| \gg \epsilon_g/P$ the surface branch in question asymptotically approaches the bulk light-hole band: $\epsilon_{s1} \to \sqrt{2/3}P|k_y| + (3\sqrt{3}/32\sqrt{2})(\epsilon_g^2/P|k_y|)$.

2. Evolution of surface states with the strength of interaction between Γ_6 and heavy-hole band

To better understand the origin of different branches of surface states presented in Fig. 3(b), it is instructive to follow
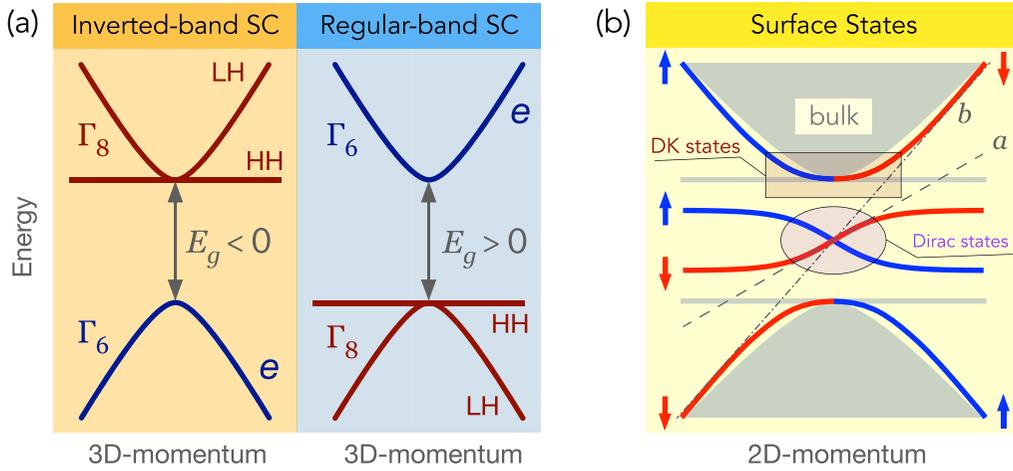
FIG. 3. (a) An interface with an antisymmetric variation of the bulk band structure. For simplicity, the strain is omitted here and the heavy-hole mass is set to infinity. (b) The surface states arising from the band alignment in (a), showing two types of states near the center of the Brillouin zone: linear Dirac states and quadratic Dyakonov-Khaetskii (DK) states. The projected bulk spectrum is shown by the shaded gray regions. All surface states are singly degenerate with a distinct spinor structure originating from one of the two time-reversal-related blocks of the Hamiltonian as shown by the spin-up and spin-down arrows. The asymptote of the Dirac states at small in-plane momenta (dashed line *a*) differs strongly from the asymptote of the DK states at large momenta (dot-and-dashed line *b*), refuting the picture of a narrow-in-energy interruption of the Dirac states by the heavy-hole band.

the evolution of these states as a function of the strength of the interaction of the Γ_6 band with the heavy-hole band. To this end we multiply the corresponding matrix elements in Eq. (2) by the factor α, where α can take the values in the interval between 0 and 1, obtaining the Hamiltonian

$$\hat{H}_\alpha = \begin{bmatrix} \epsilon_{c1,2} & \frac{(2\hat{k}_x + ik_y)P}{\sqrt{6}} & i\alpha Pk_y/\sqrt{2} \\ \frac{(2\hat{k}_x - ik_y)P}{\sqrt{6}} & \epsilon_{v1,2} + \Delta/2 & 0 \\ -i\alpha Pk_y/\sqrt{2} & 0 & \epsilon_{v1,2} - \Delta/2 \end{bmatrix}. \quad (13)$$

Then for the symmetric case we obtain the following equation, which describes the surface energy spectrum:

$$\epsilon\sqrt{(1+3\alpha^2)k_y^2 + \frac{6}{P^2}\left(\frac{\epsilon_g^2}{4} - \epsilon^2\right)} = -\frac{k_y}{2}\epsilon_g. \quad (14)$$

This equation reproduces Eq. (3) and Eq. (11) in the limits $\alpha \to 0$ and $\alpha \to 1$, correspondingly. It is useful to take the square of this equation to obtain

$$\left(\frac{\epsilon_g^2}{4} - \epsilon^2\right)\left(k_y^2 - \frac{6\epsilon^2}{P^2}\right) = 3\alpha^2 k_y^2 \epsilon^2. \quad (15)$$

Let us investigate the solutions in the $\alpha \to 0$ limit. There are two of them with the anticrossing point $\tilde{k}_y = \sqrt{3/2}(\epsilon_g/P)$, see Fig. 4(b). The first solution has the dispersion $\epsilon = \epsilon_g/2 + (\alpha^2 P^2 k_y^2/2\epsilon_g)$ at small k_y. [Note that this dispersion coincides with the one given by Eq. (12) at $\alpha = 1$.] At large k_y this branch is close to the Dirac solution $Pk_y/\sqrt{6}$. The second solution coincides with the Dirac one at small k_y and is close to the heavy-hole band ($\epsilon_g/2$) at large k_y. These two branches repel each other with increasing α and eventually approach the configuration shown in Fig. 3(b) (red curves). For example, the asymptote of the DK state at large k_y is $Pk_y\sqrt{(1+3\alpha^2)/6}|_{\alpha=1} = Pk_y\sqrt{2/3}$ [asymptote *b* in Fig. 3(b)].

Thus we see that the surface states presented in Fig. 3 (see also Fig. 5 below) are intimately related to the Dirac states. In particular, DK surface state [5] predicted within the Luttinger model is the result of strong hybridization of the Dirac state with the heavy-hole states (see also Ref. [37]). The authors of Ref. [21] claimed the same, though they referred to the states only as special rather than identify them as the DK states previously established in Ref. [5].
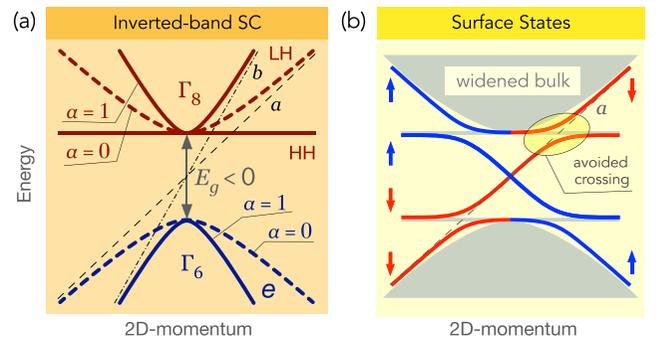

FIG. 4. Dirac cone influenced by a weak coupling to the heavy-hole band. (a) The widening of the light-hole (LH) and electron (e) branches in the in-plane momentum direction upon weakening of the coupling to the heavy-hole branch. The coupling to the heavy-hole branch is full for $\alpha = 1$ (solid lines), corresponding to a pristine semiconductor, and absent for $\alpha = 0$ (dashed lines), when dispensing with the heavy-hole branch. The asymptotes *a* and *b* delimit the bands for $\alpha = 0$ and 1, respectively. (b) The surface states obtained in the limit $\alpha \ll 1$, showing an avoided crossing of the Dirac cone with the heavy-hole bands. Here, we consider the symmetric case of Fig. 3 for a small α introduced alike for the inverted-band-gap and regular semiconductors.
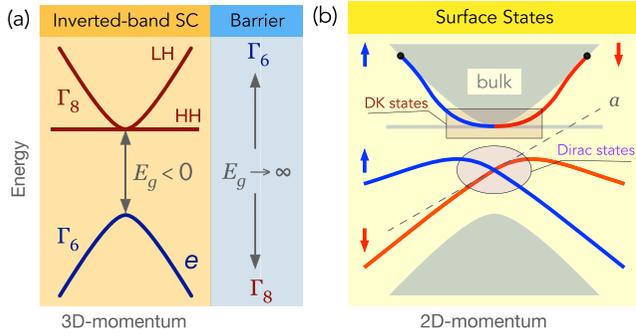
FIG. 5. Effect of a high barrier on the surface states. (a) An interface between an inverted-band-gap SC ($E_g < 0$) and a regular SC with a large fundamental gap ($E_g \to +\infty$). (b) Similar to Fig. 3, two types of surface states occur in the limit of small in-plane momenta, namely the linear Dirac states and the quadratic DK states. At large momenta, the lower cone of the Dirac states becomes sharper than the original low-momentum asymptote a, whereas the upper cone reverses slope and becomes blunter. The DK states terminate at a critical value of momentum $k_y \sim \epsilon_g/P$ (see solid circles), at which the localization length diverges ($\kappa_1 \to 0$) and the surface state flows tangentially into the bulk continuum.

3. High barrier between inverted and direct materials

The surface states spectra depend strongly on the height of the barrier between inverted and direct materials. If the fundamental gap in the direct material ϵ_{g2} is much larger than the modulus of the gap in the inverted one, then Eq. (10) (and time-reversed one) reads

$$\frac{\epsilon_{g1}}{2} - \epsilon - \sqrt{\left(\frac{\epsilon_{g1}}{2}\right)^2 - \epsilon^2 + \frac{2}{3}P^2k_y^2} = \pm\frac{Pk_y}{\sqrt{6}}. \quad (16)$$

As before, we first consider here $\Delta = 0$ case and symmetric energy diagram: $\epsilon_{v1} = -\epsilon_{c1} = \epsilon_{g1}/2 > 0; \epsilon_{c2} = -\epsilon_{v2} = \epsilon_{g2}/2 \to \infty$. Equation (16) can be easily solved analytically and surface curves are presented in Fig. 5. We see that the surface branches, which pass through the center of the gap (Dirac point) do not saturate at large $|k_y|$ but go downward. This happens because the heavy-hole band in the barrier region is located deeply below in energy, which strongly suppresses repulsion from it. These two surface branches have the following exact dispersions:

$$\epsilon_{\downarrow} = \frac{\epsilon_{g1}}{4} + \frac{Pk_y}{2\sqrt{6}} - \sqrt{\left(\frac{\epsilon_{g1}}{4}\right)^2 - \frac{Pk_y\epsilon_{g1}}{4\sqrt{6}} + \frac{7}{24}P^2k_y^2}$$

$$\epsilon_{\uparrow} = \frac{\epsilon_{g1}}{4} - \frac{Pk_y}{2\sqrt{6}} - \sqrt{\left(\frac{\epsilon_{g1}}{4}\right)^2 + \frac{Pk_y\epsilon_{g1}}{4\sqrt{6}} + \frac{7}{24}P^2k_y^2}. \quad (17)$$

As for the DK surface branch, it has the quadratic dispersion for small $k_yP \ll \epsilon_{g1}$ with the same effective mass as it was for the case of equal gaps of direct and inverted materials, see Eq. (12). The behavior of this branch at larger k_y can be traced with the exact solution of Eq. (16) (for positive k_y we take the minus sign in the right-hand side of this equation). This

solution reads

$$\epsilon_{s1} = \frac{\epsilon_{g1}}{4} + \frac{Pk_y}{2\sqrt{6}} + \sqrt{\left(\frac{\epsilon_{g1}}{4}\right)^2 - \frac{Pk_y\epsilon_{g1}}{4\sqrt{6}} + \frac{7}{24}P^2k_y^2}; \quad k_y \geqslant 0.$$

$$(18)$$

It can be easily checked from Eq. (16) that this solution is only valid at $Pk_y \leqslant \sqrt{2/3}\epsilon_{g1}$. At this critical k_y the DK surface state merges with the light-hole band at the energy value, which is equal to $(5/6)\epsilon_{g1}$, see Fig. 5(b).

4. Finite effective mass of the heavy-hole band

So far we have been considering the case when an effective mass of the heavy-hole (HH) band is infinite. It is important to realize that in this limit there is a true gap (i.e., density of states is zero) between the valence Γ_8 bands and conduction Γ_6 band, Figs. 3, 5. When effective mass of heavy holes is finite, their energy starts to depend on the k_x-wave vector component perpendicular to the interface. This means that the surface states located within the $\Gamma_8 - \Gamma_6$ gap lie on top of the continuum of the heavy-hole states corresponding to different k_x and strong hybridization between them and surface states will occur. As a result, the electrons occupying the surface states located within the $\Gamma_8 - \Gamma_6$ gap will not show in transport measurements the properties exclusive for topological surface states not hybridized with any bulk states. Thus, they cannot be distinguished as a separate group of carriers. (We should mention, however, that at small k_y, at which the overlap of these states with heavy holes is weak, they can be detected with the ARPES experiments, see Refs. [31,38].) On the other hand, the surface states lying in the projected gap between the light-hole and heavy-hole bands found in Ref. [5] have true physical meaning. (The similar ideas about the hybridization and its effect on the surface states were developed in Ref. [21].) Now we will study how these states are transformed in the presence of the strain, which opens a true gap between the light-hole and heavy-hole bands.

C. Strain applied to the inverted region

We now analyze the same surface-states problem with strain applied to the inverted region, thereby opening the gap Δ between light-hole and heavy-hole bands, see Fig. 2. We consider the case $\Delta \ll \epsilon_{g1}$, which is of practical importance, and will assume first the symmetric case, $\epsilon_{v1} = -\epsilon_{c1} = \epsilon_{c2} = -\epsilon_{v2} = \epsilon_g/2 > 0$. Clearly, the physical picture of the states lying within a $\Gamma_8 - \Gamma_6$ gap cannot change appreciably with application of strain, see Figs. (3), (5). Indeed, the perturbation theory with respect to the parameter $\Delta/\epsilon_g \ll 1$ shows that the dispersion curves within the $\Gamma_8 - \Gamma_6$ gap shift only slightly (for example, the Dirac point shifts from the middle of the gap by the value of the order of Δ). The only strong change of the spectrum occurs in the energy interval of the order of Δ near the top of the valence band of HgTe. Strain opens a gap; therefore, when the Fermi level is located within this gap, there is no conductivity (at low temperature) in the bulk and no elastic scattering between surface and bulk states. We stress, however, that the topological character of this surface state, Eq. (12), exists even in the absence of the strain (as we have already mentioned above). In particular, all the characteristic
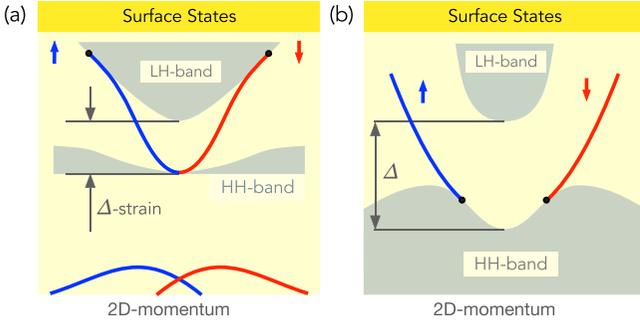
FIG. 6. Effect of a tensile in-plane strain on the surface states studied in two models. (a) Kane model with heavy-hole flat band: The strain splits and intermixes the bulk light-hole and heavy-hole bands, opening a global gap of magnitude $\Delta/4$ and inducing a dispersion in the heavy-hole band. The DK states depart from the heavy-hole band at zero momentum and merge with the light-hole band at large momenta. (b) Luttinger model (finite heavy-hole mass but no conduction bands): Qualitatively, the same result as in (a), except that the DK states depart from the heavy-hole band at finite momenta.

features of the topological state, such as a spin-momentum locking and absence of the back scattering, exist already for the solution, Eq. (12), obtained without strain.

Let us find the energy spectrum of the DK surface state in the presence of strain. From the system of equations (8), (9), (10) it is easy to find that at $k_y \to 0$ the solution has the form $\epsilon = \epsilon_{v1} - \Delta/2 + \delta P^2 k_y^2$ with $\kappa_1^2 = (3\Delta/2P^2)[\epsilon_g - \Delta/2 - 1/(2\delta)]$ and $\kappa_2^2 = (3\Delta/4P^2)(\epsilon_g - \Delta/2)$. Then from Eq. (10) we can calculate δ and, finally, the surface-state spectrum at $k_y \to 0$ reads

$$\epsilon = \epsilon_{v1} - \frac{\Delta}{2} + \frac{P^2 k_y^2}{2\epsilon_g - \Delta - \frac{2\Delta^2}{(2\epsilon_g - \Delta)}}. \tag{19}$$

We note that Eq. (19) is valid at arbitrary values of strain gap Δ, not necessarily much smaller than ϵ_g.

For the case of a very large gap of the direct material ($\epsilon_{g2} \to \infty$), we obtain a similar equation for the surface-state dispersion at $k_y \to 0$

$$\epsilon = \epsilon_{v1} - \frac{\Delta}{2} + \frac{P^2 k_y^2}{2\epsilon_{g1} - 3\Delta}. \tag{20}$$

At large k_y when the energy of the surface branch becomes much larger than Δ, the surface curve coincides with the one obtained without any strain.

Thus, the surface branch starts at the top of the heavy-hole band and at large k_y approaches the light-hole band, see Fig. 6(a) and similar picture in the Supplementary Material of Ref. [21].

The results of the k-P model agree very well with the results obtained from more sophisticated methods, such as the *ab initio* calculations, for Hg-terminated HgTe (001) surface. For Te-terminated surface it is not necessarily the case, see Refs. [39,40].

III. SURFACE STATES WITHIN LUTTINGER MODEL WITH STRAIN

As we have demonstrated, the surface states that lie within the gap opened by strain between light- and heavy-hole bands are of special importance. These states can be studied in more detail within the Luttinger model, which describes the valence Γ_8 light- and heavy-hole bands. This model allows exact analytical results for the spectrum and spinors of the surface states for an arbitrary ratio between the effective masses of light and heavy holes, and represents a more accurate description compared to the model we used above. In Ref. [5] the solution for those states was found in the absence of strain.

The Luttinger Hamiltonian for HgTe, which describes the group of states $|lh - 1/2, hh + 3/2>$ has the form

$$\hat{H}_L = \begin{bmatrix} k^2(\gamma - \tilde{\gamma}) - \Delta/4 & \sqrt{3}\tilde{\gamma}k_+^2 + \sqrt{3}\Delta/4 \\ \sqrt{3}\tilde{\gamma}k_-^2 + \sqrt{3}\Delta/4 & k^2(\gamma + \tilde{\gamma}) + \Delta/4 \end{bmatrix}. \tag{21}$$

The energy is counted from the degeneracy point of light-hole and heavy-hole bands in the HgTe region in the absence of strain ($\epsilon_{v1} = 0$), $\Delta > 0$ is the gap value due to strain, $\hat{k}_\pm = \hat{k}_x \pm ik_y$, $\hat{k}_x = -id/dx$, $\hat{k}^2 = k_y^2 + \hat{k}_x^2$, and wave vector within the boundary is again directed along the y axis, γ, $\tilde{\gamma}$ are the Luttinger constants, which determine the bulk spectra of light and heavy holes in the absence of strain: $\epsilon_{l,h} = k^2(\gamma \pm 2\tilde{\gamma})$. We assume that γ, $\tilde{\gamma} > 0$ and $(\gamma - 2\tilde{\gamma}) < 0$, $(\gamma + 2\tilde{\gamma}) > 0$ (gapless material). As before, the Hamiltonian describing the time-reversed solutions is obtained from Eq. (21) by replacing k_y by $-k_y$.

We consider the HgTe sample, which occupies the $x > 0$ space and $x = 0$ is the boundary with the vacuum or wide gap direct material (infinite barrier). We then seek for the surface states located within the gap created by strain. The x components of the wave vectors corresponding to the light and heavy holes for a given energy ϵ of the surface state are given by $k_{l,x} = i\kappa_l$, $k_{h,x} = i\kappa_h$, $\kappa_{l,h} > 0$, and the state exponentially decays towards the bulk of HgTe region. The expressions for $\kappa_{l,h}$ are

$$k_y^2 - \kappa_{l,h}^2 = \frac{\gamma\epsilon + \tilde{\gamma}\Delta \mp \sqrt{(\gamma\Delta/2 + 2\tilde{\gamma}\epsilon)^2 + 3\tilde{\gamma}\Delta k_y^2(4\tilde{\gamma}^2 - \gamma^2)}}{(\gamma^2 - 4\tilde{\gamma}^2)}. \tag{22}$$

Here $-$ corresponds to κ_l and $+$ to κ_h. The corresponding two-component spinor wave functions look like

$$\Psi_{l,h}(x) \propto e^{ik_y y} e^{-\kappa_{l,h}x} \begin{pmatrix} \sqrt{3}\Delta/4 - \sqrt{3}\tilde{\gamma}(k_y + \kappa_{l,h})^2 \\ \epsilon + \Delta/4 - (\gamma - \tilde{\gamma})(k_y^2 - \kappa_{l,h}^2) \end{pmatrix}. \tag{23}$$

A. Boundary condition and solution

The wave function of the surface state is the superposition of the functions Eq. (23): $\Psi_s(x) = a\Psi_l(x) + b\Psi_h(x)$. For an infinite barrier we use zero boundary conditions at $x = 0$: $\Psi_s(x = 0) = 0$, obtaining the following equation for the

energy ϵ of the surface state:

$$\left[\epsilon + \Delta/4 - (\gamma - \tilde{\gamma})(k_y^2 - \kappa_h^2)\right]\left[\Delta/4 - \tilde{\gamma}(k_y + \kappa_l)^2\right]$$
$$= \left[\epsilon + \Delta/4 - (\gamma - \tilde{\gamma})(k_y^2 - \kappa_l^2)\right]$$
$$\times \left[\Delta/4 - \tilde{\gamma}(k_y + \kappa_h)^2\right]. \qquad (24)$$

Positive real values of $\kappa_{l,h}$ found from Eq. (22) should be used while solving Eq. (24).

The exact solution for the surface-state dispersion reads:

$$\epsilon = -\left(\frac{\gamma}{2\tilde{\gamma}}\right)\frac{\Delta}{2} + k_y^2(\gamma + 2\tilde{\gamma})\left[1 - \frac{(1 + \sqrt{3\beta})^2}{4}\right], \quad (25)$$

where $\beta = (2\tilde{\gamma} - \gamma)/(\gamma + 2\tilde{\gamma}) > 0$ is the ratio of the bulk effective masses of the light and heavy holes. The solution Eq. (25) (we call it further as DK 1 branch) coincides at $\Delta = 0$ with the solution obtained in Ref. [5]. It is interesting that the spectrum in the presence of the strain differs from the one at $\Delta = 0$ only by the constant shift proportional to the strain value. Using this dispersion we can find from Eq. (22) the values of $\kappa_{l,h}$:

$$\kappa_{l,h} = \frac{(1 + \beta)\sqrt{3}}{4\sqrt{\beta}}k_y \mp \frac{k_y}{\sqrt{3\beta}}\sqrt{\left[1 - \frac{(1 + \sqrt{3\beta})^2}{4}\right]^2 + \frac{3\beta\Delta}{4\tilde{\gamma}k_y^2}}. \qquad (26)$$

This solution is valid at $k_y > 0$. Moreover, from the condition that $\kappa_l > 0$ we get the critical value of $k_y^\star$, thus the surface state actually exists at $|k_y| > k_y^\star$, see Fig. 6(b).

$$k_y^\star = \sqrt{\frac{\Delta}{\tilde{\gamma}\left[2 + \sqrt{\frac{3}{\beta}}(1 - \beta)\right]}}. \qquad (27)$$

It is easy to see that at this critical k_y the surface state curve merges with the bulk heavy-hole band (corresponding to $k_x = 0$). In accordance with the results obtained above, in the limit $\beta = 0$ (infinite mass of the heavy holes) the critical value $k_y^\star \to 0$. Moreover, since $k_y^\star$ must be real, we obtain the following condition: $\beta < 3$. It coincides with the one obtained in Ref. [5] for the case $\Delta = 0$. As we know from Ref. [5], in the case $\Delta = 0$ there are two solutions for the surface states. This is true also in the presence of the strain. For the case of $\Delta \geqslant 0$ the second solution (DK 2 branch) exists only at $\beta > 1/3$. Since this case does not correspond to real HgTe or α-Sn materials, we do not consider it here. However, in the case of $\Delta < 0$, which corresponds to the compressive in-plane strain, the situation is qualitatively different. This latter Dirac semimetal case is considered in Sec. IV.

B. Spin structure of the surface state

Let us describe the spin structure of the solution obtained in this section. We have mentioned already that even without strain the DK surface state [5] has all the features of a topological surface state for a 3D topological insulator: momentum-spin locking and absence of back scattering. Indeed, from Eqs. (23), (25) at $\Delta = 0$ we find the following wave function at $k_z = 0$, $k_y > 0$:

$$\Psi_\downarrow(x) = (e^{-\kappa_l x} - e^{-\kappa_h x}) \cdot \frac{1}{2\sqrt{1 + \beta}}\begin{pmatrix}\sqrt{3} + \sqrt{\beta} \\ \sqrt{3\beta} - 1\end{pmatrix}. \quad (28)$$

Let us calculate the average value of the different components of spin using this function. While doing so we should remember that the upper coefficient of the spinor Eq. (28) corresponds to $-1/2$ projection over the z axis and the lower coefficient to the $+3/2$ projection. Coefficients corresponding to $+1/2$, $-3/2$ projections are equal to zero. Then using 4×4 matrices $\hat{J}_x, \hat{J}_y, \hat{J}_z$ corresponding to the angular momentum $3/2$, we obtain $\langle J_x \rangle = 0$, $\langle J_y \rangle = 0$, and

$$\langle J_z \rangle = \frac{1}{4(1 + \beta)}\left[\frac{3}{2}(\sqrt{3\beta} - 1)^2 - \frac{1}{2}(\sqrt{3} + \sqrt{\beta})^2\right]$$
$$= \frac{\beta - \sqrt{3\beta}}{(1 + \beta)}. \qquad (29)$$

Thus, the average spin is parallel to the boundary and directed perpendicular to the electron momentum. For the opposite momentum direction within the plane one obtains the average spin, which is the opposite sign of the value given by Eq. (29); therefore, one has the momentum-spin locking for this surface state.

Finally, we present here the result obtained within the Luttinger model for the average spin of the surface electron in the presence of strain, $\Delta > 0$. At the conditions $\sqrt{\beta} \ll 1$ and $\Delta/(\tilde{\gamma}k_y^2) < \sqrt{3/\beta}$, we obtain

$$\langle J_z \rangle = -\sqrt{3\beta}\frac{1}{1 - \Delta/(9\tilde{\gamma}k_y^2)}. \qquad (30)$$

As before, $\langle J_x \rangle = 0$, $\langle J_y \rangle = 0$, i.e., spin is parallel to the boundary and perpendicular to the momentum direction. [At $\Delta = 0$ and $\beta \ll 1$ the result Eq. (30) coincides with the one given by Eq. (29).] It is interesting that in contrast to the case without strain, the average spin now increases with the lowering of energy, getting near the $k_y^\star$, Eq. (27), the maximum value, which is of order of unity even for $\beta \ll 1$.

We consider the case of isotropic Luttinger model; therefore, the spin-momentum locking is an exact characteristic of the surface states, which can be probed experimentally. By adding higher-order terms one can create some warping. (In practice, it can be done by growing film of topological material on the substrate material with the proper symmetry.) In the particular case of the hexagonal warping, Ref. [41], the component of the spin normal to the plane appears; however, the component of the spin parallel to the plane still strongly correlates with the momentum direction. Therefore, in a sense, complete breaking of the locking does not happen.

IV. SURFACE STATES IN THE DIRAC SEMIMETAL REGIME (LUTTINGER MODEL)

So far we have considered the case of tensile in-plane strain ($\Delta > 0$), which opens the gap between the light- and heavy-hole bands, creating the topological insulator in the material such as HgTe. In the opposite case, $\Delta < 0$, one has the situation of the Dirac semimetal where the light-hole band and the heavy-hole band overlap, forming two Dirac points. The bulk energy dispersion near those points is a linear function of the in-plane wave vector, see Fig. 7. The case of the Dirac semimetal has received much less attention in the literature, especially its analytical treatment. We consider here the problem of surface states in the Dirac semimetal regime
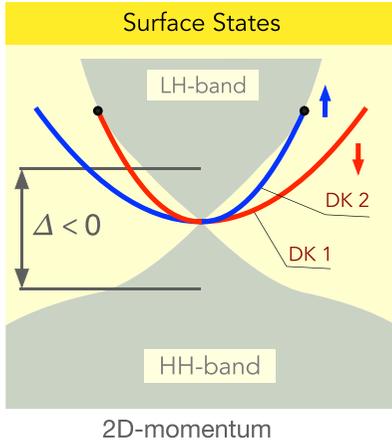
FIG. 7. Case of compressive in-plane strain ($\Delta < 0$). Two parabolic surface-state branches (DK 1 and DK 2) arise around the conic point of the bulk projected states located at the energy $\epsilon_D = (\gamma|\Delta|/4\tilde{\gamma})$. The DK 2 branch flows into the bulk spectrum at the value of in-plane momenta given by Eq. (36). At $\Delta = 0$, the DK 2 branch arises only for $\beta > 1/3$, see Ref. [5].

within the Luttinger model. Similar to the $\Delta > 0$ case, it is possible to find an exact analytical solution for the spectrum and wave functions of the surface states within the Luttinger model (i.e., for energies smaller than the $\Gamma_8 - \Gamma_6$ gap). This presents a comprehensive analytical solution for this problem. Note, that we do not study here the surface states located within the $\Gamma_8 - \Gamma_6$ gap. It is clear, however, that similar to the previous cases considered in this paper, those states lie on top of the continuum of the heavy-hole states and are strongly hybridized with them. Therefore, there are no linear in k-vector surface states going from s band to the light-hole band across the $\Gamma_8 - \Gamma_6$ gap, i.e., the popular picture is again incorrect.

We start with the bulk spectrum near the Dirac point at $\Delta < 0$. Again the x axis is perpendicular to the boundary, and the wave vector along the boundary is equal to k_y. From the expression for the dispersions of the light and heavy holes

$$\epsilon_{l,h} = \gamma(k_x^2 + k_y^2)$$
$$\pm 2\tilde{\gamma}\sqrt{\left(\frac{\Delta}{4\tilde{\gamma}} + k_x^2\right)^2 + k_y^2\left(k_y^2 + 2k_x^2 - \frac{\Delta}{4\tilde{\gamma}}\right)} \quad (31)$$

one can easily see that at $k_x^2 = |\Delta|/4\tilde{\gamma}$ (Dirac point) the bulk spectrum is linear

$$\epsilon_b(k_y) = \frac{\gamma|\Delta|}{4\tilde{\gamma}} \pm k_y\sqrt{3\tilde{\gamma}|\Delta|}; \quad |k_y| \ll \sqrt{|\Delta|/\tilde{\gamma}}. \quad (32)$$

We consider now the surface states. We will show that the two branches of parabolic surface states start from the bulk Dirac point. It is easy to realize that we can use the same Eq. (24) for the dispersion of the surface states. The only difference now is that the strain is negative: $\Delta < 0$. Therefore, the solutions Eqs. (25), (26) are still valid. We see that the first (DK 1) branch, Eq. (25), starts at the energy of the bulk Dirac point: $\gamma|\Delta|/4\tilde{\gamma}$, see Fig. 7, and has the same mass in the absence of strain. The expressions for the $\kappa_{l,h}$ are still given by Eq. (26),

the only difference from the case $\Delta > 0$ is that now quantities $\kappa_{l,h}$ become complex at small enough values of k_y. Consider the positive k_y, the character of the surface wave function is different depending on whether k_y is smaller or bigger than k_{y1}, where

$$k_{y1} = 2\sqrt{\frac{\beta|\Delta|}{\tilde{\gamma}}} \cdot \frac{1}{|1 - \sqrt{3\beta}|(\sqrt{3} + \sqrt{\beta})}. \quad (33)$$

If $k_y < k_{y1}$, then $\kappa_l = \kappa_h^\star = a - ib$, and the surface function has the form $\Psi_s(x) \propto \sin(bx) \cdot \exp(-ax)$ with a, b to be found from Eq. (26). In the case $k_y > k_{y1}$ both $\kappa_{l,h}$ are real and positive, and the surface wave function has the form $\Psi_s(x) \propto (e^{-\kappa_l x} - e^{-\kappa_h x})$, with $\kappa_{l,h}$ given by Eq. (26) at $\beta < 3$.

The interesting feature of the problem is that there is a great difference in the number of DK surface branches, which exist at positive or negative Δ. In the case of $\Delta > 0$, at $\beta < 1/3$ (which is the case of HgTe and α-Sn materials) there is only one branch DK 1 of surface states. We will show now that in the case of $\Delta < 0$, considered here, there are always two DK surface branches (even for $\beta < 1/3$). Thus, in the case of Dirac semimetal, one always has two DK surface branches, which for the unstrained sample were found in the original work Ref. [5]. We should mention that the authors of Ref. [42] found only one DK branch in the regime of a Dirac semimetal. Moreover, these two branches continue until the $k_y = 0$ point, in strong contrast to the conclusion made in Ref. [42].

Let us discuss the second surface branch, which exists at $\Delta < 0$. It is the solution of Eq. (24) where the change $k_y \rightarrow -k_y$ is made, i.e., for positive k_y it is obtained from the second block of basis states and has an opposite spin compared to the first (DK 1) solution. The exact solution for the dispersion of the second surface branch (DK 2) reads:

$$\epsilon = \left(\frac{\gamma}{2\tilde{\gamma}}\right)\frac{|\Delta|}{2} + k_y^2(\gamma + 2\tilde{\gamma})\left[1 - \frac{(1 - \sqrt{3\beta})^2}{4}\right]. \quad (34)$$

The corresponding values of $\kappa_{l,h}$ are

$$\kappa_{l,h} = \frac{(1+\beta)\sqrt{3}}{4\sqrt{\beta}}k_y \mp \frac{k_y}{\sqrt{3\beta}}\sqrt{\left[1 - \frac{(1-\sqrt{3\beta})^2}{4}\right]^2 + \frac{3\beta\Delta}{4\tilde{\gamma}k_y^2}}. \quad (35)$$

This solution is valid at $k_y > 0$. For small enough k_y the quantities $\kappa_{l,h}$ are complex again, and one can have a localized state even at $\beta < 1/3$, in strong contrast to the cases $\Delta \geqslant 0$. At large enough k_y, however, when both $\kappa_{l,h}$ are real, from the condition that $\kappa_l > 0$ we get the critical value $k_y = \tilde{k}_y$, after which the second surface branch ceases to exist (it merges with the bulk band), see Fig. 7. The expression for $\tilde{k}_y$ reads

$$\tilde{\gamma}\tilde{k}_y^2 = \frac{\sqrt{\beta}|\Delta|}{(1 - \sqrt{3\beta})(\sqrt{3} + \sqrt{\beta})}; \quad \beta < 1/3. \quad (36)$$

Thus we see that the dispersions of the surface states found by us here are very much different from the ones usually used, see, for example, Refs. [33,34]. The authors of Refs. [33,34] came to the conclusions that the effects they observe are due to the Dirac semimetal surface states of α-Sn layer rather than due to the bulk states of this material. One of the effects is

the switching of a magnet by spin-orbit torque from a Dirac semimetal by a charge current at room temperature, without an external magnetic field. This shows that the topological Dirac semimetal α-Sn is a promising material for the spintronic applications. In that respect, the results obtained by us are potentially important for the correct interpretation of the current and future experiments.

V. CONCLUSIONS

We have reconsidered the physical origin of surface states in inverted-band semiconductors, such as HgTe and α-Sn. By considering analytically two simple and exactly solvable models we have clarified several important issues concerning the nature of the states, among them the role of band structure and the role of strain. In particular, we have shown that neglecting the coupling to the heavy-hole band leads to the incorrect physical picture of the surface states obtained in many recent publications. Such a procedure results in the simple picture of linear in k-vector Dirac surface states crossing the whole gap between the conduction and light-hole bands, see Fig. 1. This latter picture does not reflect the true nature of the surface states. We have also shown that in a topological insulator regime an applied strain is a smooth deformation to the surface states that does not lead to any drastic change of the physical properties in these materials, in contrast to what has recently been published in the literature. In the Dirac semimetal regime, however, the physics of surface states, which arises around the conic point of the bulk projected states crucially depends on the applied strain.

We have reconciled different analytical solutions found in the 1980s with the results obtained recently by many groups using numerical simulations. In particular, we have demonstrated that the DK surface state [5] predicted for HgTe within the Luttinger model is the result of strong hybridization of the Dirac state with the heavy-hole states. We have also shown that these states [5] possess all the essential features of the topological states in the presence of strain that induces a gap between the light- and heavy-hole bands. As such, they should play the most important role in the experiments that probe the topological properties of the sample when the Fermi level is located within this strain-induced gap. In contrast, the surface states located within the $\Gamma_8 - \Gamma_6$ gap lie on top of the continuum of the heavy-hole states and strong hybridization with them makes those surface states marginal in observable phenomena. They still can be detected with the ARPES experiments at small 2D momenta along the interface where the overlap of these states with heavy holes is weak, see Refs. [31,38]. The results obtained by us can be used for the correct interpretation of the current and future experiments with materials such as α-Sn and HgTe.

ACKNOWLEDGMENTS

[1] I. E. Tamm, Phys. Z. Sowjetunion **1**, 733 (1932).

[2] W. Shockley, Phys. Rev. **56**, 317 (1939).

[3] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).

[4] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).

[5] M. I. Dyakonov and A. V. Khaetskii, JETP Lett. **33**, 110 (1981).

[6] B. A. Volkov and O. A. Pankratov, JETP Lett. **42**, 178 (1985).

[7] N. A. Cade, J. Phys. C: Solid State Phys. **18**, 5135 (1985).

[8] R. A. Suris, Sov. Phys. Semicond. **20**, 1258 (1986).

[9] O. A. Pankratov, S. V. Pakhomov, and B. A. Volkov, Solid State Commun. **61**, 93 (1987).

[10] B. A. Bernevig and S.-C. Zhang, Phys. Rev. Lett. **96**, 106802 (2006).

[11] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).

[12] X. Dai, T. L. Hughes, X. L. Qi, Z. Fang, and S.-C. Zhang, Phys. Rev. B **77**, 125319 (2008).

[13] Y. Ando, J. Phys. Soc. Jpn. **82**, 102001 (2013).

[14] H. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, Nature Phys. **5**, 438 (2009).

[15] Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature Phys. **5**, 398 (2009).

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

[17] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang, Science **318**, 766 (2007).

[18] C. Liu, T. L. Hughes, X.-L. Qi, K. Wang, and S.-C. Zhang, Phys. Rev. Lett. **100**, 236601 (2008).

[19] S. S. Krishtopenko, S. Ruffenach, F. Gonzalez-Posada, G. Boissier, M. Marcinkiewicz, M. A. Fadeev, A. M. Kadykov, V. V. Rumyantsev, S. V. Morozov, V. I. Gavrilenko, C. Consejo, W. Desrat, B. Jouault, W. Knap, E. Tournie, and F. Teppe, Phys. Rev. B **97**, 245419 (2018).

[20] D. Hsieh, Y. Xia, D. Qian, L. Wray, J. H. Dil, F. Meier, J. Osterwalder, L. Patthey, J. G. Checkelsky, N. P. Ong, A. V. Fedorov, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature (London) **460**, 1101 (2009).

[21] C. Brüne, C. X. Liu, E. G. Novik, E. M. Hankiewicz, H. Buhmann, Y. L. Chen, X. L. Qi, Z. X. Shen, S.-C. Zhang, and L. W. Molenkamp, Phys. Rev. Lett. **106**, 126803 (2011).

[22] D. A. Kozlov, Z. D. Kvon, E. B. Olshanetsky, N. N. Mikhailov, S. A. Dvoretsky, and D. Weiss, Phys. Rev. Lett. **112**, 196801 (2014).

[23] J. G. Checkelsky, Y. S. Hor, M.-H. Liu, D.-X. Qu, R. J. Cava, and N. P. Ong, Phys. Rev. Lett. **103**, 246601 (2009).

[24] J. G. Analytis, J.-H. Chu, Y. Chen, F. Corredor, R. D. McDonald, Z. X. Shen, and I. R. Fisher, Phys. Rev. B **81**, 205407 (2010).

[25] Z. Ren, A. A. Taskin, S. Sasaki, K. Segawa, and Y. Ando, Phys. Rev. B **84**, 165311 (2011).

[26] B. Xia, P. Ren, A. Sulaev, P. Liu, S.-Q. Shen, and L. Wang, Phys. Rev. B **87**, 085442 (2013).

[27] Y. Pan, D. Wu, J. R. Angevaare, H. Luigjes, E. Frantzeskakis, N. de Jong, E. van Heumen, T. V. Bay, B. Zwartsenberg, and Y. K. Huang, New J. Phys. **16**, 123035 (2014).

[28] Y. Ohtsubo, P. Le Fevre, Francois Bertran, and A. Taleb-Ibrahimi, Phys. Rev. Lett. **111**, 216401 (2013).

[29] A. Barfuss, L. Dudy, M. R. Scholz, H. Roth, P. Hopfner, C. Blumenstein, G. Landolt, J. H. Dil, N. C. Plumb, M. Radovic, A. Bostwick, E. Rotenberg, A. Fleszar, G. Bihlmayer, D. Wortmann, G. Li, W. Hanke, R. Claessen, and J. Schafer, Phys. Rev. Lett. **111**, 157205 (2013).

[30] S. Küfner, M. Fitzner, and F. Bechstedt, Phys. Rev. B **90**, 125312 (2014).

[31] V. A. Rogalev, T. Rauch, M. R. Scholz, F. Reis, L. Dudy, A. Fleszar, M. A. Husanu, V. N. Strocov, J. Henk, I. Mertig, J. Schafer, and R. Claessen, Phys. Rev. B **95**, 161117(R) (2017).

[32] Z. Shia, X. Wanga, C. Xu, P. Wang, Y. Liu, and T.-C. Chiang, Phys. Lett. A **384**, 126782 (2020).

[33] J. Ding, C. Liu, V. Kalappattil, V. Kalappattil, Y. Zhang, O. Mosendz, U. Erugu, R. Yu, J. Tian, A. DeMann, S. B. Field, X. Yang, H. Ding, J. Tang, B. Terris, A. Fert, H. Chen, and M. Wu, Adv. Mater. **33**, 2005909 (2021).

[34] J. Ding, C. Liu, Y. Zhang, V. Kalappattil, R. Yu, U. Erugu, J. Tang, H. Ding, H. Chen, and M. Wu, Adv. Funct. Mater. **31**, 2008411 (2021).

[35] L. G. Gerchikov and A. V. Subashiev, Phys. Stat. Sol. (b) **160**, 443 (1990).

[36] G. Bir and G. E. Pikus, *Symmetry and Strain-induced Effects in Semiconductors* (John-Wiley and Sons, Chichester, 1974).

[37] S. S. Krishtopenko, M. Antezza, and F. Teppe, J. Phys.: Condens. Matter **32**, 165501 (2020).

[38] T. Rauch, V. A. Rogalev, M. Bauernfeind, J. Maklar, F. Reis, F. Adler, S. Moser, J. Weis, T. L. Lee, P. K. Thakur, J. Schafer, R. Claessen, J. Henk, I. Mertig, Phys. Rev. Mater. **3**, 064203 (2019).

[39] T. Rauch, S. Achilles, J. Henk, and I. Mertig, Phys. Rev. B **96**, 035124 (2017).

[40] S.-C. Wu, B. Yan, and C. Felser, Europhys. Lett. **107**, 57006 (2014).

[41] L. Fu, Phys. Rev. Lett. **103**, 266801 (2009).

[42] O. V. Kibis, O. Kyriienko, and I. A. Shelykh, New J. Phys. **21**, 043016 (2019).