PHYSICAL REVIEW B **104**, 195407 (2021)

Hard-wall edge confinement in two-dimensional topological insulators and the energy of the Dirac point

P. C. Klipstein*
Semiconductor Devices, P.O. Box 2250, Haifa 31021, Israel

(Received 23 December 2020; accepted 8 October 2021; published 8 November 2021)

In 2D topological insulators (TIs) based on semiconductor quantum wells such as HgTe/CdTe, hard wall spin-polarized edge states are calculated from an 8×8 linear-k multiband (LKMB) Hamiltonian based directly on $\mathbf{k} \cdot \mathbf{p}$ theory, and from a 4×4 BHZ Hamiltonian derived from it by the elimination of remote states using perturbation theory. Both approaches lead to similar results with standard boundary conditions obtained by integrating the eigenstate equation across an interface (SBCs). In contrast, open boundary conditions (OBCs) yield unphysical results and do not work for the LKMB Hamiltonian. Their failure is traced to a spurious solution introduced by the elimination process. In the BHZ treatment, SBCs are shown to be consistent with perturbation theory on both sides of the boundary, and a wall hybridization parameter is estimated for a vacuum using a basis of empty crystal free electron states. A Dirac point is not expected for a vacuum, but can exist when a thin passivation layer is used whose midgap energy is nearly degenerate with that of the TI. In the absence of interface band mixing (IBM), the Dirac point is then very close to midgap and virtually independent of the TI band asymmetry. IBM introduces a significant energy shift, which decreases monotonically with edge state wave vector. Using SBCs, the paradoxical extension of the BHZ edge states across the topological phase transition is also resolved.

DOI: 10.1103/PhysRevB.104.195407

I. INTRODUCTION

Surface states occur in many areas of condensed matter physics, often when some material parameter changes sign across a boundary. Two classic examples are surface plasmons and surface optical phonons, where a sign reversal of the dielectric function leads to edge confinement with an amplitude that decays away from the edge, both into the material and into the surroundings [1–5]. More recently, electronic edge states have been discovered in topological insulators (TIs), where a change in sign of the band gap parameter reverses the ordering of even and odd parity crystal periodic basis functions [6]. These edge states also have the unique property that their direction of motion depends on the electron spin, leading to unusual phenomena such as dissipationless ballistic transport and the quantum spin Hall effect [7,8]. Their experimental observation, however, remains a challenge and in some cases more advanced surface passivation techniques may be needed, in order to eliminate parallel conduction paths through trivial edge states due, for example, to unsatisfied dangling bonds [9].

A popular treatment for the spin-polarized edge states in two dimensional TIs is based on the $\mathbf{k} \cdot \mathbf{p}$ theory, where useful results can be obtained with the simple four-band BHZ Hamiltonian based on 2×2 blocks for each spin direction [7]. Each 2×2 block is associated with a Chern number whose value on each side of the boundary can be related to the number of spin-polarized edge states. At a hard wall, the nature of the edge state confinement can depend strongly on the boundary conditions used. Standard boundary conditions for the wave function and its derivative (SBCs) and other related approaches can lead to strong confinement, where the wave function has a peak amplitude at the edge, similar to the classic surface states described above [10–14]. On the other hand, open boundary conditions (OBCs) lead to weak confinement, where the amplitude is zero at the edge and only reaches a peak typically 10–100 Å away [8,15,16]. While deep quantum well (QW) states also have negligible amplitude at the edge of the well, the bound state vanishes when one of the barriers is removed and this should not be confused with true edge confinement. Of these methods, the OBC approach is by far the most popular because it apparently avoids any explicit treatment of the wall [8,15–49]. Notwithstanding the greater simplicity of OBCs, this author has previously argued that the spin-polarized edge states in TIs cannot have zero amplitude at the edge, and in this respect they are no different from their classic plasmon or phonon counterparts. The weakly confined wave function in the OBC treatment arises from a mathematically correct but physically spurious solution of the BHZ Hamiltonian, which not only gives an unphysical wave function but also leads to other unphysical properties. The present paper provides further evidence for this point of view, based on a simple multiband $\mathbf{k} \cdot \mathbf{p}$ model that reduces to the four band BHZ model using Pikus-Bir perturbation theory. It is shown that edge states are not possible in the multiband model using OBCs, while the BHZ and multiband results correspond very well when SBCs are used. In contrast to some alternatives to OBCs, which are phenomenological in nature and do not consider the wall explicitly [13,38,40]. SBCs

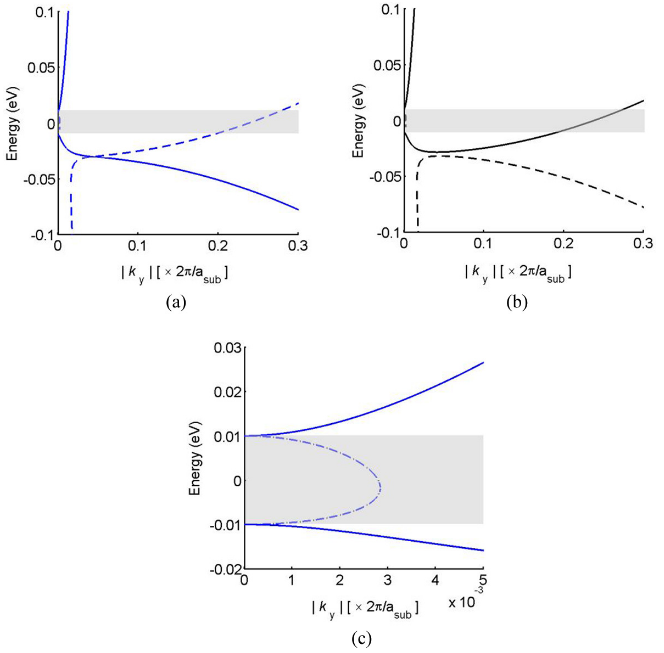
---

[*] philip_k@scd.co.il

provide a realistic physical picture on both sides of the boundary. They can also give a physical edge dispersion whose Dirac point is close to midgap, even when band structure asymmetry would cause the Dirac point to be near a bulk band edge in the four band OBC treatment.

In most cubic semiconductors, one of the crystal periodic basis functions of the BHZ Hamiltonian E_1 is based on antibonding s orbitals, while the other H_1 is based on bonding p orbitals, corresponding to the conduction and valence bands, respectively. In two-dimensional TIs, usually based on semiconductor QWs such as HgTe/CdTe or InAs/GaSb/AlSb, their order can be reversed by increasing the QW width. Spin-polarized edge states are then predicted according to the change in Chern number at the boundary with the wall or other semiconductor material [50]. For a TI with symmetrical bands, the 2×2 spin up Hamiltonian can be written as $H_{2\times2\uparrow} = A(\sigma_x k_x - \sigma_y k_y) + \sigma_z(M + \mathbf{k} \cdot B\mathbf{k})$ and the Chern number is given by $N_C = -\frac{1}{2}[\mathrm{sgn}(M) - \mathrm{sgn}(B)]$, where $2M$ is the semiconductor band gap, B^{-1} is related to the band effective mass, and $\mathbf{k} = (k_x, k_y)$ is the in-plane wave vector [28]. The symmetrical operator ordering used in the quadratic term has been justified in Ref. [51]. For a boundary between a TI material with $M < 0$, $B > 0$ and a wall with M and B parameter values, $M_0 > 0$, $B_0 > 0$, the change in Chern number is $\Delta N_C = 1$, supporting a single spin-polarized edge state. If the sign of B_0 is reversed, $\Delta N_C = 2$ and two edge states are predicted. Examples of both types will be discussed in this paper.

Band asymmetry can be incorporated into the 2 x 2 spin up Hamiltonian by adding an additional quadratic term proportional to the identity matrix $H'_{2\times2\uparrow} = H_{2\times2\uparrow} + I_{2\times2}\mathbf{k} \cdot D\mathbf{k}$ and parameter values can be found, usually empirically, which provide a good description of the small wave vector states near the band edges of most semiconductor materials. There are also solutions in the band-gap energy range, $E < |M|$. Considering the dispersion in the y direction ($k_x = 0$), the "middle states" have a small imaginary wave vector $k_y = i\sigma_m(E)$ and describe tunneling, for example when the semiconductor is used as a thin barrier material. The "wing states", on the other hand, have a wave vector $k_y = i\sigma_w(E)$ that is imaginary or real, depending on whether $D < B$ or $D > B$, respectively. When the difference between B and D is small, the decay parameter at zero energy is given by the simple formula: $\sigma_w(0) = \pm\sqrt{(2MB + A^2)/2B(B - D)}$, demonstrating that the state is only evanescent when $D < B$. This is shown in Fig. 1, which compares two TI Hamiltonians with $M = -0.01$ eV, $A = 4$ eV Å, and $B = 200$ eVÅ^2, and where extended and evanescent states are depicted as solid and dashed lines, respectively. The blue plot in Fig. 1(a) is for $D = 199.5$ eV Å^2, while the black plot in Fig. 1(b) is for $D = 200.5$ eV Å^2. The same color scheme is used in Fig. 1(c), where the results for both D values are superimposed in the small wave vector region. The switch between large imaginary and real wing wave vectors of the same magnitude shows clearly in the shaded band gap region of Figs. 1(a) and 1(b), occurring when the crossing of extended and evanescent states below the band gap changes to an anti-crossing. In contrast, the middle states with small imaginary wave vector in Fig. 1(c) are indistinguishable for the two cases and totally insensitive to the increase in D.



(a)
(b)
(c)

FIG. 1. Dispersions of the spin up BHZ Hamiltonian in the y direction with $M = -0.01$ eV, $A = 4$ eV Å, $B = 200$ eV Å^2 and (a) $D = 199.5$ eV Å^2 (blue curves), (b) $D = 200.5$ eV Å^2 (black curves), and (c) both D values (blue and black superimposed). Note the shorter range of wave vector in (c). Extended (evanescent) states are depicted as solid (dashed) lines ($a_{\mathrm{sub}} = 6.0954$ Å).

The spurious nature of the wing states was first discussed by White and Sham [52], and later by Schuurmans and t'Hooft [53]. Both the extended and evanescent wing solutions do not correspond to any plausible dispersion and are obviously unphysical [54]. As discussed in these same references, and demonstrated explicitly in Sec. II, the wing solutions arise because the number of basis states is too small. The effect of remote states, which have been omitted from the Hamiltonian is incorporated in the k-quadratic terms (proportional to B and D) using perturbation theory, and this always leads to a spurious solution. Moreover, since the range of real or imaginary wave vectors must then be limited to magnitudes, $k_{\max} < \sqrt{\frac{|M|}{|B|+|D|}}$, where the perturbation terms are less than the typical band energy, the spurious solution is usually found to lie well beyond this limit (see Sec. III). For the example in Fig. 1, the valid range corresponds to $k_{\max} < 0.005 \times \frac{2\pi}{a_{\mathrm{sub}}}$, which is the range plotted for the middle solution in Fig. 1(c). The valid range thus includes the middle solution but is about 50 times smaller than the band-gap wave vectors of the wing solutions in Figs. 1(a) and 1(b).

The behavior shown in Fig. 1 is quite general, as can be seen from the simple formula given above for $\sigma_w(0)$, which always gives a switch been real and imaginary wing wave vectors when $D = B$, regardless of the sign of M, or the magnitude of B and D. When B and D are small, the wing solution can lie far outside the Brillouin zone, while the middle solution, $\sigma_m(0) \simeq M/A$, is very insensitive to their values. The eigenvectors for the σ_m and σ_w solutions at a given energy are usually different. However, in a TI material with $M < 0$ and

$D < B$, it turns out that they are identical at energy $E^{\rm D}_{\rm OBC} = -M\frac{D}{B}$ (see Sec. III), which is just below the conduction band edge in the example of Fig. 1(a) [15]. This has led to the common but unfortunate practice of adopting OBC boundary conditions, where both solutions are treated completely seriously and combined into a single wave function with an envelope of the form $\psi = e^{-\sigma_m(E^{\rm D}_{\rm OBC})y} - e^{-\sigma_w(E^{\rm D}_{\rm OBC})y}$ so that the amplitude is zero at the sample boundary, which is assumed to be a hard wall at $y = 0$. Although this would be correct if $H'_{2\times2\uparrow}$ yielded only physical solutions, the inclusion of the spurious solution in the wave function leads to several problematic results, some of which have been discussed previously while others will be elaborated below, including a paradoxical extension of the edge states into the topologically trivial phase (see Appendix B).

It is shown in Sec. II that the BHZ Hamiltonian is a perturbation approximation of a larger linear-k multiband Hamiltonian (LKMB), which does not satisfy OBCs, because the eigenvectors of the gap-states are always different. An extended version of the BHZ Hamiltonian that was recently proposed is also discussed [43]. The equality of the eigenvectors in the standard and extended BHZ Hamiltonians is an artifact introduced by the approximation. The problem with OBCs has been highlighted previously by the author and an alternative approach was proposed for the BHZ model, based on SBC boundary conditions [10,55]. Although SBCs treatments have been reported by other workers, they are generally based on a soft wall and evolve into OBCs when the wall potential increases, because the wave function still includes the spurious solution [31,49]. The SBC approach proposed by the author is for a hard wall and does not include the spurious and physical solutions in the same wave function. It was shown in previous work that two exponential edge solutions can be found when B_0 (and D_0) is (are) negative, consistent with $\Delta N_C = 2$ when $D_0 = 0$. One of these involves only σ_m solutions in the TI and wall, and the other only σ_w solutions. The former yields a simple edge-state dispersion in the useful limit, $M_0 \to \infty$ and $B_0, D_0 \to 0$, close to $E^{\rm SBC}_{2\times2\uparrow} \simeq -Ak_x + Dk_x^2$, which merges smoothly with the bulk band edges, while the latter shows unphysical merging behavior and is rejected as spurious. It will be shown in Sec. II that these results are entirely consistent with the LKMB model, which also puts the Dirac point close to midgap when SBCs are used. When B_0 (and D_0) is (are) positive a single, nonexponential solution exists, again consistent with both models.

In previous work based on the BHZ model, the SBC edge dispersion $E^{\rm SBC}_{2\times2\uparrow}(k_x)$ was obtained by solving a characteristic equation numerically, for both strongly hybridized HgTe/CdTe and weakly hybridized InAs/GaSb/AlSb QWs [55]. In Sec. III of this paper, an analytical solution is obtained for the dispersion, which provides a simple expression for the energy of the Dirac point. The consistency of the SBC treatment with perturbation theory is also analyzed in detail, considering both the TI and wall regions. The dependence of the Dirac point on the wall hybridization parameters is then compared for the BHZ and LKMB models in Sec. IV. Although the wall hybridization parameters are found to have a relatively small effect, a suitable model of the wall region is lacking. Therefore, in Sec. V, an approach is proposed for a consistent $\mathbf{k}\cdot\mathbf{p}$ treatment. This approach yields a wall hybridization parameter for the BHZ model that is fairly similar to that in the TI material, and much smaller than the free electron Dirac value, which was previously suggested as a possibility. It also highlights the importance of an edge passivation material and introduces interface band mixing terms, which may cause a significant additional shift of the Dirac point. In Sec. VI, conclusions are summarized. Even though this paper is focused primarily on the physical nature of the wave function confinement and its influence on the energy of the Dirac point, a treatment of the full SBC edge dispersion is presented for a toy LKMB model in Appendix A, where it is shown to be quite consistent with the BHZ results. In Appendix B, it is also shown that when SBCs are used, the paradoxical extension of the BHZ edge states across the topological phase transition [41] is resolved.

II. LINEAR-K MULTIBAND MODEL

A. k · p Hamiltonian

In 1955, Luttinger and Kohn derived the $\mathbf{k}\cdot\mathbf{p}$ theory in Fourier space and showed how it could be used to treat the potential of an impurity atom in a bulk crystal [56]. This was extended by Volkov and Takhtamirov in the 1990s, who used the same approach to treat semiconductor superlattices [57–59]. When transformed into real space and with a few basic assumptions, their Hamiltonian can be written [51] as

$$E\widetilde{F}_n(\mathbf{r}) = \left(-\frac{\hbar^2}{2m_0}\nabla^2 + E_n\right)\widetilde{F}_n(\mathbf{r}) - \sum_{n'}\frac{i\hbar}{m_0}\mathbf{p}_{nn'}\cdot\nabla\widetilde{F}_{n'}(\mathbf{r}) + \sum_{n'}H^{\rm mod}_{nn'}(y)\widetilde{F}_{n'}(\mathbf{r}), \tag{1}$$

where m_0 is the free electron mass, E_n is the band edge of the zone center basis state $|n\rangle$, $\widetilde{F}_n(\mathbf{r})$ is the envelope function, which only contains Fourier components in the first Brillouin zone, and $\mathbf{p}_{nn'} = \langle n| - i\hbar\nabla|n'\rangle$ is a momentum matrix element between crystal periodic basis states $|n\rangle$ and $|n'\rangle$. For an infinite bulk material, $\widetilde{F}_n(\mathbf{r})$ is a plane wave. The term $H^{\rm mod}_{nn'}(y)$ represents additional terms introduced when the crystal potential is modulated along the y direction, for instance at a boundary between two different materials [51,57–59]. This term is not included in the present bulk treatment, but is discussed further in Sec. V.

Equation (1) can be written in matrix form with elements $H_{nn'}$ and for a bulk material with $H^{\rm mod}_{nn'} = 0$, it is essentially an exact description of the crystal if enough basis states are included. For example, in their seminal paper of 1966, Cardona and Pollak were able to model the whole Brillouin zones of silicon and germanium with 15 zone center basis states [60]. However, in cases where only a small local region of the Brillouin zone is of interest, it is possible to eliminate a large number of the remote states using perturbation theory, leaving only those states in the local energy range. The Bir and Pikus expression for the local Hamiltonian, can be written up to second order as [61]

$$\bar{H}_{mm'} = H_{mm'} - \frac{1}{2}\sum_s H'_{ms}H'_{sm'}\left(\frac{1}{E_s - E_m} + \frac{1}{E_s - E_{m'}}\right) + \dots, \tag{2}$$

in which the local states are $|m\rangle$, $|m'\rangle$, etc., the remote states are $|s\rangle$, $|s'\rangle$, etc., and $H'_{ms} = \mathbf{k}\cdot\langle m \mid \frac{\hbar}{m_0}\mathbf{p} \mid s\rangle$ is an off-diagonal matrix element of $H_{nn'}$ where $\mathbf{k} = -i\nabla$. This reduces the size of the Hamiltonian, but introduces additional k-quadratic terms into the Hamiltonian of local states $\bar{H}_{mm'}$.

The BHZ Hamiltonian for a 2D quantum well $H'_{2\times2\uparrow}$ discussed in Sec. I, is thus derived from a larger number of basis states, using Eq. (2). As a simple toy model, consider the following unperturbed 4×4 spin up Hamiltonian based on Eq. (1), with a basis, $|n\rangle$:

$$H'_{4x4\uparrow}(\mathbf{k}) = I_{4\times4}D'k^2 + \begin{bmatrix} \Delta_1 & 0 & iQ_1k_+ & Q_3k_- \\ 0 & M & Ak_+ & -iQ_2k_- \\ -iQ_1k_- & Ak_- & -M & 0 \\ Q_3k_+ & iQ_2k_+ & 0 & -\Delta_2 \end{bmatrix}, \tag{3}$$

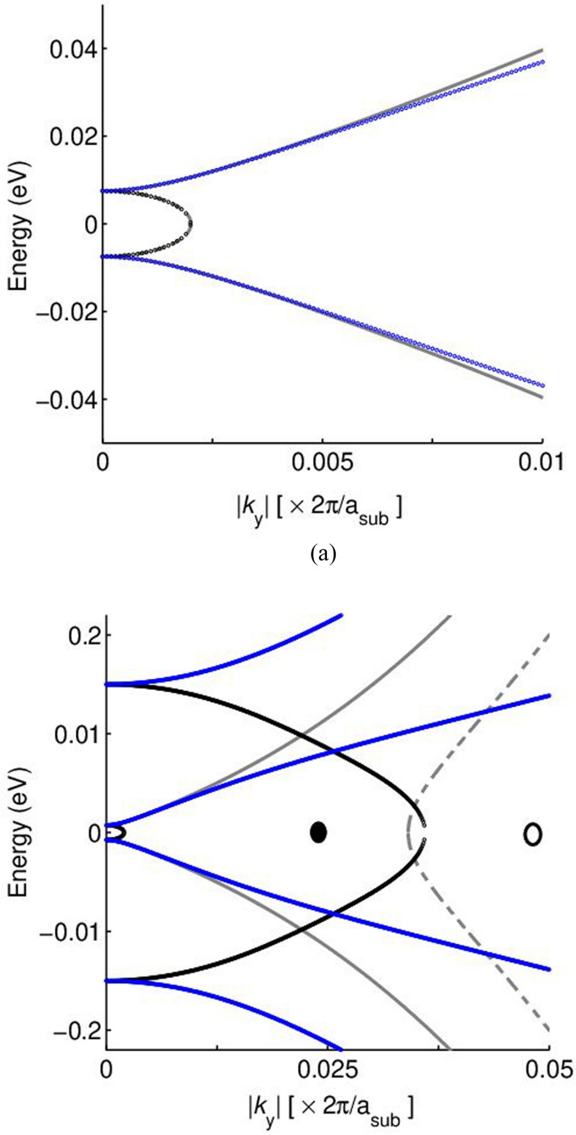
where $k_\pm = k_x \pm ik_y$, $D' = \frac{\hbar^2}{2m_0} = 3.8\,\text{eV}\,\text{Å}^2$, and $n = 1, 2, ..., 4$. The states $|2\rangle$ and $|3\rangle$ are the $E_1\uparrow$ and $H_1\uparrow$ states, discussed in Sec. I. These interact with remote states $|1\rangle$ and $|4\rangle$, at energies, Δ_1 and $-\Delta_2$, respectively. In this example, the remote states are taken to behave like the third confined electron state $E_3\uparrow$ and the first confined light-hole state $L_1\downarrow$ with parameter values $Q_3 \ll Q_1 \simeq Q_2 \simeq \frac{1}{2}A$. These are the nearest-interacting subbands if electron-light hole mixing due to quantum confinement is ignored. This type of mixing is addressed in the next subsection where an extended LKMB treatment is developed with a more realistic dispersion [e.g., see Fig. 3(c)]. However, even with four states, Eq. (3) does capture the kind of strong valence band non-parabolicity that is absent in the BHZ model (e.g., see Fig. 9).

Because D' is quite small, the diagonal quadratic term in the first line of Eq. (3) will be ignored. Moreover, its inclusion would lead to nonexponential hard wall edge state solutions, which are not considered in this paper. The effect of including this term will be discussed further at the end of Sec. V. Using the perturbation expression in Eq. (2), the LKMB Hamiltonian, $H'_{4\times4\uparrow}(\mathbf{k})$ is then reduced to the BHZ Hamiltonian, $H'_{2\times2\uparrow}(\mathbf{k})$ where

$$B = \frac{1}{2}\left(\frac{Q_2^2}{\Delta_2 + M} + \frac{Q_1^2}{\Delta_1 + M}\right), \tag{4a}$$

$$D = \frac{1}{2}\left(\frac{Q_2^2}{\Delta_2 + M} - \frac{Q_1^2}{\Delta_1 + M}\right). \tag{4b}$$

If more states that interact with bands $|2\rangle$ and $|3\rangle$ are included in Eq. (3), there will simply be an additional term from each band in the expressions for B and D. This procedure is analogous to that used to derive the BHZ Hamiltonian in the supplemental material of Ref. [7], where B and D were calculated from the Luttinger parameters. The Luttinger parameters depend on interactions with remote states, which can be expressed in a similar form to Eq. (4) [62]. Note that B and D have no dependence on Q_3. This term only makes a weak k^3 contribution to the off diagonal terms when higher orders are included in Eq. (2), but which are ignored in the BHZ Hamiltonian.

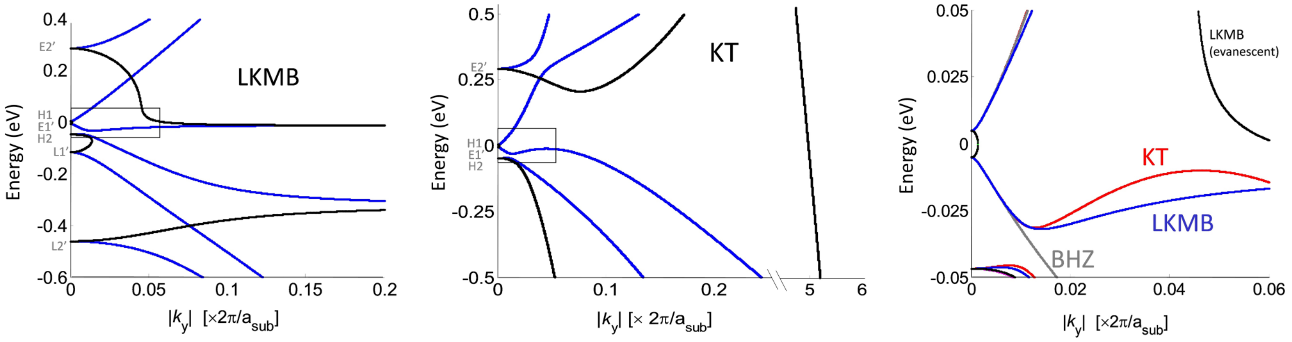


(a)

(b)

FIG. 2. Dispersions in the y direction with $M = -0.0075\,\text{eV}$, $A = Q_1 = Q_2 = 3.83\,\text{eVÅ}$, $Q_3 = 0$, and $\Delta_1 = \Delta_2 = 0.15\,\text{eV}$ for the spin up LKMB Hamiltonian over (a) a short, and (b) a wide, range of wave vector, onto which are superimposed the equivalent BHZ results with $B = 102.9\,\text{eV}\,\text{Å}^2$ and $D = 0$. Extended (evanescent) states are depicted as small blue (black) circles and as solid (dashed) gray lines for the LKMB and BHZ Hamiltonians, respectively. The large solid (open) black circles show the imaginary wave vectors of the zero-energy wing states for $Q_3 = -2$ ($Q_3 = 1$ or 6.7) eVÅ.

Equation (4) shows that the size of the reciprocal mass terms, B and D, in $H'_{2\times2\uparrow}(\mathbf{k})$ is determined by the interaction with the remote states, $|1\rangle$ and $|4\rangle$ that have been eliminated. If these states are removed to infinity, i.e., Δ_1, $\Delta_2 \to \infty$, the reciprocal mass terms are reduced to zero. This highlights a first problem with OBCs. As discussed in Sec. I, the Dirac point of the edge states in the OBC model occurs at $E^{\text{D}}_{\text{OBC}} = -M\frac{D}{B}$. Thus for a given ratio $\frac{D}{B}$ the shift of the Dirac point from midgap remains fixed, even when D and B become vanishingly small. In fact, the whole edge dispersion remains fixed



(a) (b) (c)

FIG. 3. Dispersions for (a) the extended LKMB model and (b) the KT model. The regions inside the boxes near the zone center are depicted in (c), where they are compared with the standard BHZ model. In (a) and (b) extended states are shown in blue and evanescent states in black. In (c) different colors are used to distinguish between the three models, which are labeled accordingly, with black uppermost for the evanescent states of the extended LKMB model. Parameters for the extended LKMB and KT models are given in the Supplemental Material [63]. Values for the standard BHZ and KT models are very close to those used by Krishtopenko and Teppe in Ref. [43]. In (a) and (b) the different subbands are identified at zero wave vector in gray. Note the break in the wave vector scale in (b).

(see Sec. III). This kind of behavior is not physical, because remote states, which are far away in energy cause the same shift as when they are much closer.

B. Failure of OBCs

Figure 2 shows the extended and evanescent states, in blue and black, respectively, calculated from $H'_{4\times4\uparrow}(\mathbf{k})$ with $k_x = 0$, for the case of a symmetric inverted band gap where $M = -0.0075\,\text{eV}$, $A = Q_1 = Q_2 = Q = 3.83\,\text{eV}\,\text{Å}$, $Q_3 = 0$, and $\Delta_1 = \Delta_2 = \Delta = 0.15\,\text{eV} = -20M$. In addition to the middle states with a small decay parameter connecting bands $|2\rangle$ and $|3\rangle$, as discussed in Sec. I for the BHZ model, there are now physical evanescent states with a large decay parameter connecting bands $|1\rangle$ and $|4\rangle$. For nonzero Q_3, there is a change in the magnitude of this decay parameter, which is demonstrated at zero energy by the two black circles, solid for negative Q_3 and open for positive Q_3. For all values of Q_3, however, the middle states are totally unaffected. Since Q_3 is typically very small and in any case makes no contribution to the BHZ Hamiltonian, it will henceforth be set to zero.

Superimposed in gray in Fig. 2 are the solutions of the BHZ Hamiltonian with the same M and A values as for the multiband case, and with $B = 102.9\,\text{eV}\,\text{Å}^2$, $D = 0$, calculated from the M, Q, and Δ values using Eq. (4). It can be seen in Fig. 2(a) that the two dispersions correspond very well for the middle states and for the conduction and valence band edges out to $k_y = 0.007 \times \frac{2\pi}{a_{\text{sub}}}$, which is consistent with the valid range of wave vectors for the BHZ model, $|k_y| < k_{\text{max}}$, discussed in Sec. I. However Fig. 2(b) shows that the two dispersions are completely different in the vicinity of the wing states. This highlights the spurious nature of the BHZ wing dispersion, even when $D = 0$. The two models only agree near imaginary wave vector $k_y = i\sigma_w(0)$ (with perfect agreement in the limit $M \to 0$), demonstrating that the spurious branch is a phantom like dispersion associated with bands $|1\rangle$ and $|4\rangle$ that have been eliminated. It cannot merge with these band edges as in the LKMB model, because they are no longer there, so adopts a meaningless trajectory towards the Brillouin zone boundary. It should be emphasized that wing solutions with similar behavior also exist in the tight binding version of the BHZ Hamiltonian, which apparently spans the full Brillouin zone [7,8]. In all cases, the key factor is the reduced number of bands.

The zero-energy spin-up eigenvectors of the middle and wing solutions of the BHZ model are both $\frac{1}{\sqrt{2}}[1, -1]$, which enables their combination into an OBC wave function at the Dirac point, as discussed in Sec. I. Such combination is unfortunate, because a proper description of the wing solution should contain significant amplitudes from the absent states, $|1\rangle$ and $|4\rangle$. This can be seen in the zero energy eigenvectors of the LKMB model, which are calculated in Sec. II D and given in the first two columns of Table II for the parameters used in Fig. 2, namely $[i0.037, -0.706, 0.706, i0.037]$ and $[i0.486, -0.513, 0.513, i0.486]$. These eigenvectors contain contributions from the relevant bands and represent the different physical nature of the two states correctly. However, in consequence, they can no longer be combined into a wave function, which satisfies OBCs, showing that there is a fundamental problem with the OBC approach.

In the next subsection an extended version of the LKMB Hamiltonian is developed with a larger basis than in the toy model, to give a better correspondence to the real band dispersions in a HgTe/CdTe QW. This shows that the absence of an OBC solution is quite general and does not depend on the number of basis states included in the LKMB Hamiltonian.

C. Extended LKMB Hamiltonian

Krishtopenko and Teppe (KT) have recently used an extended version of the BHZ model in which E_2 and H_2 states of opposite spin are added to each "spin" block, with k-quadratic interactions with the E_1 and H_1 states [43]. The KT model is able to reproduce the in-plane camelback dispersion in the valence band of HgTe/CdTe TI QWs quite well [10,43]. Krishtopenko and Teppe also propose that a more realistic edge state dispersion can be obtained by applying OBCs to this extended BHZ model. In this subsection, the KT model is compared with an extended version of the LKMB Hamiltonian, developed in the Supplemental Material of this paper

[63], which has 6 states in each spin block. The "spin-up" block contains the 4 states of the KT model, $E_1' \uparrow$, $E_2' \downarrow$, $H_1 \uparrow$ and $H_2 \downarrow$, and also the first two light hole states, $L_1' \downarrow$ and $L_2' \uparrow$. The primes indicate that $E_1' \uparrow$ is a mixture of $E_1 \uparrow$ with $L_2 \uparrow$, and $E_2' \downarrow$ is a mixture of $E_2 \downarrow$ with $L_1 \downarrow$, which is a result of the quantum confinement [43,64]. As in the previous subsection, the key difference between the two models is that the extended LKMB model contains only k-linear terms, while light holes and other more remote states have been eliminated in the KT model using Eq. (2), leading to the introduction of k-quadratic terms.

Figures 3(a) and 3(b) compare the dispersions for the two extended models over a wide range of energy and wave vector. The regions near the zone center in the vicinity of the band gap and marked by the boxes are expanded and superimposed in Fig. 3(c), where they are compared with the dispersion of the standard BHZ model. In the expanded plot, it can be seen that both extended models reproduce the anticrossing between the E_1'- and H_2-valence bands quite well, which leads to the camelback in the KT model at a wave vector of about $k_y \simeq 0.045 \times 2\pi/a_{\rm sub}$. The extended LKMB model shows the correct curvature but does not produce a camelback. At $|k_y| > k_{\rm max}$, the standard BHZ model deviates quite strongly from the extended models, which both extend the useful range beyond the anti-crossing. However, due to their limited basis they still diverge quite significantly from a full 8 band Kane model, especially outside the boundaries of Fig. 3(c) (see Fig. S1 within the Supplemental Material [63]). Although the KT model reproduces the camelback in the valence band of the Kane model rather better than the extended LKMB model, the reverse is true for the two conduction sub-bands, where the KT model shows strong non-parabolicity that is absent in the other models.

In Fig. 3(c) the "middle" gap states are superimposed, using different colors for the three models with black for the extended LKMB model uppermost, and all are virtually indistinguishable. This is because, as shown in Sec. II B, they are close to the zone center and their eigenvectors contain almost no contributions from bands other than the fundamental E_1' and H_1 bands that form the TI band gap. This is in contrast to the wing states, which extend further from the zone center and show the same distinguishing features as discussed earlier for the simpler models. The k-quadratic KT model in Fig. 3(b) still has a spurious wing dispersion, which does not connect with any of the bulk band edges. Moreover, it has an imaginary wave vector in the band gap region that is well beyond the boundary of the first Brillouin zone. As demonstrated in Ref. [43], OBCs can be satisfied by combing this spurious wing solution with the middle solution, just as in the standard BHZ model. This again leads to an unphysical wave function, as discussed above. In contrast, the extended LKMB model in Fig. 3(a) has a physical wing dispersion with large eigenvector components from the higher energy bands with which it connects including E_2', so it cannot satisfy OBCs. These results confirm the conclusions from the simpler toy model of Sec. II B, namely that OBCs only work when remote states are eliminated and a spurious wing dispersion is created with unphysical eigenvector components. Extending the BHZ model therefore does not lead to a more realistic description of the edge states when OBCs are used.

In the rest of this paper, edge states are developed using SBCs applied only to the physical middle solutions, which agree in all three of the models shown in Fig. 3(c) including the standard BHZ model. They are all so close to the zone center that extending the BHZ model does not lead to a more accurate result. Since the middle solutions of the standard BHZ model also agree with those of the LKMB toy model discussed in Sec. II, the toy model will be used in the final part of this Section to develop a simple analytical treatment of LKMB edge states using SBCs. It will then be shown in Sec. III, that these edge states are entirely consistent with the SBC edge states derived for the BHZ model. This is true even though the SBCs used in the two cases are different. In the LKMB model, only wave function continuity is required, while in the BHZ model an additional condition exists for the wave function derivative.

D. SBC edge states of the LKMB model

Although the rest of this paper deals mainly with BHZ edge states, the remainder of this section is dedicated to an SBC treatment of the basic LKMB model (with $Q_3 = 0$, as discussed in Sec. II B). Although a toy model with a limited number of states, it still provides a useful insight into the behavior of the BHZ model, to which it is related through Eq. (4).

With SBCs, the wall region is treated explicitly, because the edge state wave functions decay on both sides of the boundary. Since the eigenvectors of the middle and wing solutions are different in the LKMB model, it is necessary to ensure continuity of the two vector components independently. This is entirely consistent with the separation of the middle and wing solutions in the BHZ treatment discussed below, and leads to similar results. In contrast to the OBC result, the Dirac point always remains very close to midgap, even when the semiconductor band dispersion becomes asymmetric. It is assumed that the wall has a large fundamental band gap $2M_0$ that is symmetrically disposed about that of the semiconductor. How this is realized in practice, and the consequences of when this is not the case are discussed in Sec. V. It turns out that there are only exponential edge state solutions when the wall parameters have a certain relationship with those in the semiconductor. Since the number of eigenvalues is conserved when some system parameter is varied, solutions exist for other combinations of wall and semiconductor parameters but they are no longer exponential. An exponential solution exists, however, when $M_0 \gg |M|$ and the wall band gap is effectively infinite, so this is the limit that is used in both the LKMB and BHZ models. It can be reached by a trajectory in parameter space that involves only exponential solutions (as in this paper) or otherwise. At the same time, a band gap that is truly infinite results in an unphysically rapid decay of the wave function in the wall region. Typical band-gap values for the wall based on real physical systems are discussed in Sec. V.

Two specific cases are now treated for the LKMB spin up model. The simplest case is for a symmetric semiconductor band structure. Next, band asymmetry is included. In both treatments, the same values are used for the electron-hole hybridization parameters in the semiconductor and the wall, and it is shown that this always puts the Dirac point exactly at midgap, even when the band structure is asymmetric. The multiband spin up treatment where these parameters are different is discussed briefly in Sec. IV, after the corresponding results are derived for the BHZ model in Sec. III.

Table I gives expressions for the pair of zero-energy decay parameters $\sigma_\pm$ in the y direction and the corresponding eigenvectors $[a, b, c, d]_\pm$ for the symmetric spin up LKMB model, expressed in terms of the following quantities:

$$S_\pm = A^2 + 2\frac{M}{\Delta}Q^2 \pm \mathrm{sgn}(\Delta)A\sqrt{A^2 + 4\frac{M}{\Delta}Q^2}, \tag{5a}$$

$$R = 1 - 2\frac{M}{\Delta}\frac{Q^2}{S_\pm}, \tag{5b}$$

$$N = \sqrt{\left(Q^2 + \frac{S_\pm}{2}\right)\left(\frac{R^2}{A^2} + \frac{2}{S_\pm}\right)}. \tag{5c}$$

The parameter θ has a value of +1 for the semiconductor and –1 for the wall. In contrast to the BHZ model, where boundary conditions exist for both the wave function and its derivative [10,55] the only boundary condition in the LKMB model is the continuity of the wave function, because $H'_{4\times4\uparrow}(\mathbf{k})$ is linear in k_y. For an edge at $y = 0$, this means that edge solutions must have the same eigenvectors on each side of the boundary. Since the zero-energy eigenvectors in Table I only depend on the ratio of the two band-gap parameters, M and Δ, it is always possible to match the solutions with small or large decay parameters, independently, for a given A and Q, provided the band gap parameters in the semiconductor and wall have the same ratio. Examples are shown in Table II for the same parameters used in Fig. 2, with $\frac{\Delta}{M} = \frac{\Delta_0}{M_0} = -20$, where $|2M_0|$ and $|2\Delta_0|$ are the band gaps between the inner and outer bands, respectively, in the wall material. Note that Δ_0 is negative, because M and M_0 have opposite signs. In the Table, the eigenvectors for the small decay parameter (large decay parameter) on each side of the boundary are written with a small (large) typeface, so that the correspondence between the eigenvectors in each group can be seen clearly. Inserting $\frac{\Delta}{M} = \frac{\Delta_0}{M_0}$ into Eq. (4) gives $B_0 = B\frac{M}{M_0} < 0$ for the BHZ model. It was shown in previous work that for this value of B_0 there are two exponential edge states corresponding to the same matching of solutions with either a small or a large decay parameter, respectively [10,55]. Both models thus obey the same condition for the two exponential edge solutions. Moreover, $B_0 \to 0$ or $\Delta_0 \to -\infty$ as $M_0 \to \infty$, both of which are realistic descriptions of a hard wall.

TABLE I. Zero energy decay parameters and corresponding eigenvectors for the spin up LKMB Hamiltonian with a symmetric band structure, where $\Delta = \Delta_1 = \Delta_2$ and $Q = Q_1 = Q_2$. The functions S, R and, N are defined in Eq. (5).

Decay parameter	Eigenvector
$\sigma_\pm = \theta\,\mathrm{sgn}(\Delta)\frac{\Delta}{Q^2}\sqrt{\frac{S_\pm}{2}}$	$a = \theta\,\mathrm{sgn}(\Delta)\frac{i}{N}\frac{R}{A}\sqrt{\frac{S_\pm}{2}}$ $b = -\theta\,\mathrm{sgn}(\Delta)\frac{Q}{N}\sqrt{\frac{2}{S_\pm}}$ $c = \frac{Q}{N}\frac{R}{A}$ $d = \frac{i}{N}$

Positive Δ_0 solutions (corresponding to positive B_0 in the BHZ model) are shown in the last column of Table II, where there is only fairly close agreement between eigenvectors for the small decay parameters on each side of the boundary. The values for b, c on each side are almost identical while a, d although of opposite sign, are very small. Presumably a small deformation of the wave function might lead to a perfect match. This can only be checked using numerical methods, since the deformed wave function will no longer be exponential. In contrast, the solutions with large decay parameters have completely different eigenvectors in the semiconductor and the wall, with roughly equal magnitudes for all four components, but where the ratios b/c and a/d have opposite signs on each side of the boundary. The solutions with large decay parameters are thus unlikely to satisfy wave function continuity under any circumstances. These results appear to be consistent with the single edge state predicted from the Chern numbers in the spin up BHZ model. Although there is only one nonexponential solution in this model when B_0 is positive, it must converge to the same physical solution as for negative B_0 when the wall band gap is infinite in each case and $|B_0| \to 0$ [55]. The wing solutions for negative B_0 or Δ_0 are thus rejected as unphysical. This point is also demonstrated by the diagram in Fig. 4, where a wall layer with positive Δ_0 ($= \Delta'_0$) is sandwiched between the semiconductor and another wall layer with negative Δ_0 ($= \Delta''_0$). As these parameters become very large, $|\Delta'_0| = |\Delta''_0| > 4 \times 10^7$eV, only the eigenvectors of the wing solutions (listed below the dashed wave function) are equal at the interface between the two wall materials, while only the eigenvectors of the middle solutions (listed below the solid wave function) are close in value at the interface with the semiconductor. If the central layer thickness is expanded to infinity, then only

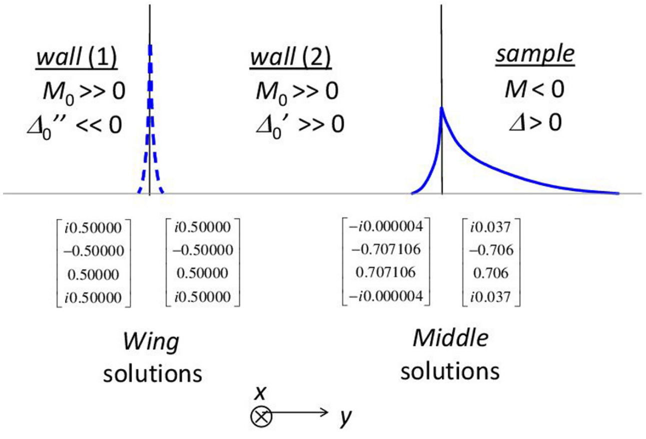


FIG. 4. Schematic depiction of Dirac point edge state wave functions at a compound, double layer wall, based on the middle (solid) and wing (dashed) solutions of the symmetric spin up LKMB Hamiltonian. The inner band-gap parameter of the wall $M_0 \gg |M|$ is constant and positive, while the outer band-gap parameter changes sign in each of the two wall layers, with values $\Delta'_0 \gg 0$ ($\Delta''_0 \ll 0$) closest (furthest) from the semiconductor. Similar zero-energy eigenvectors on each side of an interface are listed below the diagram for the semiconductor and wall parameters listed in Table II, except that $|\Delta'_0| = |\Delta''_0| > 4 \times 10^7$ eV.

TABLE II. Zero-energy decay parameters and eigenvectors for the spin up symmetric LKMB Hamiltonian at $k_x = 0$, with two types of wall in which Δ_0 has opposite signs. Middle (small type face) and wing (large type face) solutions are listed below the material parameters. The equivalent BHZ results are given in the last four rows. The solutions are calculated for $M = -0.0075$ eV and $M_0 = 200$ eV, but the middle or wing eigenvectors in the wall and semiconductor are equal for any M_0 when $\Delta_0 < 0$ and $\Delta/M = \Delta_0/M_0$. When $\Delta_0 > 0$, they are similar (dissimilar) for the middle (wing) solution, implying that only a single nonexponential middle solution exists.

Parameter	Semicond.		Wall (–ve Δ_0)		Wall (+ve Δ_0)	
	LKMB model					
A, A_0 (eV Å)	3.83		3.83		3.83	
Q, Q_0 (eV Å)	3.83		3.83		3.83	
Δ/M, Δ_0/M_0	−20		−20		50	
$\sigma_\pm$, $\sigma_{0\pm}(\times\frac{2\pi}{a_{sub}})$	0.0020	0.0360	−53.48	−959.7	−49.68	−2582.6
a	i0.037	i0.486	i0.037	i0.486	−i0.014	i0.505
b	−0.706	−0.513	−0.706	−0.513	−0.707	−0.495
c	0.706	0.513	0.706	0.513	0.707	−0.495
d	i0.037	i0.486	i0.037	i0.486	−i0.014	−i0.505
	BHZ model					
B, B_0(eVÅ^2)	102.9		−0.00386		0.00144	
$\sigma_\pm$, $\sigma_{0\pm}(\times\frac{2\pi}{a_{sub}})$	0.0020	0.0341	−53.65	−908.9	−49.7	−2633.3
a	0.707	0.707	0.707	0.707	0.707	0.707
b	−0.707	−0.707	−0.707	−0.707	−0.707	0.707

the single-edge state based on the small decay parameters of the middle solutions remains at the semiconductor boundary, while if the central layer thickness is reduced to zero, there are two edge states localized at an interface between the semiconductor and a wall with negative Δ_0, which is the situation already discussed above. This provides additional support for the conclusion that a wall with negative Δ_0 supports both the physical and spurious edge states [65]. Figure 4 is analogous to the equivalent treatment for the BHZ case, in Fig. 3 of Ref. [55].

Exponential solutions can also be found for the zero-energy edge states when $\Delta_1 \neq \Delta_2$, and/or $Q_1 \neq Q_2$, and the band structure is asymmetric. In the LKMB model, the ratio of the outer to inner band energies for a wall with hybridization parameters A_0, Q_{10}, and Q_{20}, that correspond to a particular eigenvector $[a, b, c, d]$ at energy E, are given as a function of k_x by Eq. (A2) in Appendix A. When the hybridization parameters on each side of the boundary are equal, namely $Q_1 = Q_{10}$, $Q_2 = Q_{20}$ and $A = A_0$, the expressions for the eigenvector components in the right hand column of Table IV of Appendix A can be used in Eq. (A2) with $k_x = 0$ to show that the zero energy eigenvectors in the semiconductor and wall are equal, provided $\frac{\Delta_{10}}{M_0} = \frac{\Delta_1}{M}$ and $\frac{\Delta_{20}}{M_0} = \frac{\Delta_2}{M}$, consistent with the symmetric case discussed above. Therefore, as in the symmetric case, solutions with either small or large decay parameters, can be matched independently, when the outer band parameters in the wall, Δ_{01} and Δ_{02}, are negative. Note that in the BHZ model, the semiconductor band asymmetry is defined by the B and D parameters. Eq. (4) shows that a given asymmetry can be achieved by various combinations of the semiconductor Q and Δ parameters in the LKMB model. In all these cases, the Dirac point does not move from gap center when the hybridization parameters in the semiconductor are equal to those in the wall, in good agreement with the BHZ model when SBCs are used, but in complete contrast to the large shift observed with OBCs. The BHZ model is discussed in the next section, where the effect of unequal hybridization parameters in the wall and semiconductor is now considered ($A_0 \neq A$).

III. SBC EDGE STATES OF THE BHZ MODEL

It has been shown previously [10,15] that in the band gap region of the four band BHZ Hamiltonian, the two exponential decay parameters for each 2×2 spin block can be expressed in terms of the wave vector parallel to the edge k_x and energy E as

$$\sigma_\pm(k_x, E) = \sqrt{k_x^2 + F \pm \sqrt{F^2 - G}}, \tag{6}$$

where $F = \frac{A^2+2(MB+ED)}{2B_+B_-}$, $G = \frac{M^2-E^2}{B_+B_-}$ and $B_\pm = B \pm D$. At $k_x = 0$, these are just the decay parameters of the middle and wing solutions discussed in Sec. I, i.e., $\sigma_{m,w}(E) = \sigma_\pm(0, E) = \sqrt{F \pm F^2 - G}$. The spin up eigenvectors corresponding to each solution [10] can be set equal to give the following characteristic equation for the OBC edge state energy:

$$\frac{M + E + B_-\left(k_x^2 - \sigma_-^2\right)}{A(k_x + \sigma_-)} = -\frac{A(k_x - \sigma_+)}{M - E + B_+\left(k_x^2 - \sigma_+^2\right)}. \tag{7}$$

Making use of the transformation, $E \to -E$, $D \to -D$, $k_x \to -k_x$ and solving Eq. (7) for the energy yields the well-known OBC dispersion formula for spin up [15,29]:

$$E^{\mathrm{OBC}}_{2\times2\uparrow} = -Ak_x\sqrt{1 - \frac{D^2}{B^2}} - M\frac{D}{B}. \tag{8}$$

As discussed in Sec. II A, the OBC solution is unphysical because it only depends on the ratio of reciprocal mass parameters, and not on their size.

The characteristic equation for both SBC edge solutions, physical and spurious, is derived from the boundary conditions for the wave function and its derivative with negative $B_0, D_0 \to 0$ in the limit $M_0 \to \infty$. It has been reported previously by the author and has the form [55]

$$\frac{M + E + B_-(k_x^2 - \sigma_\pm^2)}{A(k_x + \sigma_\pm)} = \frac{\sigma_\pm D - \sqrt{\sigma_\pm^2 D^2 + A_0^2}}{A_0}. \tag{9}$$

Equation (9) agrees with the treatment by Raichev [66] for the [001] direction, who found an analytical expression for the special case, $D = B$. In both cases, an exponential solution requires a decay parameter in the wall given by the formula above Eq. (9) in Ref. [10]. Equation (9) can be solved for the energy in an analogous way to the OBC case, yielding a dispersion

$$E^{\mathrm{SBC}}_{2\times2\uparrow} = -Ak_x\sqrt{1 + \frac{\sigma^2 D^2}{A_0^2}} + Dk_x^2 + D\sigma^2\left(\frac{A}{A_0} - 1\right), \tag{10}$$

where σ is the smaller decay parameter in the semiconductor corresponding to the physical solution. Strong hybridization is assumed, as in the previous Section, where the decay parameter is real (the weak case is discussed below). As demonstrated in Ref. [10], $\sigma = \sigma_-$ ($\sigma = \sigma_+$) when $D < B$ ($D > B$). When $D > B$, the σ_- solution in Eq. (6) is imaginary and corresponds to the spurious gap solution with a large real wave vector discussed in Sec. I. Note that for spin down, the sign of the first term reverses in both Eqs. (8) and (10).

For $D < B$, a numerical solution of Eq. (9) was used previously to demonstrate that the σ_+ solution has a larger phase velocity than the σ_- solution and does not merge smoothly with the bulk band edges (see for example, Fig. 2(a) in Ref. [10]). Since this decay parameter is equal to the wing solution when $k_x = 0$, this was taken as further evidence of the spurious nature of the edge state based on the larger decay parameter. On the other hand, the physical solution based on the smaller decay parameter does merge smoothly with the bulk band edges, at which point $\sigma \to 0$. If the wave vector at which the bulk and edge states merge is k_x^m, then Eq. (10) shows that the merging energy is $E^{\mathrm{SBC}}_{2\times2\uparrow} \xrightarrow{\lim \sigma \to 0} -Ak_x^m + D(k_x^m)^2$. Comparing this energy with the bulk dispersion, given in Eq. (6) of Ref. [10], yields $k_x^m = \pm\sqrt{\frac{-M}{B}}$. Note that substituting the merging energy and $k_x = k_x^m$ into Eq. (6) yields $\sigma = 0$, as required.

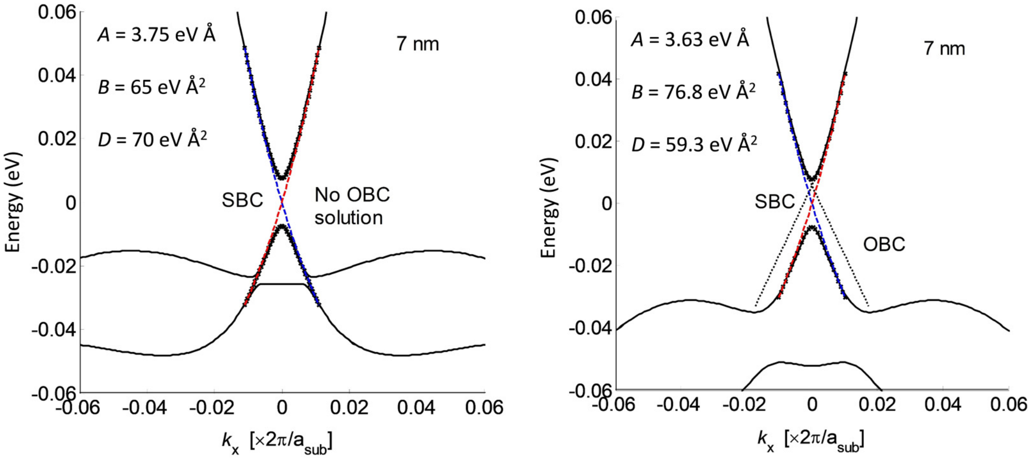
Equation (10) agrees with the result of the LKMB model discussed at the end of the previous section, namely that when the hybridization parameters in the semiconductor and wall are equal, the Dirac point is at midgap, regardless of the degree of band asymmetry. On the other hand, it was previously pointed out that it is unlikely that these parameters are exactly equal [55]. This is confirmed in Sec. V where a hybridization parameter is estimated for the wall. Equation (10) then puts the Dirac point at $E^{\mathrm{D}}_{2\times2\uparrow} = D(\frac{A}{A_0} - 1)\sigma_{\mathrm{D}}^2$, which is slightly below midgap when A_0 is typically slightly larger than A. The physical zone center decay parameter σ_{D} can be evaluated by solving Eq. (6) with $k_x = 0$ and $E = E^{\mathrm{D}}_{2\times2\uparrow}$. In the limit $A_0 \gg A$, it turns out to be independent of D, and is then given by

$$\sigma_{\mathrm{D}} = \frac{A}{2B} - \sqrt{\frac{A^2}{4B^2} + \frac{M}{B}}, \tag{11}$$

which is the same expression as for the $D = 0$ case, given in equation (3b) of Ref. [10]. Since the Dirac point is always close to midgap where the decay parameter varies slowly with energy [see Fig. 1(c)], Eq. (11) should be a good approximation for any value of $A_0 \geqslant A$.

For equal hybridization parameters in the TI and wall, the band gap ratios, $\frac{\Delta_{i0}}{M_0} = \frac{\Delta_i}{M}$, at the end of Sec. II can be substituted into Eq. (4) to show that $B_0 = \frac{M}{M_0}B$ and $D_0 = \frac{M}{M_0}D$. Substituting these relations into the expression for $k_{\max}$ in Sec. I yields the valid range of wave vectors in the wall, $k_{\max,0} = \frac{M_0}{|M|}k_{\max}$. In a wall where M_0 is effectively infinite and the physical decay parameter is comparable to the size of the Brillouin zone, its value can be estimated from $\sigma(0,0)$ in Eq. (6) in the limit of vanishing B_0, D_0, giving $\sigma_{0-} \simeq -\frac{M_0}{A}$. While it is clear that the SBC wave function in the TI is consistent with perturbation theory, to show that this is also the case in the wall it is required that $|\sigma_{0-}| < k_{\max,0}$, i.e., $\frac{M_0}{A} < \frac{M_0}{|M|}\sqrt{\frac{|M|}{B+D}}$. This expression can be rearranged to give $\frac{A^2}{B^2} > \frac{|M|}{B}$, which is always true in strongly hybridized TIs, where σ_{D} in Eq. (11) is real. Thus, the physical SBC edge state does not violate perturbation theory. In contrast, since the wing decay parameter in the TI is greater than or equal to $\frac{A}{B}$ [52,55], this inequality also proves that the wing solution is outside the valid range, as already pointed out in Sec. I. While it is sometimes argued that the inclusion of the wing solution in the OBC wave function is benign [42], especially when it decays over several lattice spacings [47,49], this shows that it cannot be included without violating perturbation theory. Further anomalies related to OBCs are discussed in Appendix B.

Before comparing the Dirac point predicted above for $A_0 \neq A$ with the LKMB result, a few more properties of the newly derived dispersion relation in Eq. (10) are discussed, all of which have been confirmed previously from a direct solution of Eq. (9), [10,55]. First, in strongly hybridized TIs such as HgTe/CdTe, it was noted above that the edge-state velocity for the spurious solution with $D < B$ is larger than for the physical solution. This is confirmed by replacing the decay parameter in Eq. (10) with $\sigma_+ \gg \sigma_-$, which leads to a larger square root term and hence a larger velocity. Second, in weakly hybridized InAs/GaSb/AlSb, the decay parameters near $k_x = 0$ are complex conjugates. As discussed previously, the physical and spurious wave functions are obtained from real symmetric and antisymmetric combinations of these two solutions [67]. However, it was noted in Ref. [55] that the characteristic equation in this regime does not have an exact solution when $D \neq 0$, although the error is small. This is now understood from Eq. (10), where the energy $E^{\mathrm{SBC}}_{2\times2\uparrow}$ is no longer real when $\sigma_\pm^2 D^2/A_0^2$ is complex. The absence of an exact solution when D is finite shows that the edge state wave function is no longer purely exponential, and numerical methods must be used to obtain a precise solution. Note also, that for the exponential case with $D = 0$, SBCs give $B_0 = \frac{M}{M_0}B$



(a) (b)

FIG. 5. 8-band Kane model of the band gap region of a "7 nm" (67Å) HgTe/CdTe QW plotted along [1,0,0], calculated with the parameter sets listed in Table SIII of the Supplemental Material [63] based on (a) Ref. [10] and (b) Ref. [43] ($a_{\mathrm{sub}} = 6.479$Å). The bulk BHZ bands are superimposed as crosses and plotted to the merging points with the SBC edge states, which are shown as dashed lines (BHZ parameters are indicated on the plots, and $A_0 = A$). The BHZ band gap parameter in each case is $M = -0.0075$ eV. In (a) there is no OBC edge solution, because $B<D$, while in (b) the OBC edge solution is plotted as dots.

and $|\mathrm{Re}(\sigma_0)| = \frac{A}{2|B_0|}$ when $A_0 = A$ [10]. The condition for a complex TI decay parameter is $\frac{A^2}{4B^2} < \frac{|M|}{B}$, yielding $|\mathrm{Re}(\sigma_0)| < k_{\mathrm{max},0}$ and confirming, also for weak hybridization, that the SBC wave function in a wall with effectively infinite M_0 is still consistent with perturbation theory. For an exponential solution with $A_0 > A$, the wall Hamiltonian is non-Hermitian except in the infinite limit, when it can again be shown that the solution is consistent with perturbation theory.

An important distinction between strongly and weakly hybridized TIs based on HgTe/CdTe and InAs/GsSb/AlSb QWs, respectively, is that the valence band in the second case has a strong splitting due to structural inversion asymmetry (SIA). Based on perturbation arguments, it was previously argued that the edge state dispersion is essentially unaffected by SIA, so that it merges with the edge of the unsplit valence band [67]. The Dirac point is then close to, or even hidden below, the split valence band edge. This can be understood more directly by noting that because the valence band profile is not symmetric in the growth or z direction, it behaves like an effective electric field $\mathbf{F} = \hat{\mathbf{z}}F_z$. Ignoring k-independent and weaker anisotropic contributions, the leading Rashba term in the valence band behaves as $H_R \propto \mathbf{J} \cdot \mathbf{k} \times \mathbf{F}$, where $\mathbf{J}$ is the total angular momentum [68]. This leads to split valence band energies for total angular momentum aligned parallel or antiparallel to the tangent of a circle in the x-y plane [68]. In contrast, the edge states move only in the x direction with total angular momentum components quantized along z (see Table I of Ref. [10]) so they have no Rashba interaction.

To illustrate the results of this section and compare with a realistic dispersion model, Fig. 5 shows the bulk band structure of a HgTe/CdTe QW with a width of 7 nm, calculated with an 8-band Kane model [10] according to the two different sets of model parameters listed in Table SIII within the Supplemental Material [63], which are based on previous work by the author [10] and the set used by Krishtopenko and Teppe [43], respectively. In the second case, modified Luttinger values are given [69], which have been converted into the standard values using the relations in Ref. [70]. The camelback discussed earlier in Sec. II C can be seen clearly, where it was shown to arise from an anti-crossing between the second heavy-hole band H_2 and the E_1' valence band edge. In the second case the H_2 band lies deeper in energy and the anticrossing is larger, although both of these properties have been shown to be quite sensitive to the lattice parameter of the substrate [43]. Bulk BHZ bands have been fitted to the Kane bands according to the BHZ parameters given on the plots, which are also similar to the values given in each reference. In Fig. 5(a) it can be seen that the band asymmetry is such that $D > B$ and there is therefore no OBC solution, as discussed above. On the other hand, SBC solutions exist in both cases, and these are plotted with the bulk BHZ bands up to the merging points. They occur within the small range of wave vectors allowed by perturbation theory, $|k| < k_{\mathrm{max}} \simeq 0.008 \times 2\pi/a_{\mathrm{sub}}$, and evolve smoothly as the Luttinger and energy parameters are varied. The differences in the two sets of parameters are actually quite small, the main variation being for the Luttinger parameters, γ_1, γ_2, γ_2, and energy E_{P}. As described in Ref. [71], the method for determining the parameters used in Fig. 5(a) has been shown to work very well for narrow band gap superlattices of various types, even when strong band bowing is present. These parameter values are also much closer to the values given in the widely cited paper by Lawaetz [62], which are also listed in Table SIII [63]. This is just to demonstrate that both band structures in Fig. 5 correspond to a realistic TI with an inverted band gap, so the existence of their edge states should not depend on the particular choice of Luttinger parameters. Moreover, for a smooth evolution of the bulk dispersion from Fig. 5(b) to Fig. 5(a), the OBC edge states vanish at the point where their phase velocities decrease to zero, when they no longer merge with the valence band edge, even far beyond the Brillouin zone boundaries (see also Fig. 2 in Ref. [55]).

The extension of the edge states into the camelback region and beyond before vanishing completely, when the bulk band structure is slightly deformed, highlights the unphysical nature of the OBC solution, as already predicted from the LKMB model in Sec. II. In contrast, Fig. 5 shows that the standard BHZ model provides a consistent description of the edge states when SBCs are used. The failure of OBCs shows that it is necessary to take the wall into account. In the next section, the sensitivity of the SBC dispersion to the wall hybridization parameters is compared for the LKMB and BHZ treatments, and this is followed in Sec. V by a discussion of physical models for the wall, consistent with the BHZ Hamiltonian.

IV. SENSITIVITY TO WALL HYBRIDIZATION

In this section the energies of the Dirac point are compared for the BHZ and LKMB models, when one or more of the electron-hole hybridization parameters in the semiconductor are unequal to those in the wall. As shown in the previous section, when $A_0 > A$, the BHZ model predicts a negative Dirac point energy of $E^{\mathrm{D}}_{2\times2\uparrow} = D(\frac{A}{A_0} - 1)\sigma^2_{\mathrm{D}}$, where σ_{D} is given to a very good approximation by Eq. (11).

Since the Dirac point is no longer at $E = 0$, it is not a simple matter to find an analytical solution for its energy in the LKMB model. Instead a numerical solution can be performed based on Eq. (A2), with $k_x = 0$. By inserting the eigenvector for the middle solution of the semiconductor at energy E into Eq. (A2), and using the predicted ratio of the outer to inner band energies in the wall to calculate its middle solution at the same energy, the energy of the Dirac point $E^{\mathrm{D}}_{4\times4\uparrow}$ is found when the wall and semiconductor eigenvectors are equal. The results are shown in Table III for typical semiconductor parameters and two different values of the wall hybridization parameter A_0. The effect of changing the secondary wall hybridization parameters Q_{10} and Q_{20} is also studied. The calculation was performed such that each of the eigenvector components in the semiconductor and wall agree to better than 0.001%. Results for the BHZ model are also shown in the lower part of the Table, for comparison.

In the LKMB model there is a dependence on the secondary hybridization parameters and almost perfect correspondence exists with the BHZ model when the wall values are very small (not shown in Table III). For the parameter values shown in the Table, the correspondence between models is reasonable, with agreement for the energy of the Dirac point to better than 1.0 meV or just a few percent of the TI band gap. Note that the Dirac point even has a small positive energy when all hybridization parameters in the wall are equal. It can be concluded, however, that the shift of the Dirac point from midgap is extremely small in both models, showing that the BHZ model provides a reasonable estimate.

The large wall hybridization parameter of 1973.5 eV Å in the right hand column of Table III corresponds to $A_0 = \hbar c$ in the relativistic Dirac equation. It was previously suggested by the author that this value might be used for a vacuum wall [55]. In the next section, however, it is argued that the relativistic value cannot be justified and an alternative picture is presented. This predicts a smaller value for the hybridization parameter, closer to that in the semiconductor.

TABLE III. The eigenvectors of the spin up LKMB middle solutions in three types of wall with negative Δ_{i0} and $A_0 \neq A$, at the energy of the Dirac point $E^{\mathrm{D}}_{4\times4\uparrow}$ for the semiconductor parameters defined in the first column. The solutions are calculated for $M = -0.0075$ eV and $M_0 = 200$ eV, but the eigenvectors and Dirac point are essentially unchanged for any $M_0 > 1$ eV. The equivalent BHZ results are given in the lower rows. The ratio of the Dirac point energies calculated by the two methods is shown in the last row. The listed outer band energies of the wall were calculated using Eq. (A2). For other values, the solutions are not exponential.

Parameter	Semicond.	Wall (1)	Wall (2)	Wall (3)
		LKMB model		
A, A_0 (eV Å)	3.83	11.0	11.0	1973.5
Q_1, Q_{10} (eV Å)	2.0	2.0	11.0	2.0
Q_2, Q_{20} (eV Å)	2.0	2.0	11.0	2.0
Δ_1/M, Δ_{10}/M_0	−20.0	−6.723	−40.18	−0.037
Δ_2/M, Δ_{20}/M_0	−3.2	−1.066	−6.52	−0.0058
a		$i0.0195$	$i0.0200$	$i0.0193$
b		−0.6971	−0.6769	−0.7016
c		0.7064	0.7268	0.7017
d		$i0.1211$	$i0.1149$	$i0.1225$
$E^{\mathrm{D}}_{4\times4\uparrow}$(eV)		−0.00020	0.00027	−0.00030
		BHZ model		
B, B_0(eVÅ^2)	135.3	−0.153	−0.063	0.0204
D, D_0(eVÅ^2)	107.2	−0.149	−0.047	−0.00032
$E^{\mathrm{D}}_{2\times2\uparrow}$(eV)		−0.00031	−0.00031	−0.00048
$E^{\mathrm{D}}_{4\times4\uparrow}/E^{\mathrm{D}}_{2\times2\uparrow}$		0.629	−0.856	0.633

V. DISCUSSION OF THE BHZ WALL REGION

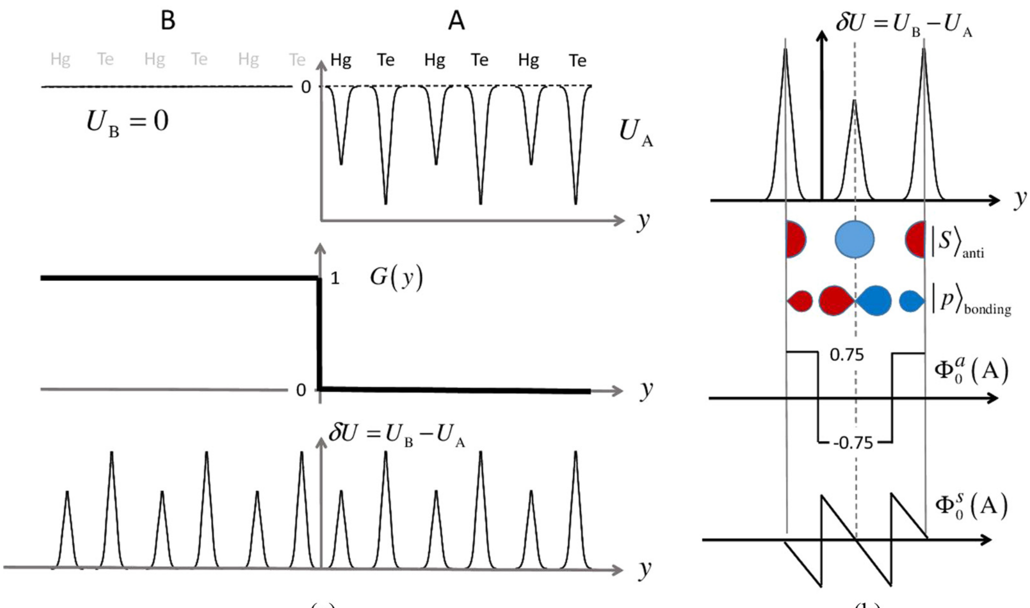
A. k · p treatment of the interface

Although the 2D Dirac Hamiltonian for a free electron is similar in form to the BHZ Hamiltonian, a treatment is required for the wall that is conceptually consistent within a non-relativistic **k·p** framework. In this section an attempt is made to address this issue.

The final term in Eq. (1), was not required for the preceding treatment of the evanescent band-gap states, which are essentially properties of the bulk material. However, this term is required when considering properties related specifically to the sample edge. If we ignore derivative of a delta-function terms, and consider only a single boundary at $y = 0$, then $H^{\mathrm{mod}}_{nn'}(y)$ is given to a reasonable approximation by the following expression:

$$H^{\mathrm{mod}}_{nn'}(y) = \delta U_{nn'}\widetilde{G}(y) - D_{0,nn'}\widetilde{\delta}(y), \tag{12}$$

where $G(y)$ is a step function as shown in the middle panel of Fig. 6(a) and $\delta(y)$ is a Dirac delta function, whereas $\widetilde{G}(y)$ and $\widetilde{\delta}(y)$ are the same functions with Fourier components limited to the first Brillouin zone [51]. While the first pair of functions are mathematically abrupt, the second pair change over a distance of about one monolayer, or $a' = \frac{a_{\mathrm{sub}}}{2}$. The matrix element $\delta U_{nn'}$ is defined as $\langle n|\delta U|n'\rangle$ where $\delta U = U_B - U_A$ and U_A and U_B are the microscopic crystal potentials for materials



(a) (b)

FIG. 6. (a) Schematic microscopic potential of mercury telluride near the edge of a HgTe/CdTe two dimensional QW. The upper panel in (a) shows the microscopic crystal potential U_A in the QW, and $U_B = 0$ is the potential in the wall, where $U_B = U_A + G(y)\delta U$. The step function $G(y)$ is shown in the middle panel and the perturbing potential $\delta U = U_B - U_A$ is depicted in the lower panel. (b) Components $\Phi_0^a(y)$ and $\Phi_0^s(y)$ of the crystal periodic interface function Φ_0 are shown over a distance of one monolayer (width a') for a mathematically abrupt interface, together with the s-antibonding and p-bonding crystal periodic functions, and the perturbation δU. These functions are used to calculate the interface potentials $D_{0,nn'} = \langle n|\Phi_0 \delta U|n'\rangle$.

A and B on each side of the interface. Material A is treated as a reference crystal and δU is thus the perturbation that transforms material A into material B. E_n in Eq. (1) is then a local band edge in the reference crystal, and the perturbation terms $\delta U_{nn'}\widetilde{G}(y)$ in Eq. (12) can be used with the Pikus-Bir formula in Eq. (2) to generate the band edge positions of the local states in the other material. Together with the interface terms $D_{0,nn'}$, which are discussed in detail below, it is possible to describe the local band edges and basis states in materials A or B, and their evolution on passing from one material to the other [58,59].

Figure 6(a) illustrates how materials A and B are defined in the present treatment. Material A represents the TI semiconductor, which is depicted schematically along the y direction in Fig. 6 for mercury telluride in the central plane of a HgTe/CdTe QW grown in the z direction. If U_A is the microscopic crystal potential in the QW, $U_B = 0$ represents the potential in the wall, which is treated as the vacuum or "empty crystal" with the same lattice parameter as material A. Thus $\delta U = -U_A$, which is depicted schematically in the lower panel of Fig. 6(a). Local antibonding s- and bonding p-like band edge states in the semiconductor should thus evolve into empty crystal states of the same symmetry in the vacuum.

B. Four band empty crystal

In the empty crystal, antibonding s- and bonding p-like states with a given spin can be constructed directly from degenerate free electron states with wave vectors, $\frac{\pi}{a'}[\pm 2, 0, 0]$, $\frac{\pi}{a'}[0, \pm 2, 0]$ and $\frac{\pi}{a'}[0, 0, \pm 2]$, normalized to an effective unit cell volume a'^3 as follows [1,60]:

$$u_s = i\sqrt{\frac{2}{3a'^3}}\left[\cos\left(\frac{2\pi x}{a'}\right) + \cos\left(\frac{2\pi y}{a'}\right) + \cos\left(\frac{2\pi z}{a'}\right)\right], \tag{13a}$$

$$u_p = \sqrt{\frac{1}{a'^3}}\left[\sin\left(\frac{2\pi x}{a'}\right) + i\sin\left(\frac{2\pi y}{a'}\right)\right]. \tag{13b}$$

The hybridization parameter in the wall is thus

$$A_0 = -\frac{\hbar^2}{m_0}\langle u_p|\frac{\partial}{\partial y}|u_s\rangle, \tag{14}$$

which yields

$$A_0 = \hbar v_w, \tag{15a}$$

$$v_w = \frac{h}{\sqrt{6}m_0 a'}. \tag{15b}$$

Taking $a' = 3$Å gives a value of $v_w = 9.90 \times 10^5$ m/s. Thus, the hybridization parameter for the wall calculated from Eq. (15b) is 6.5 eV Å, which is much smaller than the free electron Dirac value of 1973.5 eV Å.

C. The need for passivation

Unfortunately, a vacuum wall based on the empty crystal wave functions defined in Eq. (13) does not obey a Dirac-like Hamiltonian, because these wave functions are degenerate with an energy that is much higher than the energies of the band gap states in the semiconductor. Instead the wall behaves for spin up as

$$H^{\text{wall}}_{2\times2\uparrow} = A_0(\sigma_x k_x - \sigma_y k_y) + \sigma_z M_0 + I_{2\times2} M_1 \tag{16}$$

with $M_0 = 0$ and $M_1 = \hbar^2(2\pi/a')^2/2m_0 = 16.7\,\text{eV}$. Although topological edge states are predicted from the change in Chern number or $\mathbb{Z}_2$ index of the semiconductor when the band gap inverts [7,28,38], this wall Hamiltonian does not ensure the existence of a Dirac point. For example, the SBC wave function has no exponential solution in the wall with a real energy at $k_x = 0$.

Treatments that lead to an edge state dispersion with a Dirac point, including those based on OBCs [31], and SBC treatments presented in this paper or elsewhere [14], generally require a wall with a large band gap that overlaps that of the semiconductor. Since this does not occur with a vacuum, a passivation material is required, which should also have band edge states with opposite parity, ideally with *s*- and *p*-like symmetry, so that it behaves for spin up like $H^{\text{wall}}_{2\times2\uparrow}$ with a large value of M_0, and with $M_1 \simeq 0$ (additional quadratic terms, $\sigma_z \mathbf{k}\cdot B_0\mathbf{k} + I_{2\times2}\mathbf{k}\cdot D_0\mathbf{k}$, can also be included). A good candidate appears to be silicon dioxide (SiO_2), which is the material deposited on the delineated bar-shaped sample in at least two cases where the observation of one dimensional edge states has been reported [8,72]. Harrison [73] discusses the band structure of this material, which has a band gap of approximately 9 eV. The upper valence band is composed of states, which have bonding *p*-like symmetry, and in the cubic β-cristobalite phase, they mimic the $(X \pm iY)/\sqrt{2}$ basis of the TI Hamiltonian. The conduction band is composed of an antibonding state with *s*-like symmetry on both the silicon and oxygen sites. The lattice parameter of the cubic unit cell is about 7.1Å, but a tetragonal version exists with *a* and *c* parameters of 5.0 and 6.9 Å [74]. The cubic lattice parameters for HgTe/CdTe or InAs/GaSb/AlSb QWs, are 6.5 Å and 6.1 Å, respectively, which are fairly similar to these values. It is reasonable to suppose that in the ideal case the first few atomic layers of the passivation layer will grow in registration with the TI semiconductor lattice before the SiO_2 structure adopts a more complex lower energy phase. A SiO_2 band gap of $\sim$9 eV corresponds to $M_0 = 4.5\,\text{eV}$. In addition, the difference between the electron affinities of SiO_2 and HgTe or InAs is about $4.3 \pm 0.3\,\text{eV}$ [75–77], so $M_0 \gg M_1$. Since the band-edge states have the same *s*- and *p*-like symmetries as the empty crystal states in Eq. (13), the value of $A_0 = 6.5\text{eV Å}$ calculated above should give a correct order of magnitude. The wave function decay parameter in the SiO_2 is thus approximately $\frac{M_0}{A_0} \simeq 0.7\text{Å}^{-1}$. This corresponds to a decay length of the order of one effective lattice parameter a', which is reasonably consistent with a $\mathbf{k}\cdot\mathbf{p}$ wave function containing only Fourier components in the first Brillouin zone. It also shows that only a few SiO_2 monolayers are required before the wave function has fully decayed, consistent with the structural assumptions made above.

If a passivation layer such as silicon dioxide is not used intentionally, it is still quite possible that a thin native oxide with a large band gap can form at an untreated wall after sample delineation, e.g., mercury oxide and/or tellurium oxide for an HgTe/CdTe QW sample. Even here, the conduction and valence bands may be composed of the *s*-and *p*-like orbitals of the oxygen and semiconductor atoms, so it may still be possible to define a wall Hamiltonian with $M_0 \gg M_1, M$. An additional complication is the possible formation of trivial edge states, whose presence may well depend on the choice of surface treatment and passivation material. For example, bar samples have been fabricated from InAs/GaSb/AlSb QWs by different groups, using silicon oxide or silicon nitride [72,78], and aluminum oxide or hafnium oxide [9]. In the second case there is evidence of edge conduction in the normal phase, suggesting that trivial edge states are present and calling into question whether this is generally the case, or dependent on which passivation material is used.

D. Interface band mixing

The final term of Eq. (12) contains interface potentials $D_{0,nn'}$, which are analogous to the symmetry based surface parameters of Asmar *et al.* [14], and which arise because the orbitals on the boundary layer of atoms experience a different microscopic potential to their immediate neighbors. The interface potentials can be evaluated as $D_{0,nn'} = \langle n|\Phi_0 \delta U|n'\rangle$, where the crystal periodic interface function, $\Phi_0(y) = \Phi_0^a(y) + \Phi_0^s(y)$, is the sum of two components, the first of which has even parity and the other, odd parity, with respect to the boundary atomic plane [51]. In the example in Fig. 6 with an interface to the vacuum at $y = 0$, this is the plane of mercury atoms closest to the origin. The functions, $\Phi_0^{s/a}(y)$, are depicted over one unit cell in the lower part of Fig. 6(b), for a mathematically abrupt interface. For a more realistic interface with a finite width they become significantly dampened, as shown in Fig. 3 of Ref. [51] for an interface with a width of 0.8Å. Also shown in Fig. 6(b) are the crystal periodic perturbing potential, δU, as in Fig. 6(a), and the antibonding *s*- and bonding *p*-like crystal periodic functions of the reference crystal, which are depicted blue when positive and red when negative. The example in Fig. 6 is for a vacuum interface, but the perturbing potential $\delta U = U_B - U_A$ can easily be modified to represent an interface with a passivation material, such as silicon dioxide whose microscopic potential is then represented by U_B. This material is assumed to be periodic for the first one or two monolayers as discussed above, and so can be treated here with perfect periodicity because we are only interested in the interface region over the width of the delta function $\widetilde{\delta}(y)$, which is about one monolayer. Inspection of the symmetries in Fig. 6(b) shows that there will be four finite contributions to $D_{0,nn'}$, namely $D_{0,\text{SS}}$, $D_{0,\text{XX}}$ and $D_{0,\text{YY}}$, which have a finite matrix element with $\Phi_0^a(y)$, and $D_{0,\text{SY}}$, which has a finite matrix element with $\Phi_0^s(y)$. Ignoring electron-light hole mixing [63], the spin up edge state wave function is the product of an eigenvector based on E_1 and H_1 QW states [10]: $|+\rangle \simeq \frac{1}{\sqrt{2}}|iS\rangle\uparrow - \frac{1}{2}(X+iY)\uparrow$ and a *y*-dependent envelope function: $\psi_{\uparrow\sigma(k_x)}(y)$, so the energy shift of the Dirac point due to the interface band mixing is approximately

$$\begin{aligned}\delta E &= \eta|\psi_{\uparrow\sigma(0)}(0)|^2 \\ &\quad\times\left[\frac{2D_{0,\text{SS}} + D_{0,\text{XX}} + D_{0,\text{YY}}}{4} - \frac{D_{0,\text{SY}}}{\sqrt{2}}\right] \\ &= \eta|\psi_{\uparrow\sigma(0)}(0)|^2\overline{D}_0,\end{aligned} \tag{17}$$

in which $\sigma(k_x)$ is the decay parameter of the physical edge state in the semiconductor. The dependence on the squared amplitude of the envelope function at $y = 0$ is due to the delta function in Eq. (12), where $\widetilde{\delta}(y)$ has been replaced by $\delta(y)$. Since the wave function in the wall decays at the about the same rate as $\widetilde{\delta}(y)$, this will lead to an overestimate, so a correction factor η has been introduced into Eq. (17) where $\eta \simeq 0.5$. Note that overlap integrals of QW envelope functions of order unity have not been included explicitly. The value of $\overline{D}_0$ depends on the magnitudes and signs of the four contributions in the square bracket, which can only be estimated using microscopic calculations. Estimates for these interface band mixing potentials at a superlattice interface vary widely, and are typically in the range $0.1 - 2\,\text{eV\AA}$ [79,80]. Noting that $|\psi_{\uparrow\sigma(0)}(0)|^2 = 2\sigma_\text{D}$ where σ_D may be estimated using Eq. (11), and assuming a relatively large value of $\overline{D}_0 = 2\,\text{eV\AA}$, the Dirac point is predicted to shift by $\delta E = 0.004$ eV when $B = 135.3\,\text{eV\AA}^2$ and $D = 107.2\,\text{eV\AA}^2$ (as in Table III). An important point to note is that the shift of the edge state dispersion is largest at the Dirac point and decreases with increasing edge state wave vector, because it is proportional to $2\sigma(k_x)$, which vanishes at the merging points with the bulk band structure. Thus interface band mixing may shift the Dirac point and distort the edge state dispersion, but the merging points will remain fixed.

E. Effect of the D' term

Finally, the significance of the quadratic term proportional to D' in the first line of the LKMB Hamiltonian of Eq. (3) must be considered. In the present paper this term has been ignored because it is quite small. For example, its inclusion causes a virtually imperceptible shift of the middle states in Fig. 2(a), of only $-17\,\mu$eV at midgap reducing to zero at the band edges. Moreover its matrix element with the edge-state wave functions diverges due to a discontinuity in the first derivative of the wave function at the boundary, and to a contribution in the wall, which tends to infinity with increasing wall potential, M_0. Assuming that the energy of the Dirac point varies smoothly with increasing D', this shows that a nonzero value of D' must lead to a nonexponential wave function, with a continuous first derivative and with a more linear mode of decay in the wall. This is consistent with the BHZ model, where inclusion of the quadratic term in Eq. (3) results in the addition of $D' = \frac{\hbar^2}{2m_0}$ to the expression for D in Eq. (4b). As for the multiband case, this has negligible effect on the dispersion of the middle states in the semiconductor. In the wall, however, an exponential solution requires that the band asymmetry parameter varies as $D_0 = \frac{\sigma}{\sigma_0}D$, vanishing when $M_0 \to \infty$ [10]. If instead $D_0 \to D'$ and does not vanish, the solution will again be non-exponential. There will thus be a shift in the edge-state energy in both models, compared with the exponential solutions calculated for $D' = 0$. Based on the small value of D', and in the absence of an exact numerical solution, this shift is assumed to be small.

VI. CONCLUSIONS

The four-band BHZ Hamiltonian provides a simple but realistic description of the band edge states in 2D TIs such as HgTe/CdTe and InAs/GaSb/AlSb. However, the validation of hard wall boundary conditions appropriate to such materials has remained elusive. The most popular choice is OBCs, which avoid any explicit treatment for the wall, while other boundary conditions tend to be phenomenological in nature, so the connection with the microscopic structure of the wall remains unclear. In this paper, OBCs have been ruled out, because they fail when the remote states are included explicitly in the semiconductor Hamiltonian. At the same time, the other boundary conditions show that a wide variety of edge-state dispersions are possible, depending on the values of the phenomenological parameters. Therefore a different approach has been adopted here, based on SBCs, which address apriori both the microscopic properties of the wall, and the conditions that can lead to the existence of a Dirac point.

The use of OBCs leads to contactless edge confinement, and a Dirac point with an unphysical dependence on the TI band parameters. Other unphysical results have also been reported when these boundary conditions are used [55], and it is shown in Appendix B that they lead to an unphysical extension of the edge states into the topologically trivial phase. Although mathematically correct, it is known that one of the two gap solutions for a given spin direction is spurious. Unfortunately, this solution must be combined with the physical gap solution in order to satisfy OBCs, and such combination is only possible because both solutions have the same eigenvector. The spurious solution is related to diagonal k-quadratic terms, which are introduced into the BHZ Hamiltonian when remote states are eliminated using perturbation theory. If the remote states are not eliminated but included in a larger, LKMB Hamiltonian, the eigenvectors of the two gap solutions are no longer equal and an OBC solution is impossible. This is because the spurious solution is replaced by a physical solution with large eigenvector components from the remote states. It has been shown that this also applies to an extended version of the BHZ Hamiltonian that was recently proposed to model the bulk band edges over a wider range, including a camelback in the valence band.

An alternative approach suggested previously by the author is to use SBCs. These boundary conditions match both the wave function and the product of its derivative and a reciprocal mass term at the interface between the semiconductor and the wall, and are obtained by integrating the BHZ eigenstate equation across the boundary region. Unlike most other SBC treatments, which have been applied to a soft wall and which evolve into OBCs when the wall becomes hard, the present SBC approach is for a hard wall and results in wave function confinement with a large amplitude at the edge. This type of confinement is typical of other classic surface phenomena, such as surface plasmons and phonons. The SBC edge state wave function is constructed from just the physical gap solutions on each side of the boundary. It has been verified in the present paper by comparison with a LKMB solution, with which it is in complete agreement. In both cases, the Dirac point has a dependence on the TI band parameters that is physically justifiable.

One of the challenges of the SBC approach is to establish the strength of the electron-hole hybridization in the wall. When the semiconductor and wall have the same midgap energies but different hybridization parameters, the Dirac point is slightly shifted from midgap but rather insensitive to the difference. In this paper, a free electron basis that matches the symmetry of the antibonding *s* states and bonding *p* states in the semiconductor is used to estimate the wall hybridization parameter, which behaves like the free electron Dirac value $A_0 = \hbar c$ but with the speed of light c replaced by a velocity $v_w = \frac{h}{\sqrt{6} m_0 a'}$ in which a' is the monolayer thickness in the semiconductor. The wall hybridization parameter is then about 6.5 eV Å, which is comparable to the semiconductor value in a TI with strong hybridization. An empty crystal Hamiltonian constructed from this basis represents a hard vacuum like wall, but unfortunately exhibits no splitting between the electron and hole states, which are also much higher in energy than those in the semiconductor. In order to ensure an edge-state dispersion with a Dirac point in the BHZ band gap, it appears that a thin passivation layer is required, which may be the native oxide of the material, or an externally deposited dielectric material such as silicon dioxide, for which successful observations of the quantum spin Hall effect have been reported. Silicon dioxide indeed has bands of the correct *s*- and *p*-like symmetry separated by a large band gap that overlaps that of the semiconductor fairly symmetrically. Assuming a pseudomorphic cubic phase immediately next to the semiconductor, the hybridization parameter will be fairly close to the empty crystal value, which can be used without introducing a significant error. The nature of the passivation layer may also be important to avoid the presence of trivial edge states.

Another challenge in the SBC approach is that the mode of decay of the edge state wave function is generally nonexponential. Nevertheless, it is possible to find useful exponential solutions when the wall is effectively infinite, $M_0 \gg |M|$, and the edge state decay parameter is real as in HgTe/CdTe QWs. Although this has been the main focus of the present paper, it was also confirmed that weakly hybridized systems with a complex decay parameter, such as InAs/GaSb/AlSb, exhibit nonexponential behavior when the band structure is asymmetric. Even for weakly hybridized systems with a symmetric band structure, SBCs have been shown previously to exhibit non-exponential behavior in narrow samples, except at certain characteristic widths [55].

The shift of the Dirac point from midgap is generally small when SBCs are used with wall passivation. However, several additional factors could affect its position, including the small quadratic D' term left out of the LKMB Hamiltonian (see Sec. V), the difference in the midgap energies of the semiconductor and passivation materials [$M_1 \neq 0$ in Eq. (16)], and the effect of interface band mixing. The first two factors should not be too significant, especially if a passivation material is used with $M_1 \ll M_0$ (the effect of changing M_1 is shown in Appendix A), but the third factor could be important, producing a shift, which is greatest at the Dirac point and which steadily decreases to zero at the merging wave vectors with the bulk band edges. Microscopic calculations of the interface band mixing potentials are therefore needed to establish the size of this effect for different passivation materials. The position of the Dirac point could be important since it has recently been suggested that a Dirac point hidden below a camelback or SIA split valence band edge may increase the robustness of edge state conduction in the presence of magnetic fields [44,45].

Although the present work is based on the $\mathbf{k} \cdot \mathbf{p}$ approach, it should also be relevant to tight binding models. Fukui [81] shows that based only on nearest neighbor interactions, edge states can be found in graphene for zigzag and bearded edges, where the wave function behaves like the single BHZ solution for SBCs with $B = D = 0$ [12], finite in the crystal and vanishing within one nearest neighbor distance beyond the boundary. In contrast, this does not occur for an armchair edge when next-nearest-neighbor interactions are included, and an OBC approach is proposed based on combining degenerate solutions. This is also the approach proposed for the Wilson-Dirac model [8,24], which is closely analogous to the BHZ Hamiltonian, and where one of the two solutions is spurious.

In summary, the present work has demonstrated that OBCs lead to unphysical behavior and must be replaced by boundary conditions such as SBCs, that properly take the wall into account and allow stronger edge confinement, with a large amplitude at the sample edge. In addition, a passivation layer is often present, either intentionally or due to the formation of a native oxide, and may even be an essential way of ensuring the existence of a Dirac point. For this case, the standard BHZ model with SBCs can provide an adequate description of the edge states in both the TI and the wall without violating its range of validity. Although scattering from edge imperfections is suppressed by time reversal symmetry, a practical consequence of the present hard wall SBC approach is that stronger wave function confinement, combined with significant disorder in the passivation layer, could both lead to a lower threshold for the breakdown of dissipationless transport than previously thought [82].

ACKNOWLEDGMENTS

The author acknowledges useful correspondence with Dr. M. V. Durnev and valuable discussions with Prof. N. H. Lindner.

APPENDIX A: FULL EDGE STATE DISPERSION IN THE TOY LKMB MODEL

The full spin up edge state dispersion, $E(k_x)$ is calculated by finding solutions for $H'_{4\times 4\uparrow}(\mathbf{k})$ with the same eigenvectors on both sides of the boundary. Assuming $D' = 0$, and writing $k_y = i\sigma$, Eq. (3) can be solved for the decay parameter, σ, and eigenvector, $[a, b, c, d]$ in the semiconductor in terms of the energy E and wave vector k_x to yield the relations in Table IV, where

$$Z_{\pm} = (M + E)\Delta_1' Q_2^2 \pm (M - E)\Delta_2' Q_1^2, \tag{A1a}$$

$$N = \sqrt{T_1 T_3^2 + T_2}, \tag{A1b}$$

$$T_1 = 1 + \left(\frac{Q_1(k_x - \sigma)}{\Delta_1'}\right)^2, \tag{A1c}$$

TABLE IV. Decay parameters in the y direction at energy E and edge wave vector k_x, and the corresponding eigenvectors, for the spin up LKMB Hamiltonian ($Q_3 = 0$). The functions, $Z_\pm$, Δ_1', Δ_2', and N are defined in Eq. (A1).

Decay parameter	Eigenvector
$\sigma_\pm = \theta\sqrt{k_x^2 + \frac{\Delta_1'\Delta_2'A^2 + Z_+ \pm \sqrt{\Delta_1'\Delta_2'A^2\{\Delta_1'\Delta_2'A^2 + 2Z_+\} + Z_-^2}}{2Q_1^2Q_2^2}}$	$a = i\frac{Q_1}{AN\Delta_1'}\left(\frac{Q_2^2(k_x^2 - \sigma^2) + (M - E)\Delta_2'}{Q_2(k_x - \sigma)}\right)$ $b = \frac{1}{N}\frac{\Delta_2'}{Q_2(k_x - \sigma)}$ $c = -\frac{1}{AN}\left(\frac{Q_2^2(k_x^2 - \sigma^2) + (M - E)\Delta_2'}{Q_2(k_x - \sigma)^2}\right)$ $d = \frac{i}{N}$

$$T_2 = 1 + \left(\frac{\Delta_2'}{Q_2(k_x - \sigma)}\right)^2, \tag{A1d}$$

$$T_3 = \frac{Q_2^2\left(k_x^2 - \sigma^2\right) + (M - E)\Delta_2'}{AQ_2(k_x - \sigma)^2}, \tag{A1e}$$

$$\Delta_1' = \Delta_1 - E, \tag{A1f}$$

$$\Delta_2' = \Delta_2 + E, \tag{A1g}$$

and $\theta = 1$ in the semiconductor. In the wall $\theta = -1$ and the corresponding parameters are A_0, M_0, Q_{10}, Q_{20}, Δ_{10}, Δ_{20}, and σ_0. The relations in Table IV can be expressed in terms of these parameters and rearranged to yield the band gap ratios in the wall that correspond to a specific eigenvector:

$$\frac{\Delta_{10}}{M_0} = -\frac{Q_{10}}{Q_{20}}\left\{\frac{c}{a}\right\}\frac{2i\frac{Q_{20}k_x}{M_0} - \left(1 - \frac{E}{M_0}\right)\left\{\frac{b}{d}\right\}}{1 - i\frac{A_0}{Q_{20}}\left\{\frac{c}{d}\right\}} + \frac{E}{M_0}, \tag{A2a}$$

$$\frac{\Delta_{20}}{M_0} = i\left\{\frac{b}{d}\right\}\frac{\frac{2Q_{20}k_x}{M_0} + i\left(1 - \frac{E}{M_0}\right)\left\{\frac{b}{d}\right\}}{1 - i\frac{A_0}{Q_{20}}\left\{\frac{c}{d}\right\}} - \frac{E}{M_0}. \tag{A2b}$$

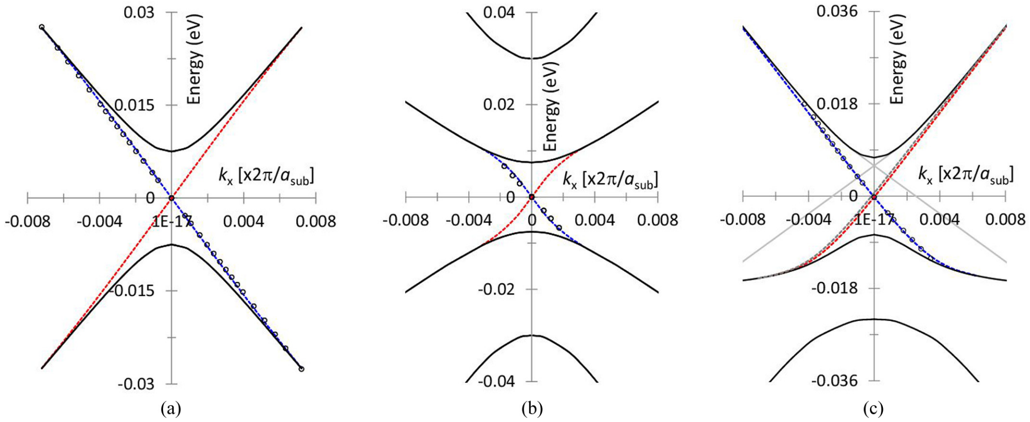
It is clear that these ratios become independent of the wall band gap $2M_0$ when $M_0 \gg |M|$, since E has the same order of magnitude as the semiconductor band gap parameter M so $|\Delta_{10}|, |\Delta_{20}| \to \infty$ as $M_0 \to \infty$. For given values of E and k_x, the semiconductor eigenvector components are calculated according to the relations in Table IV and inserted into Eq. (A2). This yields the wall band gaps for a given set of wall hybridization parameters, and M_0. These band gaps should be substituted back into the relations in Table IV and the process repeated at different energies until the wall eigenvector comes out to be the same as the semiconductor eigenvector, giving a self-consistent solution at the chosen wave vector. The values of the decay parameters in the semiconductor and wall can also be found for each point on the dispersion curve, using the formula in the left hand column of Table IV.

As discussed in Sec. II, it appears to be a typical feature of SBC edge states that an exponential solution only exists when specific wall parameters have a certain ratio with the wall band gap. The relevant BHZ parameters are D_0 and B_0, and a useful exponential solution can be found in the infinite wall limit when $D_0, B_0 \to 0$. For other ratios of the wall parameters, a numerical approach must be used. The LKMB model exhibits similar features, where an exponential solution exists for the wall band gaps given by Eq. (A2), giving a useful exponential solution in the limit: $|\Delta_{10}|, |\Delta_{20}| \to \infty$ as $M_0 \to \infty$. Practically, this "hard wall" limit corresponds to $M_0 \gg |M|$, which is already obeyed very well for $M_0 > 1\,\text{eV}$ in the examples discussed below.

The blue dashed curves in Fig. 7 show three examples of the spin up edge state dispersions calculated for symmetric or asymmetric semiconductor band structures. In each case, the wall parameters are $M_0 = 20$ eV, $A_0 = 11$ eV Å, and $Q_{10} = Q_{20} = 2$ eV Å, where M_0 and A_0 have the same order of magnitude as the values estimated for a passivation material in Sec. V. In all three cases the dispersions are insensitive to a variation in the value of the wall band gap by a factor of greater than 0.05 (corresponding to $M_0 > 1$ eV). In Fig. 7(a), the semiconductor parameters are the same as in Table II for a symmetric TI band structure. The open circles depict a variation of the form: $E = -Ak_x$, and there is virtually no difference between this variation and the blue dashed curve, showing that the dispersion calculated from the LKMB model has the same linear variation as for the BHZ case. The red dashed curve is for the spin down edge state and is obtained from the blue curve by time reversal.

In Fig. 7(b), the outer band gap in the semiconductor is reduced to $\Delta = \Delta_1 = \Delta_2 = 0.03\,\text{eV}$. Based on Eq. (4), this is equivalent to $B = 650\,\text{eV}\,\text{Å}^2$ for the BHZ case. All other parameters in the wall and semiconductor are the same as for Fig. 7(a), and except near the band edges, the LKMB result is still very close to the circles, which depict the BHZ variation, $E = -Ak_x$. Figure 7(c) shows an example with asymmetric semiconductor band parameters, as listed in the left hand column of Table III. As already discussed for this case in Sec. IV, the Dirac point is very close to midgap, with an energy of $E^{\mathrm{D}}_{4\times4\uparrow} = -0.00020\,\text{eV}$. Using this value as an estimate for the BHZ Dirac point energy, the open circles in Fig. 7(c) depict the dispersion based on the BHZ spin up result, $E = -Ak_x + Dk_x^2 + E^{\mathrm{D}}_{2\times2\uparrow}$, where the square root in Eq. (10) is very close to unity. The BHZ dispersion again agrees quite well with the calculated LKMB result.

The OBC dispersion, given by Eq. (8), may be compared with the SBC results in Fig. 7. It is the same as the open circles for the symmetric cases in Figs. 7(a) and 7(b) but very different for the asymmetric case in Fig. 7(c), where it is shown as gray lines. For the asymmetric case, the Dirac point

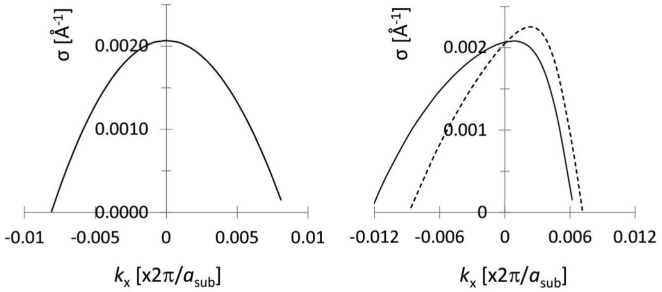


(a) (b) (c)

FIG. 7. Edge-state dispersions for the LKMB spin up (down) Hamiltonian are shown as blue (red) dashed curves superimposed on the nearby bulk bands (black curves). In each case, the wall has parameters $M_0 = 20\,\text{eV}$, $A_0 = 11\,\text{eV}\,\text{Å}$, and $Q_{10} = Q_{20} = 2\,\text{eV}\,\text{Å}$. The semiconductor parameters in (a) are the same as in Table II. In (b), the outer bands of the semiconductor are closer to the inner bands than in (a), with $\Delta_1 = \Delta_2 = 0.03\,\text{eV}$. In (c) the asymmetric semiconductor band parameters are those listed in the left hand column of Table III. Open circles depict the edge dispersions of the spin up BHZ Hamiltonian [Eq. (10)] with a Dirac point energy of zero in (a) and (b) and $-0.00020\,\text{eV}$ in (c). For comparison, the OBC dispersion is also shown in (c) as gray lines. The gray dashed line in (c) is the spin down edge dispersion after the wall band gap is upshifted by 2 eV, when the Dirac point shifts by 0.00079 eV.

is strongly shifted toward the conduction band and the edge state dispersion is linear with a much reduced phase velocity.

The wave vector dependence of the spin up decay parameters in the LKMB treatment is shown in Fig. 8, where it is compared for the symmetric and asymmetric cases shown in Figs. 7(a) and 7(c), respectively. The wave vectors of the merging points can clearly be identified where $\sigma \to 0$. In both cases, the decay parameter at the Dirac point agrees with Eq. (11), which has a value close to $\sigma_{\mathrm{D}} \simeq M/A = 0.0020\,\text{Å}^{-1}$ when $A^2 \gg 4MB$ as in these examples. For the symmetric case, the BHZ model corresponds fairly well with the LKMB model over the whole wave vector range, with merging points in the BHZ model at $\pm 0.083 \times \frac{2\pi}{a_{\mathrm{sub}}}$. For the asymmetric case, however, the BHZ model gives merging points at $\pm 0.072 \times \frac{2\pi}{a_{\mathrm{sub}}}$, while the magnitudes of the merging wave vectors in Fig. 8(b) are similar to these values but unequal. This can be attributed to stronger band nonparabolicities in the asymmetric LKMB model. The gray-dashed curve in Fig. 7(c) and the dashed curve in Fig. 8(b) show the effect of a misalignment between the wall and TI band gaps, which is equivalent to $M_1 = 2\,\text{eV}$ in Eq. (16). It can be seen that the edge dispersion is very insensitive to the misalignment, although the decay parameter is more strongly affected.



(a) (b)

FIG. 8. Decay parameters for the spin up edge states shown by the blue dashed curves in Figs. 7(a) and 7(c), respectively. The dashed line in Fig. 8(b) is when the wall band gap is upshifted by 2 eV.

All of the edge dispersions in Figs. 7 and 8 are based on middle wave vector solutions in the wall and semiconductor. As mentioned in Sec. III, when the wing solutions are used in the asymmetric BHZ model, they only yield an edge state when $D < B$, which then has an anomalous dispersion, as shown for example with $A_0 = A$ in Fig. 2(a) of Ref. [10]. This is also true in the LKMB model. Using parameters corresponding to Fig. 7(c) but with $A_0 = A$, the edge dispersion fails to cross or merge with the bulk bands, and no exponential edge state can be found for $A_0 = 11\,\text{eV}\text{Å}$. Therefore, this edge state can be considered unphysical even though, in principle, the LKMB Hamiltonian gives physical wing solutions, as shown, for example, in Fig. 2(b). This behavior is related to the large magnitude of the imaginary wing wave vector in the wall Hamiltonian, which tends to infinity as $\Delta_{10}, \Delta_{20} \to -\infty$. Even for large but finite values of these parameters, its magnitude is well beyond the boundary of the Brillouin zone.

Finally, it should be noted that the band gap ratios in Eq. (A2) that correspond to exponential edge state solutions are k dependent, so the outer bands have a significant dispersion even when the inner bands do not (i.e., constant M_0). Since all bands in the wall are effectively at infinity, and the results in Fig. 7 are totally insensitive to the positions of the inner bands for $M_0 > 1\,\text{eV}$, it is anticipated that the effect of the outer band positions on the energy dispersions should not be too significant. Nevertheless, in order to test this assumption a full numerical treatment is needed for constant values of all the wall band gaps, when the wave function will generally exhibit a nonexponential decay.

APPENDIX B: TOPOLOGICALLY TRIVIAL BHZ EDGE SOLUTIONS

It has recently been pointed out that edge states can be found for the BHZ Hamiltonian in the topologically trivial

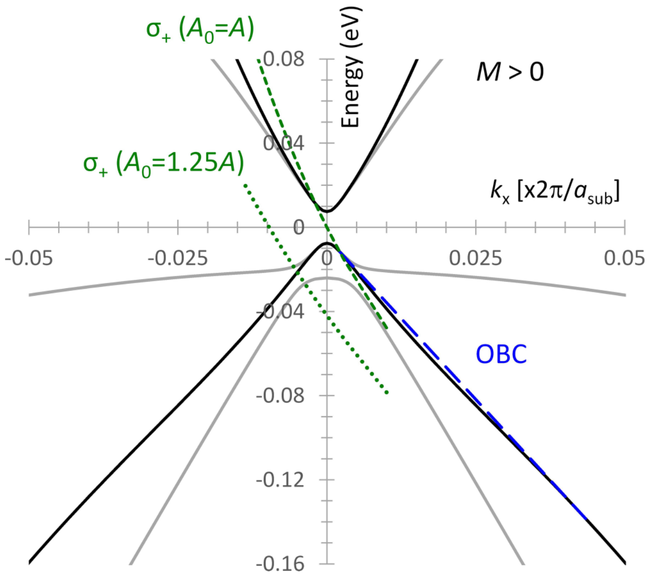


FIG. 9. BHZ band dispersions for a topologically trivial insulator with $M = 0.0075\,\text{eV}$, $A = 4\,\text{eV}\,\text{Å}$, $B = 76.2\,\text{eV}\,\text{Å}^2$, and $D = 50.8\,\text{eV}\,\text{Å}^2$(black), superimposed on the corresponding LKMB dispersions calculated from Eq. (3), with $\Delta_1 = 0.15\,\text{eV}$, $\Delta_2 = 0.024\,\text{eV}$, $Q_1 = Q_2 = 2\,\text{eV}\,\text{Å}$, and $Q_3 = 0$ (gray). There is good correspondence for $|k_x| < k_{\text{max}}$ (defined in Sec. I). A spin up OBC edge solution extends to wave vectors well beyond k_{max} (blue, long dash). Spin up SBC edge states, based on the spurious wing solution σ_+ are also plotted near the zone center for wall hybridization parameters of 4 and $5\,\text{eV}\,\text{Å}$ (green short-dash and dot, respectively). There is no SBC edge state with physical decay parameter σ_-.

phase, with $M > 0$ [41]. In this Appendix, such behavior is confirmed for both OBC and SBC boundary conditions. However, all such states include the wing solution, and are therefore unlikely to exist in any real physical system. No edge state can be found using SBCs based only on the physical middle solution.

The band structure near the bulk band gap of a topologically trivial insulator is shown in gray in Fig. 9, for an LKMB Hamiltonian with $M = 0.0075\,\text{eV}$, $A = 2Q_1 = 2Q_2 = 4\,\text{eV}\,\text{Å}$, $Q_3 = 0$, $\Delta_1 = 0.15\,\text{eV}$, and $\Delta_2 = 0.024\,\text{eV}$. The conduction and valence bands of the BHZ Hamiltonian are superimposed in black, with quadratic coefficients, $B = 76.2\,\text{eV}\,\text{Å}^2$ and $D = 50.8\,\text{eV}\,\text{Å}^2$, calculated from the LKMB parameters using Eq. (4). It can be seen that beyond the range $k_{\text{max}} \simeq 0.0075 \times \frac{2\pi}{a_{\text{sub}}}$ allowed by perturbation theory, the conduction and valence bands of the BHZ Hamiltonian deviate strongly from those of the parent LKMB Hamiltonian (see Sec. I), highlighting the fact that the results of the BHZ Hamiltonian are only reliable in this small wave vector range.

Exponential BHZ edge states can be determined using the OBC and SBC characteristic equations (7) and (9), respectively. For OBCs, a topologically trivial edge solution exists with real decay parameters when $0.0021 \times \frac{2\pi}{a_{\text{sub}}} < k_x < 0.043 \times \frac{2\pi}{a_{\text{sub}}}$. The dispersion, shown as a blue dashed line in Fig. 9, reproduces very well the form of the dispersion shown in Fig. 1(a) of Ref. [41] (see [83]). A topologically trivial SBC edge solution can also be found, for the wing decay parameter, σ_+, in Eq. (9). Its dispersion is plotted near the zone center, as green dashed and dotted lines, respectively, for wall hybridization parameters of 4 and $5\,\text{eV}\,\text{Å}$. The solution for $A_0 = A$ is in fact very similar to the spurious σ_+ solution shown for the TI phase in Fig. 2(a) of Ref. [10]. This is because $B_-\sigma_+^2 \gg |M|$ in Eq. (9), so the sign of M becomes unimportant. Noting that Eq. (9) corresponds to a wall with negative $B_0 \to 0$, a finite wall with positive B_0 can be intercalated next to the TI, so that the edge state localizes on the interface between wall materials for which $\Delta N_C = 1$, separating from the sample edge for which $\Delta N_C = 0$ (analogous to Fig. 4 and Ref. [55]). This confirms that the σ_+ edge state is not a physical solution. In addition to a wave function that includes the unphysical wing solution, both these OBC and SBC edge states extend to wave vectors far beyond the k_{max}—limit. In contrast, when SBCs are used in Eq. (9) with the physical, σ_- decay parameter, there is no topologically trivial solution. This is consistent with $\Delta N_C = 1$, allowing only a single edge state, which is the unphysical, σ_+ solution [84].

In Ref. [41], edge states appear in the topologically trivial phase for any value, apart from $\frac{\pi}{2}$, of a k_x-independent, phenomenological boundary condition parameter θ which is conserved across the phase boundary, and whose value depends on the boundary conditions used for the envelope function. For OBC edge states as in Fig. 9, θ is a function of $\frac{D}{B}$ and is indeed independent of wave vector and the sign of the band-gap parameter, M. For SBCs in the TI phase, with $A_0 = A$, and $|M|$, B, and D as in Fig. 9, the physical σ_- edge state behaves as $\theta = \chi\frac{\pi}{2}$, with $\chi = 1.018$ at $k_x = 0$, and $\chi \to 1$ at the merging points. The value of χ is quite insensitive to an increase in the wall hybridization parameter and always tends to one at the merging points. When there is charge conjugation symmetry with $D = 0$ and a linear edge dispersion, $\chi = 1$ for all k_x. Thus, even in the presence of strong band asymmetry, the behavior of the physical SBC solution is very close to the special case of $\theta = \frac{\pi}{2}$, where no edge state exists in the topologically trivial phase.

The θ parameter is directly related to the edge-state eigenvector, and its simple relationship with the edge-state dispersion is based on a linear approximation of the BHZ Hamiltonian near the phase transition, in which quadratic terms smaller than the band gap are ignored. Even for $D = 0$ in a strongly hybridized material, this does not really hold for the wing solution with decay parameter, $\sigma_w \simeq \frac{A}{B}$, since $B\sigma_w^2 > 4M$ (see Sec. III). Since the linear approximation only works for the physical σ_- solution, the paradoxical appearance of edge states in the topologically trivial phase may simply be a mathematical artifact associated with the anomalous eigenvector of the wing solution.

---

[1] P. Yu and M. Cardona, *Fundamentals of Semiconductors, Physics and Material Properties* (Springer Verlag, Berlin, 1996).
[2] R. H. Ritchie, Phys. Rev. **106**, 874 (1957).
[3] J. M. Pitarke, V. M. Silkin, E. V. Chulkov, and P. M. Echenique, Rep. Prog. Phys. **70**, 1 (2007).

[4] R. Fuchs and K. L. Kliewer, Phys. Rev. **140**, A2076 (1965).

[5] A. K. Sood, J. Menendez, M. Cardona, and K. Ploog, Phys. Rev. Lett. **54**, 2115 (1985).

[6] X.-L. Qi and S.-C. Zhang, Phys. Today **63**(1), 33 (2010).

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

[8] M. König, H. Buhmann, L. W. Molenkamp, T. L. Hughes, C.-X. Liu, X.-L. Qi, and S.-C. Zhang, J. Phys. Soc. Jpn. **77**, 031007 (2008).

[9] F. Nichele, H. J. Suominen, M. Kjaergaard, C. M. Marcus, E. Sajadi, J. A. Folk, F. Qu, A. J. A. Beukman, F. K. de Vries, J. van Veen *et al.*, New J. Phys. **18**, 083005 (2016).

[10] P. C. Klipstein, Phys. Rev. B **91**, 035310 (2015); **93**, 199905(E) (2016).

[11] G. Tkachov and E. M. Hankiewicz, Phs. Rev. Lett. **104**, 166803 (2010).

[12] G. Tkachov and E. M. Hankiewicz, Phys. Staus. Solidi B **250**, 215 (2013).

[13] A. Medhi and V. B. Shenoy, J. Phys.: Condens. Matter **24**, 355001 (2012).

[14] M. M. Asmar, D. E. Sheehy, and I. Vekhter, Phys. Rev. B **95**, 241115(R) (2017).

[15] B. Zhou, H.-Z. Lu, R.-L. Chu, S.-Q. Shen, and Q. Niu, Phys. Rev. Lett. **101**, 246807 (2008).

[16] S. Essert, Ph.D. thesis, Univ. of Regensberg, 2015.

[17] X.-L. Qi, Y.-S. Wu, and S.-C. Zhang, Phys. Rev. B **74**, 085308 (2006).

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

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

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

[21] J. Linder, T. Yokoyama, and A. Sudbø, Phys. Rev. B **80**, 205401 (2009).

[22] C.-X. Liu, X.-L. Qi, H.-J. Zhang, X. Dai, Z. Fang, and S.-C. Zhang, Phys. Rev. B **82**, 045122 (2010).

[23] H. Z. Lu, W. Y. Shan, W. Yao, Q. Niu, and S. Q. Shen, Phys. Rev. B **81**, 115407 (2010).

[24] S. Mao, Y. Kuramoto, K.-I. Imura, and A. Yamakage, J. Phys. Soc. Jpn. **79**, 124709 (2010).

[25] W. Y. Shan, H. Z. Lu, and S. Q. Shen, New J. Phys. **12**, 043048 (2010).

[26] E. B. Sonin, Phys. Rev. B **82**, 113307 (2010).

[27] S. Murakami, J. Phys.: Conf. Ser. **302**, 012019 (2011).

[28] S. Q. Shen, W. Y. Shan, and H. Lu, Spin **1**, 33 (2011).

[29] M. Wada, S. Murakami, F. Freimuth, and G. Bihlmayer, Phys. Rev. B **83**, 121310(R) (2011).

[30] P. Michetti, J. C. Budich, E. G. Novik, and P. Recher, Phys. Rev. B **85**, 125309 (2012).

[31] P. Michetti, P. H. Penteado, J. C. Egues, and P. Recher, Semicond. Sci. Technol. **27**, 124007 (2012).

[32] Y. Takagaki, J. Phys.: Condens. Matter **24**, 435301 (2012).

[33] L. Cano-Corte’s, C. Ortix, and J. van den Brink, Phys. Rev. Lett. **111**, 146801 (2013).

[34] M. Hohenadler and F. F. Assaad, J. Phys.: Condens. Matter **25**, 143201 (2013).

[35] P. Sengupta, K. Tillmann, T. Yaohua, M. Povolotskyi, and G. Klimeck, J. Appl. Phys. **114**, 043702 (2013).

[36] J. Wang, Y. Xu, and S.-C. Zhang, Phys. Rev. B **90**, 054503 (2014).

[37] D.-H. Xu, J.-H. Gao, C.-X. Liu, J.-H. Sun, F.-C. Zhang, and Y. Zhou, Phys. Rev. B **89**, 195104 (2014).

[38] V. V. Enaldiev, I. V. Zagorodnev, and V. A. Volkov, JETP Lett. **101**, 89 (2015).

[39] M. V. Durnev and S. A. Tarasenko, Phys. Rev. B **93**, 075434 (2016).

[40] M. V. Entin, L. I. Magarill, and M. Mahmoodian, Europhys. Lett. **118**, 57002 (2017).

[41] D. R. Candido, M. Kharitonov, J. C. Egues, and E. M. Hankiewicz, Phys. Rev. B. **98**, 161111(R) (2018).

[42] L. Gioia, U. Zülicke, M. Governale, and R. Winkler, Phys. Rev. B **97**, 205421 (2018).

[43] S. S. Krishtopenko and F. Teppe, Phys. Rev. B. **97**, 165408 (2018).

[44] C.-A. Li, S.-B. Zhang, and S.-Q. Shen, Phys. Rev. B **97**, 045420 (2018).

[45] R. Skolasinski, D. I. Pikulin, J. Alicea, and M. Wimmer, Phys. Rev. B **98**, 201404(R) (2018).

[46] M. V. Durnev and S. A. Tarasenko, Ann. Phys. (NY) **531**, 1800418 (2019).

[47] L. Gioia, M. G. Christie, U. Zülicke, M. Governale, and A. J. Sneyd, Phys. Rev. B **100**, 205417 (2019).

[48] J. Böttcher, C. Tutschku, and E. M. Hankiewicz, Phys. Rev. B **101**, 195433 (2020).

[49] M. V. Durnev, Phys. Solid State **62**, 504 (2020).

[50] C. H. Kane, in *Contemporary Concepts of Condensed Matter Science*, Vol. 6, edited by M. Franz and L. Molenkamp (Elsevier, Amsterdam, 2013), Chap. 1.

[51] P. C. Klipstein, Phys. Rev. B **81**, 235314 (2010).

[52] S. R. White and L. J. Sham, Phys. Rev. Lett. **47**, 879 (1981).

[53] M. F. H. Schuurmans and G. W. t’Hooft, Phys. Rev. B **31**, 8041 (1985).

[54] A physical dispersion must be evanescent in the band gap, starting and finishing at a band edge.

[55] P. C. Klipstein, J. Phys.: Condens. Matter **30**, 275302 (2018).

[56] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).

[57] V. A. Volkov and É. E. Takhtamirov, Phys. Usp. **40**, 1071 (1997).

[58] É. E. Takhtamirov and V. A. Volkov, Semicond. Sci. and Technol. **12**, 77 (1997).

[59] É. E. Takhtamirov and V. A. Volkov, JETP **89**, 1000 (1999).

[60] M. Cardona and F. H. Pollak, Phys. Rev. **142**, 530 (1966).

[61] G. L. Bir and G. E. Pikus, *Symmetry and Strain Induced Effects in Semiconductors* (Wiley, New York, 1974).

[62] P. Lawaetz, Phys. Rev. B. **4**, 3460 (1971).

[63] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.104.195407 for a derivation of the 12 band linear-*k* Hamiltonian (extended LKMB model). The KT Hamiltonian (extended BHZ model) is also presented, and the parameter values used in both models are listed. The dispersion over the range covered by the extended models is compared with an 8-band Kane model plotted for two sets of Kane/Luttinger parameters. The Supplemental Material includes Refs. [7,10,43,62,69].

[64] D. G. Rothe, R. W. Reinthaler, C.-X. Liu, L. W. Molenkamp, S.-C. Zhang, and E. M. Hankiewicz, New J. Phys. **12**, 065012 (2010).

[65] Even though the wing solution of the LKMB Hamiltonian is not necessarily spurious, the reason it gives an unphysical edge state is discussed at the end of Appendix A.

[66] O. E. Raichev, Phys. Rev. B. **85**, 045310 (2012).

[67] P. C. Klipstein, J. Phys.: Condens. Matter **28**, 375801 (2016).

[68] R. Winkler, *Spin-Orbit Coupling Effects in Two-Dimensional Electron and Hole Systems* (Springer-Verlag, Berlin, 2003).

[69] S. S. Krishtopenko, W. Knap, and F. Teppe, Nat. Sci. Rep. **6**, 30755 (2016).

[70] C. Pidgeon and R. Brown, Phys. Rev. **146**, 575 (1966).

[71] P. Klipstein, Y. Livneh, A. Glozman, S. Grossman, O. Klin, N. Snapi, and E. Weiss, J. Electron. Materials **43**, 2984 (2014).

[72] I. Knez, C. I. Rettner, S.-H. Yang, S. S. P. Parkin, L. Du, R.-R. Du, and G. Sullivan, Phys. Rev. Lett. **112**, 026602 (2014).

[73] W. A. Harrison, *Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond* (W. H. Freeman and Co., San Francisco, 1980), Chap. 11-D, pp. 267–270.

[74] S. Coh and D. Vanderbilt, Phys. Rev. B **78**, 054117 (2008).

[75] D. A. Tashmukhamedova and M. B. Yusupjanova, J. Surf. Investig. X-ray Synchrotron and Neutron Tech. **10**, 1273 (2016).

[76] A. V. Voitsekhovskii, D. I. Gorn, S. N. Nesmelov, and A. P. Kokhanenko, Opto-electron. Rev. **18**, 241 (2010).

[77] K. H. Hellwege and O. Madelung (eds.), *Physics of Group IV elements and III-V Compounds*, Landolt-Börnstein New Series, Group III Vol. 17a (Springer Verlag, Berlin, 1982).

[78] I. Knez, R.-R. Du, and G. Sullivan, Phys. Rev. B **81**, 201301(R) (2010).

[79] B. A. Foreman, Phys. Rev. Lett. **81**, 425 (1998).

[80] Y. Livneh, P. C. Klipstein, O. Klin, N. Snapi, S. Grossman, A. Glozman, and E. Weiss, Phys. Rev. B **86**, 235311 (2012); **90**, 039903(E) (2014).

[81] T. Fukui, Phys. Rev. Research **2**, 043136 (2020).

[82] A. Pezo, B. Focassio, G. R. Schleder, M. Costa, C. Lewenkopf, and A. Fazzio, Phys. Rev. Materials **5**, 014204 (2021).

[83] Note that the BHZ Hamiltonian used in Ref. [41] corresponds to spin down in this paper.

[84] This is also consistent with the LKMB model. For positive M_0, $\frac{\Delta_{i0}}{M_0} = \gamma \frac{\Delta_i}{M}$ with $\gamma = \pm 1$, semiconductor parameters as in Fig. 9, and $A_0 = A$, the only eigenvectors at zero energy and wave vector that are nearly similar on each side of the boundary are those for the wing solution, and only for $\gamma = -1$, when the outer band-gap parameters change sign (consistent with negative B_0 and $\Delta N_C = 1$ in the BHZ model). Since γ is negative there is no exponential solution (see Sec. II D). However, given that the eigenvectors are very close, namely $[i0.321, -0.174, 0.421, i0.830]$ in the wall and $[i0.343, -0.159, 0.410, i0.830]$ in the semiconductor, this suggests that a nearby nonexponential edge state exists in the LKMB model, corresponding to the wing solution of the topologically trivial phase.