PHYSICAL REVIEW B **93**, 205431 (2016)

Generic helical edge states due to Rashba spin-orbit coupling in a topological insulator

Laura Ortiz
Departamento de Física Teórica I, Universidad Complutense de Madrid, Spain

Rafael A. Molina
Instituto de Estructura de la Materia, IEM-CSIC, Serrano 123, Madrid 28006, Spain

Gloria Platero
Instituto de Ciencia de Materiales de Madrid, ICMM-CSIC, Madrid 28049, Spain

Anders Mathias Lunde
Center for Quantum Devices, Niels Bohr Institute, University of Copenhagen, Denmark

(Received 28 September 2015; revised manuscript received 15 March 2016; published 20 May 2016)

We study the helical edge states of a two-dimensional topological insulator without axial spin symmetry due to the Rashba spin-orbit interaction. Lack of axial spin symmetry can lead to so-called *generic* helical edge states, which have energy-dependent spin orientation. This opens the possibility of inelastic backscattering and thereby nonquantized transport. Here we find analytically the new dispersion relations and the energy dependent spin orientation of the generic helical edge states in the presence of Rashba spin-orbit coupling within the Bernevig-Hughes-Zhang model, for both a single isolated edge and for a finite width ribbon. In the single-edge case, we analytically quantify the energy dependence of the spin orientation, which turns out to be weak for a realistic HgTe quantum well. Nevertheless, finite size effects combined with Rashba spin-orbit coupling result in two avoided crossings in the energy dispersions, where the spin orientation variation of the edge states is very significantly increased for realistic parameters. Finally, our analytical results are found to compare well to a numerical tight-binding regularization of the model.

DOI: 10.1103/PhysRevB.93.205431

I. INTRODUCTION

An insulating bulk energy gap along with gapless edge states is a hallmark of a two-dimensional (2D) topological insulator (TI) [1–4]. At each boundary, two counterpropagating edge states with opposite spin polarization and wave numbers form Kramers pairs, i.e., two distinct degenerate states connected by time-reversal symmetry. These states are denoted helical edge states due to their connection between spin and propagation direction. Due to time-reversal symmetry, elastic backscattering of a single electron from a helical edge state (HES) to its Kramers partner is not possible by any time-reversal invariant potential, e.g., disorder [5]. Thereby an important mechanism to hinder ballistic transport is absent and quantized conductance of $2e^2/h$ per pair of HESs is within reach. The first experimentally realized 2D TI in a HgTe quantum well (QW) indeed found quantized conductance in micrometer-sized samples [6–9], along with evidence of edge-state transport in both two-terminal [6] and multiterminal [7] configurations. Prior to the experiments, HgTe QWs were in fact predicted to be 2D TIs beyond a certain QW thickness [10]. These efforts also resulted in a rather generic Dirac-like model describing the essential physics of some 2D TIs, which is now known as the Bernevig-Hughes-Zhang (BHZ) model. Recently, also InAs/GaSb double QWs were suggested theoretically to be 2D TIs described using the BHZ model [11], which afterwards have been tested experimentally [12–16].

Deviations from the quantized conductance have also been found experimentally for longer edges in both HgTe [6,7,9,17–19] and InAs/GaSb [20] QW TIs. Conduction reduction due to *inelastic* backscattering has been studied theoretically [21–28], since it is not *a priori* ruled out by time-reversal invariance. Most studies of inelastic backscattering combine some energy-exchange mechanism (e.g., phonons [24] or electron-electron interactions [21,22,24,27]) with a way to manipulate the spin (often some form of spin-orbit coupling [24,26,27]). Scattering of localized spins [29–33] such as magnetic impurities or nuclear spins [34] has also been analyzed.

In a particularly interesting proposal for inelastic backscattering, Schmidt *et al.* [21] considered HESs without axial spin symmetry. The Rashba spin-orbit coupling (RSOC) [35–37] and bulk inversion asymmetry (BIA) [11,38,39] can break the axial spin symmetry of the HESs. In this case, a pair of HESs acquire a more generic and intriguing spin structure than merely having opposite and constant spin orientations independently of energy. Time-reversal symmetry still dictates that the two counterpropagating Kramers partners have orthogonal spinors, but it does not require equal spinors at different energies as illustrated in Fig. 1. These states were

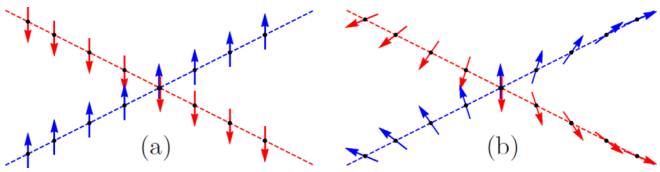

FIG. 1. Illustration of the dispersion relation and spin orientation for (a) helical edge states with constant spin orientation and (b) generic helical edge states with energy-dependent spin orientation. In this paper, we analyze the generic helical edge states and their spin orientation variation due to the Rashba spin-orbit coupling within the BHZ model.

named *generic* helical edge states (GHESs) [21]. Recently, Kainaris *et al.* [27] extended the original work [21] on transport in short GHESs with electronic interaction and disorder to longer ones. Furthermore, the spin structure of the GHESs was also shown to change the noninteracting transport properties of a point contact and of disordered 2D TI strips [40]. Moreover, the spin structure of GHESs plays a role in the umklapp-scattering-induced energy gap suggested to host parafermions, a generalization of Majorana fermions [41].

These studies show that it is worthwhile to analyze the GHESs and their microscopic origin further, which is the purpose of this paper. Very recently, Rod *et al.* [42] studied the spin texture of GHESs due to BIA within the BHZ model and also numerically for the Kane-Mele model [1,2]. In contrast, we consider how RSOC [35] can produce GHESs within the BHZ model. We develop analytical models for the GHESs appearing at an isolated boundary and in the case of a finite width ribbon, where the overlap of the edge states on different boundaries plays an important role. For an isolated edge, we are able to give an analytical formula for the so-called spin-structure parameter, which describes how much the spin orientation of the GHESs changes. This parameter was originally introduced phenomenologically [21]. Using realistic numbers for a HgTe QW, we find that an isolated edge is in fact rather robust against spin rotation produced by the RSOC. In contrast, we discover that the combination of RSOC and the finite width enhanced significantly the spin rotation versus energy of the GHESs. Throughout the paper, spin rotation refers to the spin orientation variations of the GHES. Furthermore, we show that our analytical models compare well to full numerical tight-binding calculations.

We organize the paper as follows. First, we outline the phenomenology of the GHESs (Sec. II) and the BHZ model including the RSOC that breaks the axial spin symmetry (Sec. III). Then, we consider an isolated pair of GHESs at a single boundary in Sec. IV A and finally analyze the case of a finite width ribbon both analytically and numerically (Sec. IV B). Section V summarizes the paper and the appendices give various technical details.

II. PHENOMENOLOGY OF THE GENERIC HELICAL EDGE STATES

In this section, we discuss the GHESs phenomenologically. GHESs can be modelled as two counterpropagating one-dimensional (1D) states with linear dispersion relations $\varepsilon_{k,\pm} = \pm\hbar v k$, i.e.,

$$H_0 = \sum_{k,\eta=\pm1} \eta\hbar v k c^\dagger_{k\eta} c_{k\eta}, \tag{1}$$

as in Refs. [21,27,40,41]. Here, $c^\dagger_{k\eta}$ ($c_{k\eta}$) creates (annihilates) a state $|k,\eta\rangle$ with momentum k and propagating direction η. The states of opposite k and η are Kramers partners such that elastic scattering due to, e.g., impurities is still absent. The spin s_j ($j = x,y,z$) expectation values of the Kramers partners are also opposite, i.e.,

$$\langle k,+|s_j|k,+\rangle = -\langle -k,-|s_j|-k,-\rangle, \quad j = x,y,z. \tag{2}$$

The counterpropagating states at each k can be related to the spin states $\sigma = \uparrow,\downarrow$ along a definite direction by a momentum dependent (and thereby energy dependent) SU(2) matrix B_k as [21]

$$\begin{pmatrix} c_{k\uparrow} \\ c_{k\downarrow} \end{pmatrix} = B_k \begin{pmatrix} c_{k+} \\ c_{k-} \end{pmatrix}. \tag{3}$$

Time-reversal symmetry and $B_k \in$ SU(2) lead to $B_k = B_{-k}$. Consistent with these facts, Schmidt *et al.* [21] introduced the following expansion for small $|k| \ll k_0$:

$$B_k = \begin{pmatrix} 1 - k^4/(2k_0^4) & -k^2/k_0^2 \\ k^2/k_0^2 & 1 - k^4/(2k_0^4) \end{pmatrix}, \tag{4}$$

where the spin-quantization axis is chosen such that at the band crossing point $k = 0$, we have $c_{k=0,+} = c_{k=0\uparrow}$ and $c_{k=0,-} = c_{k=0\downarrow}$ as in Fig. 1(b). In other words, a constant rotation of all the spins regardlessly of k has been removed from B_k in Eq. (4) following Ref. [21]. Such a k-independent rotation corresponds to a constant rotation matrix and can be removed by choosing a rotated basis for the spin. Importantly, a phenomenological spin-structure parameter k_0 has been introduced in the expansion (4), which measures the velocity of spin rotation in momentum space. Schmidt *et al.* [21] showed using perturbation theory that the correction to the quantized conductance due to backscattering processes possible within a pair of GHESs scales as temperature to the fourth power with a prefactor depending on k_0. In this paper, we find k_0 analytically within the BHZ model including the RSOC for an isolated edge.

To gain more insights into the spin structure of the GHESs, we also evaluate the total spin rotation of the edge states, which we define as

$$T_s = \int dk(|\langle k_1,\uparrow|k_1,+\rangle|^2 - |\langle k,\uparrow|k,+\rangle|^2). \tag{5}$$

Here, k_1 is a fixed reference momentum and the integration is over the range of k space where the edge states exist. The idea behind T_s is to quantify the total variation of the spin orientation of the edge state $|k,+\rangle$ over all relevant k. We have constructed T_s such that if $|k,+\rangle$ is a HES (i.e., $|k,+\rangle = |k,\uparrow\rangle$), then $T_s = 0$. Likewise, if the spin of $|k,+\rangle$ is rotated by the same amount for all k, then we still get $T_s = 0$. This is due to the reference term $|\langle k_1,\uparrow|k_1,+\rangle|^2$ with an arbitrary, but fixed, momentum k_1. In our calculations, we choose k_1 to be the momentum where the edge-state dispersion crosses the upper bulk band gap edge. In the cases we have analyzed, the reference term $|\langle k_1,\uparrow|k_1,+\rangle|^2$ is very close to one and quite unaffected by small changes in k_1. However, generally, the choice of k_1 does affect the numerical value of T_s, but it does not change the variations of T_s versus some physical parameter. The behavior of the spin rotation is more complex for a ribbon than for a single edge, especially for narrower ribbons, as the edge-state wave function can have components on both edges. The quantity T_s is useful in that case as this kind of behavior is difficult to capture with the parameter k_0, which quantifies the rotation close to $k = 0$. The unit of both T_s and k_0 is inverse length.

Before proceeding, we consider a simple 1D model Hamiltonian for a pair of HESs with a generic linear spin-orbit coupling, i.e.,

$$H = \hbar v k\sigma_z + (a_x\sigma_x + a_y\sigma_y)k. \tag{6}$$

Here, $\sigma_i, i = x,y,z$, are the Pauli matrices and a_x, a_y are the spin-orbit coupling strengths. By diagonalization, we see that this often used [31,32] simple model does not introduce a k-dependent B_k, since all matrix elements are linear in momentum k. Thereby, it does not give rise to energy-dependent spin orientation and to GHESs, i.e., $T_s = 0$. This is consistent with the lack of the lowest-order inelastic backscattering due to a linear spin-orbit coupling combined with a phonon exchange [23]. In order to get nontrivial GHESs, we resort to calculations for the realistic BHZ model with RSOC.

III. THE BHZ MODEL WITH RASHBA SPIN-ORBIT COUPLING

The BHZ model is an effective four-band model describing the basic physics of a 2D TI [10]. It was derived using $\mathbf{k}\cdot\mathbf{p}$ theory for the band structure of a HgTe QW and therefore valid for small wave vectors $\mathbf{k} = (k_x, k_y)$, i.e., close to the Γ point. It accounts correctly for the physics of HgTe QWs close to the critical well thickness, which marks the transition between a normal semiconductor band structure and an inverted band structure with topologically protected edge states [10]. The BHZ Hamiltonian consists of two disconnected blocks connected by time reversal symmetry. Each block has the form of a massive Dirac model in 2D in addition to quadratic terms crucial for the band inversion and thereby the topological properties of the material. In fact, the Dirac-like nature makes the BHZ model rather generic for 2D TIs—even though it grew out of a specific material choice. The BHZ basis states consist of two Kramer pairs of electronlike, $|E\pm\rangle$, and hole-like, $|H\pm\rangle$, states, respectively. The states labeled with $+$ $(-)$ are often referred to as the spin up (spin down), since they have positive (negative) total angular momentum projection [34]. In this sense, the time-reversed blocks of the BHZ model have opposite spin. In the basis $\{|E+\rangle, |H+\rangle, |E-\rangle, |H-\rangle\}$, the BHZ Hamiltonian is

$$H_0 = \begin{pmatrix} \varepsilon_k + M_k & Ak_+ & 0 & 0 \\ Ak_- & \varepsilon_k - M_k & 0 & 0 \\ 0 & 0 & \varepsilon_k + M_k & -Ak_- \\ 0 & 0 & -Ak_+ & \varepsilon_k - M_k \end{pmatrix}, \quad (7)$$

where $k_\pm = k_x \pm i k_y$, $\varepsilon_k = -Dk^2$, $M_k = M_0 - Bk^2$ and $k^2 = k_x^2 + k_y^2$. The sign of M_0/B determines the existence of the HESs [43] and $D \neq 0$ induces particle-hole asymmetry in H_0. Table I gives the parameters for two different systems modelled by the BHZ model, namely HgTe QWs [10] and InAs/GaSb double QWs [11].

In this paper, we utilize an extension of the BHZ model derived in Ref. [35] for the inclusion of structural inversion asymmetry (SIA) terms including the RSOC. Importantly, the RSOC couples the two blocks of H_0 such that the axial spin symmetry is broken. Here, we include only the most important RSOC linear in momentum, i.e.,

$$H_R = \begin{pmatrix} 0 & 0 & -iR_0k_- & 0 \\ 0 & 0 & 0 & 0 \\ iR_0k_+ & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (8)$$

TABLE I. The BHZ model parameters for two QW systems in the topological regime [44]. The parameters for the HgTe QW correspond to a well width of 7 nm, while the values for the InAs/GaSb double QWs are for equal widths of 10 nm for both wells. The RSOC constant R_0 includes a value for the external electric field ϵ_z (in meV) in the case of HgTe.

Material	HgTe	InAs/GaSb
A (meV nm)	365.0	37.0
B (meV nm^2)	-686.0	-66.0
D (meV nm^2)	-512.0	-58.0
M_0 (meV)	-10.0	-7.8
R_0 (meV nm)	$15.6\,\epsilon_z$	-7.0

and therefore our full Hamiltonian is $H = H_0 + H_R$. Interestingly, the Rashba term in H_R only couples the electronlike bands, which makes our model more complex than the simple 2×2 model Hamiltonian in Eq. (6). Moreover, GHES are now possible as we shall see below. The strength of the RSOC, R_0, depends of the amount of SIA, which is often related to an internal or external electric field. For a HgTe QW one can control the RSOC with an external field [35], whereas it is an internal field for InAs/GaSb double QWs [11]. Rothe *et al.* [35] also derives higher-order RSOC terms in momentum as we briefly discuss in Appendix E.

The HgTe has a zinc-blende crystal structure such that inversion symmetry of the crystal is lacking. Therefore bulk-inversion-asymmetric terms can, in principle, be included [38] but are often disregarded due to their small size [3,38]. However, in InAs/GaSb double QWs BIA terms are in fact significant [11], so our calculations for this system without BIA terms are not an attempt to model the real system in detail.

Before we proceed to the GHESs, we briefly comment on the bulk bands including the RSOC. By diagonalizing $H = H_0 + H_R$, we find

$$E^n_{1,2} = -Dk^2 \pm \frac{R_0 k}{2} - \frac{\sqrt{J_k + K_k^\pm}}{2}, \quad (9a)$$

$$E^p_{3,4} = -Dk^2 \pm \frac{R_0 k}{2} + \frac{\sqrt{J_k + K_k^\pm}}{2}, \quad (9b)$$

where p (n) is the bulk band with positive (negative) energy and we define $J_k = 4A^2k^2 + 4B^2k^4 + R_0^2k^2 + 4M_0^2$ and $K_k^\pm = -4Bk^2(2M_0 \pm R_0 k) \pm 4M_0R_0k$ and $k = \sqrt{k_x^2 + k_y^2} > 0$. Figure 2 shows the bulk energy bands with RSOC (together with the edge-state dispersions that we consider below). The bulk bands shift due to the RSOC such that the band gap for HgTe becomes indirect. Moreover, the size of the bulk band gap is changed slightly, but not enough to change the topology of the system.

IV. GENERIC HELICAL EDGE STATES

The RSOC breaks the spin degeneracy of the BHZ model in such a way that GHESs with energy-dependent spin orientation now become feasible. We treat the GHESs below in two cases:

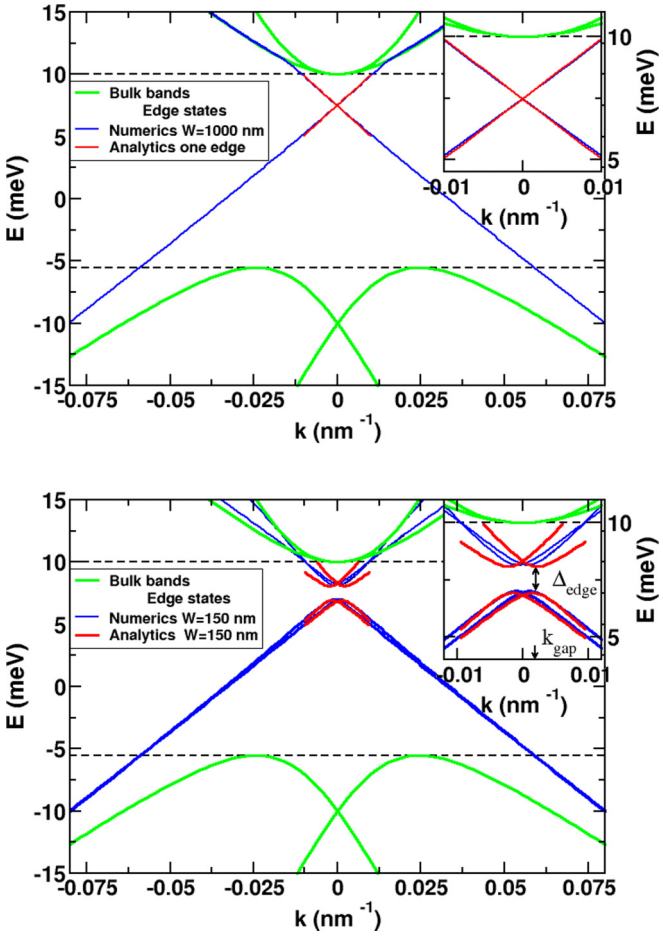

FIG. 2. Bulk and edge-state dispersion with the RSOC for a single edge (top) and a ribbon (bottom) in a HgTe QW. The parameters are given in Table I and the electric field ϵ_z is such that $R_0 = A$. The bulk gap is marked by dashed horizontal lines and the bulk bands, Eq. (9), are for $H_0 + H_R$ with periodic boundary conditions in both directions. Analytical results are not shown in the entire bulk energy gap, because our method requires the existence of both edge states at the same time. The insets show the edge-state dispersions close to $k = 0$.

(i) a single isolated edge and (ii) a finite width ribbon with two parallel edges. The isolated edge case offers more analytical insights and we are able to extract the spin-structure parameter k_0 defined in Eq. (4).

Without the RSOC, it is possible to obtain the HESs analytically for both an isolated edge and a ribbon with two edges [43]. In Appendix B, we give the details of the analytical wave functions and dispersion relations of the HESs in both cases. Including the RSOC, it becomes much more challenging to obtain exact analytical forms by the same method, since it is now a problem with four coupled differential equations, see Appendix B 1. Nevertheless, we are able to obtain analytical results by assuming that the GHESs with RSOC are combinations of the HESs without RSOC, neglecting the possible contribution of the bulk bands. This is a good approximation, since the edge states naturally have a small spatial overlap with the bulk states as long as the edge states are well-localized at the boundary. This is well satisfied especially for momenta close to zero and well into the bulk gap. We find that the bulk gap is reduced as R_0 increases and so does the range of applicability of the analytical results. Moreover, we also compare our analytical results with the solution via exact diagonalization of a tight-binding regularization of the BHZ Hamiltonian for a ribbon of width W with periodic boundary conditions in the x direction and edges at $y = -W/2$ and $y = W/2$. The details of the tight-binding formulation is discussed in Appendix D and follows Ref. [45]. This calculation allows us to unambiguously check the validity of our analytical model.

In the next subsection, we find the GHESs in the presence of RSOC for an isolated edge. We obtain explicit expressions for the spin orientation versus energy and find good agreement with the large-width limit of the numerics. We show that for a single edge, the spin orientation is only weakly dependent on energy for a real HgTe sample, i.e., the spin orientation is actually quite robust against RSOC. The following subsection is devoted to a ribbon. Now the expressions become much more complicated but the results as a function of the width of the sample show more interesting patterns, where spin rotation versus energy cannot be neglected.

A. The case of a single isolated edge

Now we find the pair of GHESs appearing at an isolated boundary of a 2D TI described by the BHZ model including the RSOC. As mentioned above, the starting point is the exact HESs without the RSOC. The HES dispersions are linear [46], i.e., $E_{\sigma k_y} = E_0 + \mathfrak{s}\hbar v k_y$, where $\mathfrak{s} = +(-)$ for $\sigma = \uparrow(\downarrow)$, v is the constant velocity and E_0 an energy shift. The HESs located at the boundary of the half-plane $x > 0$ are given by

$$\psi_{k_y\sigma}(x,y) = \frac{1}{\sqrt{L}} e^{ik_y y} g_{\mathfrak{s}k_y}(x)\hat{\phi}_\sigma, \tag{10}$$

i.e., a plane-wave running along the y axis combined with a transverse wave function $g_{\mathfrak{s}k_y}(x)$ determining the width of the HES and a k_y-independent four-component spinor $\hat{\phi}_\sigma$. There is one spinor from each time-reversed block of the BHZ model, i.e., $\hat{\phi}_\uparrow$ ($\hat{\phi}_\downarrow$) only has nonzero components on the two first (last) entries with positive (negative) total angular momentum projection. Periodic boundary conditions are used along the edge of length L. The HES wave functions and dispersions are given explicitly using the BHZ parameters in Appendix B 2.

To include the RSOC analytically, we write the full Hamiltonian $H = H_0 + H_R$ in a basis of the HESs for $R_0 = 0$ given in Eq. (10), i.e.,

$$\mathcal{H}_0 + \mathcal{H}_R = \sum_{k_y,\sigma\in\{\uparrow,\downarrow\}} E_{\sigma k_y} c^\dagger_{\sigma k_y} c_{\sigma k_y} + \sum_{k_y,k'_y}\sum_{\sigma\sigma'} \langle \psi_{k_y\sigma} | H_R | \psi_{k'_y\sigma'}\rangle c^\dagger_{\sigma k_y} c_{\sigma' k'_y}, \tag{11}$$

where $c^\dagger_{\sigma k_y}$ ($c_{\sigma k_y}$) creates (annihilates) a particle in the HES $\psi_{k_y\sigma}$. In this approach, we neglect the matrix elements between the edge and bulk states. These are presumably very small, since bulk and edge states to a very large extend are spatially

separated. This is an excellent assumption well within the bulk gap close to the Γ point, whereas the bulk states begin to play a role close to the bulk band gap edge as our numerics show. The full Hamiltonian (11) simplifies by noting that the matrix elements of H_R are diagonal in k_y due to the plane-wave part of the HESs (10). Moreover, H_R only couples opposite spins, so we find

$$\mathcal{H} = \sum_{k_y} (c^\dagger_{\uparrow k_y}, c^\dagger_{\downarrow k_y}) \begin{pmatrix} E_0 + \hbar v k_y & k_y \alpha_{k_y} \\ k_y \alpha_{k_y} & E_0 - \hbar v k_y \end{pmatrix} \begin{pmatrix} c_{\uparrow k_y} \\ c_{\downarrow k_y} \end{pmatrix}, \quad (12)$$

where an effective RSOC $\alpha_{k_y} \equiv \langle \psi_{k_y\uparrow} | H_R | \psi_{k_y\downarrow} \rangle / k_y$ is introduced. In terms of the BHZ parameters, we find

$$\alpha_{k_y} = R_0 \frac{B-D}{2Bk_y} \int_0^\infty dx\, g_{k_y}(x)[\partial_x + k_y] g_{-k_y}(x)$$
$$= R_0 \frac{(B-D)^2}{2B^2} \left(1 - a k_y^2\right) + \mathcal{O}\left[k_y^4\right]. \quad (13)$$

where $a = \frac{D^2[A^2B + 2(B^2-D^2)M_0]}{2B(B^2-D^2)M_0^2}$ and we expanded in k_y in the last step. The exact result for α_{k_y} and details of the calculation are found in Appendix C 1.

The effective RSOC (13) only includes the first-order RSOC in the BHZ basis given in Eq. (8). In Appendix E, we discuss the effects of higher-order RSOC terms. We show that the second-order term does not contribute to α_{k_y}, while the third-order term, in principle, could contribute even though we face technical difficulties in this case due to the hard wall boundary condition used to find the HESs analytically. However, the third-order RSOC term cannot introduce terms of a different order in k_y in α_{k_y} than the ones found here. Therefore it cannot change the physics of the GHESs discussed below. Moreover, the magnitude of the effects of the third-order term can partly be incorporated into the prefactor R_0.

The form of $\mathcal{H}$ in Eq. (12) is clearly very similar to the simple 1D Hamiltonian for a pair of HESs with a generic spin-orbit coupling (6), since the effective spin-orbit term $\alpha_{k_y}\sigma_x k_y$ resembles $a_x \sigma_x k$. The important difference is that our effective RSOC α_{k_y} depends on k_y and therefore gives rise to GHESs with *k_y-dependent* (or equivalently energy-dependent) spin orientation as we shall see shortly. In contrast, the spin-orbit coupling in Eq. (6) only leads to a constant wave-vector-independent spin rotation. In other words, the effective spin-orbit term $\alpha_{k_y}\sigma_x k_y$ has to be nonlinear in k_y for GHESs to arise.

By diagonalizing $\mathcal{H}$ in Eq. (12), we get the dispersion relations including the RSOC,

$$E^{\mathrm{RSOC}}_{k_y,\pm} = E_0 \pm \hbar v_{\alpha_{k_y}} k_y, \quad (14a)$$

and the eigenstates in k_y space,

$$\Psi_{k_y,\pm} = \frac{1}{\sqrt{2}} \begin{pmatrix} \pm\sqrt{1 \pm \frac{v}{v_{\alpha_{k_y}}}} \\ \sqrt{1 \mp \frac{v}{v_{\alpha_{k_y}}}} \end{pmatrix}, \quad (14b)$$

where $\pm$ corresponds to two different edge states with the renormalized velocity $v_{\alpha_{k_y}} = \sqrt{v^2 + (\alpha_{k_y}/\hbar)^2}$. For $R_0 = 0$, the states are simply $\psi_{k_y\uparrow}$ and $\psi_{k_y\downarrow}$, whereas for $R_0 \neq 0$, they become a superposition of both spins. Moreover, they are GHESs due to their k_y-dependent spin orientation. The case described by the model Hamiltonian in Eq. (6) is included here: if α_{k_y} is *in*dependent of k_y, then so are $\Psi_{k_y,\pm}$ and no GHESs appear. Due to time-reversal symmetry, the eigenstates constitute a Kramers pair with opposite spin orientations (i.e., orthogonal spinors). This is seen by applying the time-reversal operator Θ to $\Psi_{k_y,\pm}$ and using $\alpha_{-k_y}=\alpha_{k_y}$, i.e., $\Theta\Psi_{k_y,\pm} = \mp\Psi_{-k_y,\mp}$ (see Appendix A). Finally, we observe that the RSOC does not open a gap in the spectrum in accordance with time-reversal symmetry, but merely renormalizes the velocity close to $k_y = 0$ and creates a slight nonlinearity for larger k_y.

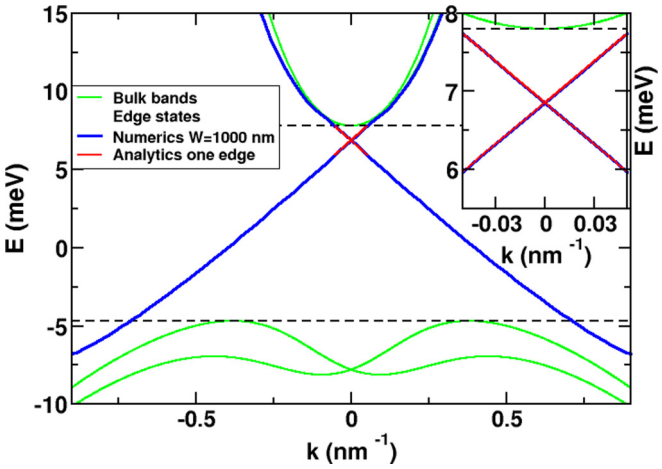

FIG. 3. Dispersion relation for InAs/GaSb QWs using the BHZ Hamiltonian with the parameters given by Table I without taking the BIA terms into account. Analytical results for an isolated edge and numerical results for $W = 1000$ nm for the edge states almost coincide.

The GHES dispersions (14a) for a HgTe QW with $R_0 = A$ are shown in the top panel of Fig. 2 along with a comparison to our numerical results using the tight-binding regularization for $W = 1000$ nm. We find that the effect of the RSOC on the dispersions is rather weak for a HgTe sample. We also present similar calculations for an InAs/GaSb double QW in Fig. 3. Our analytical method only works if both HESs without RSOC exist simultaneously, hence the dispersions do not cover the entire bulk gap as seen in Figs. 2 and 3. Although the bulk bands are quite different for the InAs/GaSb and HgTe QWs, we find that the behavior of the GHESs is very similar for similar values of R_0—both in the single-edge case and for the ribbon discussed in the next section. Therefore we do not show more figures for InAS/GaSb with the understanding that the results for the latter are similar to our results for HgTe in the presence of an electric field such that $R_0 \approx 0.2A$.

Next, we consider the k_y dependence of the spin orientation of the GHESs in the case of an HgTe TI. In Fig. 4, we show the amount of spin $\downarrow$ in the state $\Psi_{k_y,+}$, which is a spin $\uparrow$ state for $R_0 = 0$, i.e., the projection $P = |\langle k_y,\downarrow | k_y,+\rangle|^2 = |\langle \psi_{k_y\downarrow} | \Psi_{k_y,+}\rangle|^2$. We find a reasonably good comparison between the analytical results for the isolated edge and the numerical results for a large width of $W = 1000$ nm. The small discrepancy between the analytical and numerical projections could be due to the truncation of the Hilbert space in the

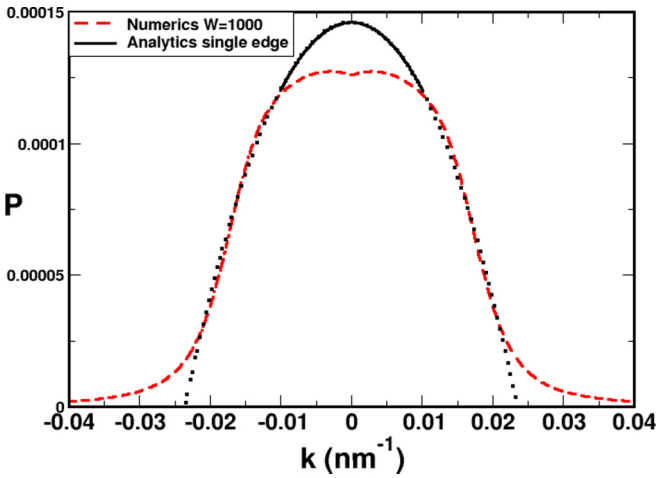

FIG. 4. The projection $P = |\langle\psi_{k_y\downarrow}|\Psi_{k_y,+}\rangle|^2$ of the GHES $\Psi_{k_y,+}$ with $R_0 = 0.5A$ into the $R_0 = 0$ spin-$\downarrow$ state as a function of k_y using the parameters of a HgTe QW in Table I. The figure shows a comparison of the analytical results with the numerics with $W =$ 1000 nm. The analytical projection is seen in both the bulk energy band gap (full black curve) and in the regime of coexistence of bulk and edge states (dotted black curve). Moreover, we have manually removed the numerical results close $k = 0$ where a very narrow peak appears due to finite size effect, see Fig. 6 and the discussion in Sec. IV B.

analytical calculation. As seen in the figure, the spin rotation is rather small in a realistic HgTe QW even for relatively large values of R_0, i.e., the spin orientation of the edge states is rather robust against large external electric field. The analytical projection is found from the GHESs in Eq. (14b) using the exact RSOC α_{k_y} in Eq. (C1) in Appendix C 1. The analytical theory requires simultaneous existence of both HESs without RSOC. The analytical projection in Fig. 4 is shown in both the bulk band gap region (full curve) and in the region of coexistence between edge and bulk states (dotted curve). In the coexistence regime, the HESs gradually widen and finally the penetration length divergences well within the bulk states as seen in Fig. 10 in Appendix B 2. By using the projection, we obtain the total spin rotation T_s, Eq. (5). From the numerical results for the entire k space, we find that T_s is proportional to R_0^2 to a good approximation.

Now, we find the analytical form of the spin structure parameter [21] k_0 in Eq. (4) for the BHZ model including the RSOC. We do this by introducing two unitary transformations, which together diagonalize $\mathcal{H}$ in Eq. (12). The first transformation is k_y-independent and rotate the spin basis such that it removes the k_y-independent part of α_{k_y}. This part does not lead to GHESs as discussed above. This rotation is convenient such that we use the same choice of spin-quantization axis as in Ref. [21], i.e., the spins point along the new rotated spin-quantization axis at $k = 0$. The second unitary transformation is k_y-dependent and transforms between the eigenstates and the new rotated spin basis. In other words, it is the matrix B_k in Eq. (3). Now we perform the steps explicitly. First, we define $\delta\alpha_{k_y} \equiv \alpha_{k_y} - \alpha_0$, where $\alpha_0 = \alpha_{k_y=0}$ is k_y-independent. Thereby, we can diagonalize the α_0 part of $\mathcal{H}$, i.e.,

$$\begin{aligned}\mathcal{H} &= \sum_{k_y} C_{k_y}^\dagger \begin{pmatrix} E_0 + \hbar v k_y & k_y(\alpha_0 + \delta\alpha_{k_y}) \\ k_y(\alpha_0 + \delta\alpha_{k_y}) & E_0 - \hbar v k_y \end{pmatrix} C_{k_y} \\ &= \sum_{k_y} C_{k_y}^\dagger U \begin{pmatrix} E_0 + \hbar v_{\alpha_0} k_y & 0 \\ 0 & E_0 - \hbar v_{\alpha_0} k_y \end{pmatrix} U^\dagger C_{k_y} \\ &\quad + \sum_{k_y} C_{k_y}^\dagger U k_y \delta\alpha_{k_y} \begin{pmatrix} \sin(\theta) & \cos(\theta) \\ \cos(\theta) & -\sin(\theta) \end{pmatrix} U^\dagger C_{k_y}. \end{aligned} \quad (15)$$

Here, $C_{k_y}^\dagger = (c_{\uparrow k_y}^\dagger, c_{\downarrow k_y}^\dagger)$, $\hbar v_{\alpha_0} = \sqrt{(\hbar v)^2 + \alpha_0^2}$ and the first k_y-independent unitary transformation U is

$$U = \begin{pmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{pmatrix}, \quad (16)$$

where $\cos(\theta) \equiv v/v_{\alpha_0}$ and $\sin(\theta) \equiv \alpha_0/(\hbar v_{\alpha_0})$. This transformation is simply a k_y-independent rotation to a new spin basis,

$$\begin{pmatrix} c_{\uparrow' k_y} \\ c_{\downarrow' k_y} \end{pmatrix} = U^\dagger \begin{pmatrix} c_{\uparrow k_y} \\ c_{\downarrow k_y} \end{pmatrix}, \quad (17)$$

where $\uparrow'$ and $\downarrow'$ are the eigenstates of $\mathcal{H}$ at $k_y = 0$. Finally, we diagonalize the Hamiltonian completely by a second unitary transformation and obtain

$$\mathcal{H} = \sum_{k_y} C_{k_y}'^\dagger \mathcal{V}_{k_y} \begin{pmatrix} E_0 + k_y \hbar v_{\alpha_{k_y}} & 0 \\ 0 & E_0 - k_y \hbar v_{\alpha_{k_y}} \end{pmatrix} \mathcal{V}_{k_y}^\dagger C'_{k_y},$$

where $C_{k_y}'^\dagger = (c_{\uparrow' k_y}^\dagger, c_{\downarrow' k_y}^\dagger)$ and $\hbar v_{\alpha_{k_y}} = \sqrt{(\hbar v)^2 + \alpha_{k_y}^2}$. As expected, we find the same dispersions as in Eq. (14a). More importantly, we acquire an analytical form of the unitary transformation $\mathcal{V}_{k_y}$, which by construction is exactly B_{k_y} from Eq. (3), i.e.,

$$B_{k_y} = \mathcal{V}_{k_y} = \begin{pmatrix} \cos(\phi/2) & -\sin(\phi/2) \\ \sin(\phi/2) & \cos(\phi/2) \end{pmatrix}, \quad (18)$$

where $\cos(\phi) \equiv [(\hbar v)^2 + \alpha_{k_y}\alpha_0]/(\hbar^2 v_{\alpha_0} v_{\alpha_{k_y}})$ and $\sin(\phi) \equiv \delta\alpha_{k_y} v/(v_{\alpha_0}\hbar v_{\alpha_{k_y}})$. Therefore we have now found the k_y-dependent matrix B_{k_y} relating the GHESs to the HESs with a fixed spin axis for a specific model, namely the BHZ model including the RSOC. We remark that $\delta\alpha_{k_y} = 0$ at $k_y = 0$ by definition, so $\mathcal{V}_{k_y=0}$ is the unity matrix and therefore $\uparrow'$ and $\downarrow'$ become eigenstates at $k_y = 0$.

We can now find the spin structure parameter k_0 in Eq. (4) controlling the amount of spin rotation for small k_y. By expanding $B_{k_y} = \mathcal{V}_{k_y}$ around $k_y = 0$, we obtain

$$\frac{1}{k_0^2} = \frac{D^2 |R_0 A(B-D)||A^2 B + 2M_0(B^2 - D^2)|}{2\sqrt{B^2 - D^2} M_0^2 \left|4A^2B^2(B+D) + (B-D)^3 R_0^2\right|}. \quad (19)$$

Thereby we have an explicit expression for k_0—a parameter originally introduced based on symmetry arguments [21]. Such an expression in terms of the BHZ parameters is valuable beyond the case of HgTe QWs due to the generic Dirac-like nature of the BHZ model. Interestingly, we observe that the particle-hole asymmetry parameter D plays an essential role for k_0, i.e., for $D = 0$, no spin rotation appears and therefore no GHESs come out in the case studied here. This is valid beyond the expansion of B_{k_y} in k_y, since the

effective RSOC in Eq. (13) is k_y-independent to all orders, $\alpha_{k_y}^{(D=0)} = R_0/2$, for $D = 0$ such that B_{k_y} is the unity matrix [see Eq. (C5) in Appendix C 1]. Curiously, the parameter D is often removed in many theoretical discussions of topology [47] and thereby the interesting physics of GHESs might be missed. Furthermore, Eq. (19) also reveals that k_0 depends on the Dirac mass M_0 and the RSOC strength R_0 in rather nontrivial ways.

Before proceeding, we briefly discuss the effect of the lowest-order BIA terms given by [3,38]

$$H_{\mathrm{BIA}} = \begin{pmatrix} 0 & 0 & 0 & -\Delta \\ 0 & 0 & \Delta & 0 \\ 0 & \Delta & 0 & 0 \\ -\Delta & 0 & 0 & 0 \end{pmatrix}, \tag{20}$$

where Δ is a constant. Including the H_{BIA} in the basis of the HESs for an isolated edge, Eq. (10), as we did for H_R in Eq. (11), we find $\langle \psi_{k_y\sigma} | H_{\mathrm{BIA}} | \psi_{k'_y\sigma'} \rangle = 0$ for all $k_y\sigma$ and $k'_y\sigma'$. Hence within our analytic approach, the lowest-order BIA terms do not affect the HESs nor their spin orientation for an isolated edge. The second-order RSOC terms have the same structure in the antidiagonal as H_{BIA} and therefore also have zero matrix elements, see Appendix E. Including the small overlaps between the bulk and edge states, a modest effect on the energy dispersions is found due to H_{BIA} close to the bulk gap edge, where these overlaps matter the most [48]. For a ribbon, the H_{BIA} was found to couple opposite edges [49]. Very recently, Rod *et al.* [42] found GHESs for both ribbon and disk geometries due to H_{BIA} in the BHZ model. For both geometries, a finite k_0^{-2} was extracted numerically in the limit of a particle-hole symmetric BHZ model (i.e., $D = 0$), where both edge and bulk states were included in their calculations.

B. The case of a ribbon with two parallel edges

In this section, we consider the GHESs for a ribbon with two parallel edges using the BHZ model including the RSOC. Thereby, four edge states come into play, since a pair of GHESs exist on each edge for well-separated edges. We pay special attention to how the finite size effects can enhance spin orientation variation of the GHESs as the width of the ribbon W gets smaller and the edge states on opposite sides begin to overlap.

Before including the RSOC, we briefly summarize the HESs and their dispersions without RSOC for a ribbon [43]. We refer to Appendix B 3 for details. An important difference between the ribbon and the single-edge case is that we do not have the energy dispersions in closed analytical forms for the ribbon, but instead as the solutions to a cumbersome equation [see Eq. (B12)]. Nevertheless, the physical consequence of the finite width is clear: a gap opens at the crossing point of the dispersions found for the isolated edge, see Fig. 9 [43]. The dispersions for a ribbon have a limiting cusp form for a wide ribbon, i.e.,

$$E_{k_x}^{e=\pm} \to E_0 \pm \hbar v |k_x| \quad \text{for} \quad W \to \infty, \tag{21}$$

where $E_{k_x}^{+}$ ($E_{k_x}^{-}$) is the energy dispersion above (below) the gap for $W \leqslant \infty$. Therefore the label $e = \pm$ should not be confused with the single-edge case, where $\pm$ often refers to the sign of the velocity. The velocity v and energy shift E_0 are identical to the single-edge case. Noticeably, $E_{k_x}^{e=\pm}$ are independent of the spin σ, since equal spins travel in opposite directions on the two edges.

A ribbon with edges at $y = \pm W/2$ has four HESs without RSOC [43] $\psi_{k_x\sigma}^{e}(x,y)$, where $e = \pm$ labels the energy $E_{k_x}^{e}$ to which the state belongs. As for an isolated edge, the states have a plane-wave part running along the edges, i.e., $\psi_{k_x\sigma}^{e} \propto e^{ik_x x}$. Only the first (last) two components of the states $\psi_{k_x\uparrow}^{e}$ ($\psi_{k_x\downarrow}^{e}$) are nonzero, corresponding to the spin-up (spin-down) block of H_0. However, in contrast to the single-edge case, the spinors are not constant, but the relative weight of the two components vary with both k_x and y. A particular state $\psi_{k_x\sigma}^{e}$ is not always localized on the same edge. Instead, the localization changes gradually from one edge to the other when crossing $k_x = 0$. For $k_x > 0$, the states

$$\psi_{k_x\uparrow}^{+},\ \psi_{k_x\downarrow}^{-} \text{ are localized close to } y = W/2 \text{ and} \tag{22a}$$

$$\psi_{k_x\uparrow}^{-},\ \psi_{k_x\downarrow}^{+} \text{ are localized close to } y = -W/2, \tag{22b}$$

and vice versa for $k_x < 0$.

As for the isolated edge, we build an analytical model using only the HESs without RSOC. Since this approach leaves out the overlaps between bulk and edge states, it becomes less good for a narrow ribbon, where bulk and edge states become comparable in spatial extend. Therefore our analytical results are most reliable for small momenta well within the bulk gap as we shall see.

By including the RSOC in the subspace of the HESs without RSOC, $\psi_{k_x\sigma}^{e}$, the Hamiltonian becomes

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_R = \sum_{\sigma,k_x,e} E_{k_x}^{e} \left(c_{k_x\sigma}^{e}\right)^{\dagger} c_{k_x\sigma}^{e} + \sum_{k_x,k'_x} \sum_{\sigma,\sigma'} \sum_{e,e'} \left\langle \psi_{k_x\sigma}^{e} \middle| H_R \middle| \psi_{k'_x\sigma'}^{e'} \right\rangle \left(c_{k_x\sigma}^{e}\right)^{\dagger} c_{k'_x\sigma'}^{e'}, \tag{23}$$

where $(c_{k_x\sigma}^{e})^{\dagger}$ [$c_{k_x\sigma}^{e}$] creates [annihilates] a particle in the HES $\psi_{k_x\sigma}^{e}$ of energy $E_{k_x}^{e}$. The RSOC Eq. (8) only couples opposite spins and the Hamiltonian is diagonal in k_x, since $\langle \psi_{k_x\sigma}^{e} | H_R | \psi_{k'_x\sigma'}^{e'} \rangle \propto \delta_{k_x,k'_x}$. We order the basis as $\{|\psi_{k_x\uparrow}^{+}\rangle, |\psi_{k_x\downarrow}^{-}\rangle, |\psi_{k_x\uparrow}^{-}\rangle, |\psi_{k_x\downarrow}^{+}\rangle\}$ such that the first two entries are localized on the opposite edge of the last two as seen in Eq. (22), i.e.,

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_R = \sum_{k_x} \mathbf{C}_{k_x}^{\dagger} \begin{pmatrix} E_{k_x}^{+} & ib & 0 & id_{+} \\ -ib & E_{k_x}^{-} & id_{-} & 0 \\ 0 & -id_{-} & E_{k_x}^{-} & -ib \\ -id_{+} & 0 & ib & E_{k_x}^{+} \end{pmatrix} \mathbf{C}_{k_x}, \tag{24}$$

where $\mathbf{C}_{k_x}^{\dagger} = [(c_{k_x\uparrow}^{+})^{\dagger}, (c_{k_x\downarrow}^{-})^{\dagger}, (c_{k_x\uparrow}^{-})^{\dagger}, (c_{k_x\downarrow}^{+})^{\dagger}]$ and we introduced the interedge matrix elements

$$id_{+} = \left\langle \psi_{k_x\downarrow}^{+} \middle| H_R \middle| \psi_{k_x\uparrow}^{+} \right\rangle, \tag{25a}$$

$$id_{-} = \left\langle \psi_{k_x\uparrow}^{-} \middle| H_R \middle| \psi_{k_x\downarrow}^{-} \right\rangle, \tag{25b}$$

and the intraedge matrix element

$$ib = \left\langle \psi_{k_x\downarrow}^{-} \middle| H_R \middle| \psi_{k_x\uparrow}^{+} \right\rangle, \tag{26}$$

which all depend on k_x. In Eq. (24), we used that the intraedge matrix elements on opposite edges are related as $ib = \langle\psi^-_{k_x\downarrow}|H_R|\psi^+_{k_x\uparrow}\rangle = -\langle\psi^+_{k_x\downarrow}|H_R|\psi^-_{k_x\uparrow}\rangle$, as discussed in Appendix C 2. Thereby, we are left with three matrix elements only, which depend on the implicitly known dispersions relations $E^\pm_{k_x}$. The detailed formulas are given in Appendix C 2.

Due to the ordering of the basis, the Hamiltonian (24) has two 2×2 blocks in the diagonal, one for each edge. Each 2×2 block resembles the Hamiltonian (12) found for an isolated edge and an effective intraedge RSOC could be introduced as b/k_x as in Sec. IV A. However, due to the limiting cusp form of the energy dispersions (21), one should instead define the effective intraedge RSOC as $\alpha^{\text{intra}}_{k_x} = -b/|k_x|$. With this definition, $\alpha^{\text{intra}}_{k_x}$ corresponds to the effective RSOC for the isolated edge in Eq. (13) in the wide ribbon limit. However, as the width gets smaller, we find an increased k_x dependence of $\alpha^{\text{intra}}_{k_x}$ for small k_x. This indicates an increased spin-orientation change for small k_x as W decreases, which we also find below by direct calculation.

The opposite edges of the ribbon are coupled by the interedge elements $d_\pm$ in the antidiagonal of $\mathcal{H}$, which vanish for $W\to\infty$. Finally, we mention that performing unitary transformations of $\mathcal{H}$ to find k_0 as in Sec. IV A is difficult, since we do not have closed formulas for $E^e_{k_x}$.

By diagonalization of the Hamiltonian (24), the dispersion relations including the RSOC become

$$E^{\text{RSOC}}_{k_x,s\tau} = \tau\tfrac{1}{2}\sqrt{\left[s(d_- - d_+) + E^-_{k_x} - E^+_{k_x}\right]^2 + 4b^2} + \tfrac{1}{2}\left[s(d_+ + d_-) + E^+_{k_x} + E^-_{k_x}\right], \tag{27}$$

where $s=\pm1$ and $\tau=\pm1$. In the wide-ribbon limit, where the interedge matrix elements $d_\pm$ are insignificant, these dispersions resemble the isolated-edge dispersions, Eq. (14a) (disregarding the cusp limit of $E^\pm_{k_x}$). For a finite width W, however, the interedge matrix elements $d_\pm$ come into play and create four different dispersions. As shown in the bottom panel of Fig. 2 and in Fig. 7, two gaps arise symmetrically with respect to $k_x=0$. Therefore we have found that the spin degeneration present for $R_0=0$ between $\psi^e_{k_x\uparrow}$ and $\psi^e_{k_x\downarrow}$ is broken by the interplay of RSOC and a finite width, where both ingredients are necessary. A similar effect of SIA combined with finite size have also been found in one dimension higher, namely for the 2D Dirac surface states on a 3D TI [50].

In Fig. 5, we show the position k_{gap} and value Δ_{edge} of the gaps as a function of W for different R_0. We compare numerical and analytical calculations on a logarithmic scale. As the width is increased, the position of the gap goes rapidly towards $k_x=0$ and the value of the gap goes to zero, such that we recover the result for an isolated edge. Interestingly enough, there are several values of the width, depending on the value of R_0, where the gap is particularly reduced. The reason is essentially that the actual transverse wave function including RSOC for an isolated edge has a form similar to an exponential times a sine. This means that the solution for an isolated edge state (including the gapless dispersions) also becomes the solution for a finite ribbon, when the zeros of the transverse wave function match the width. This destructive interference has been studied before both with [51] and without [52] RSOC. Similar physics has also been discussed

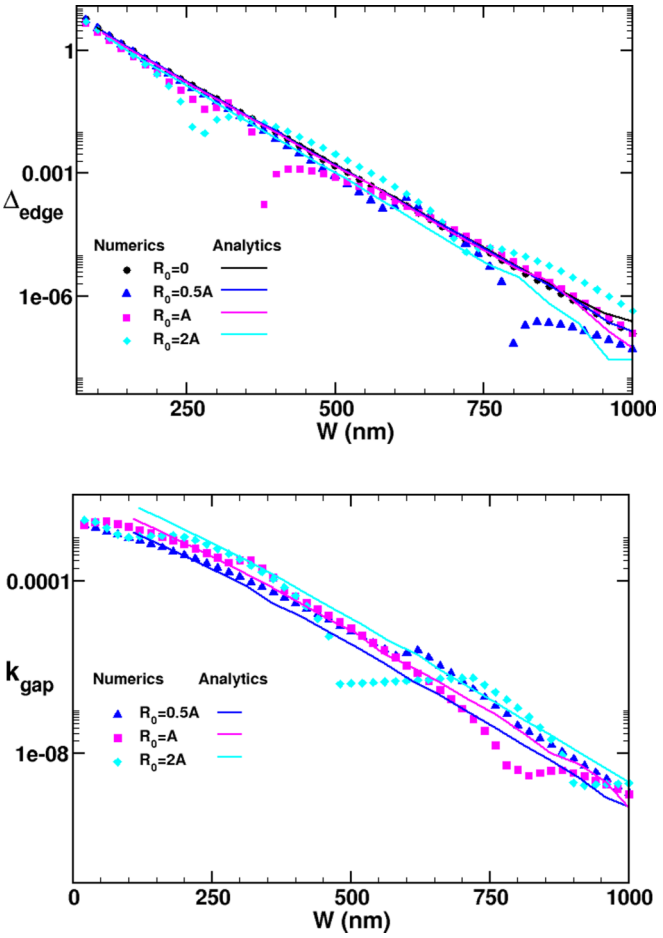

FIG. 5. Comparison of analytical and numerical results for the value of the gap Δ_{edge} (meV) vs width of the ribbon (top) and position of the gap in momentum k_{gap} (nm^{-1}) vs width of the ribbon (bottom). We use the HgTe parameters in Table I.

for thin films of 3D TIs [53–55]. In Ref. [51], Takagaki showed that the gap vanishes periodically with a period almost inversely proportional to the strength of the RSOC. In these particular values of the width, the coupling between the edges is canceled without reaching the large width limit. As shown in Fig. 5, this effect is not captured by the analytical theory, although it correctly gives the essential decaying trends of both the gap size Δ_{edge} and position k_{gap}.

The eigenvectors in k_x space in the basis presented above, i.e., $\{|\psi^+_{k_x\uparrow}\rangle, |\psi^-_{k_x\downarrow}\rangle, |\psi^-_{k_x\uparrow}\rangle, |\psi^+_{k_x\downarrow}\rangle\}$ are

$$\Psi_{k_x,s\tau} = \frac{1}{\sqrt{8b^2+2\zeta^2_{s\tau}}}\begin{pmatrix} is2b \\ -s\zeta_{s\tau} \\ i\zeta_{s\tau} \\ 2b\end{pmatrix}, \tag{28}$$

where $s=\pm$ and $\tau=\pm$ and we defined

$$\zeta_{s\tau} = s(d_+ - d_-) + E^+_{k_x} - E^-_{k_x} - \tau\sqrt{\left[s(d_- - d_+) + E^-_{k_x} - E^+_{k_x}\right]^2 + 4b^2}. \tag{29}$$

Here the two first and the two last components of $\Psi_{k_x,s\tau}$ represent spinors localized on opposite edges. The four states

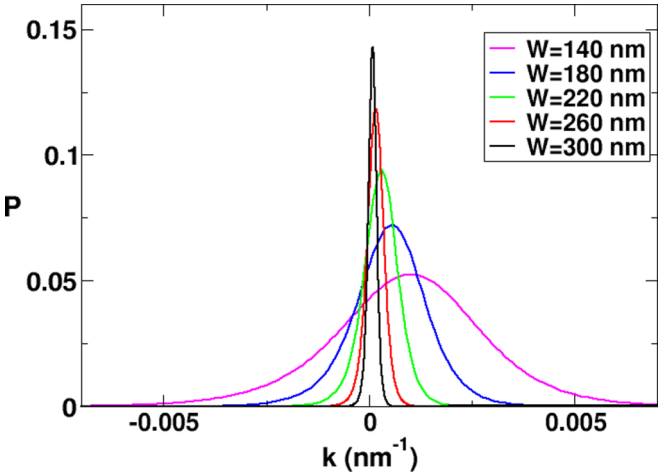

FIG. 6. The analytical projection of a ribbon edge state with primarily spin up character (on the lower edge $y = -W/2$ and $k_x > 0$) into the spin down subspace, i.e., $P = |\langle\psi^+_{k_x\downarrow}|\Psi_{k_x,+-}\rangle|^2$. In other words, the last component squared of $\Psi_{k_x,+-}$ in Eq. (28). The parameters for HgTe in Table I and $R_0 = A$ are used.

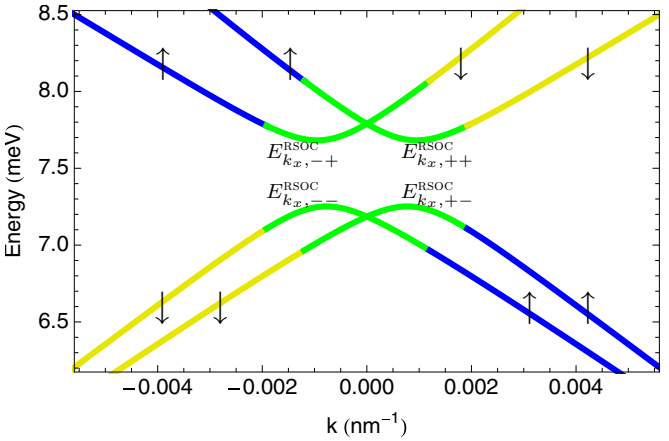

FIG. 7. The ribbon energy dispersions $E^{\rm RSOC}_{k_x,s\tau}$, Eq. (27), and the spin structure of the GHESs close to the lower edge at $y = -W/2$. The combination of finite width and RSOC produce splitting in k_x space and energy gaps at $k_{\rm gap} \neq 0$. The spin structure associated to the two avoided crossings is illustrated by the colors: the states are more than 99% pure spin ↑ (↓) on the lower edge in the blue (yellow) part of the dispersions, whereas the spin orientation rotates when all states come close together (green regions). The upper edge at $y = W/2$ has the opposite spin structure (see the main text). Therefore the states become true GHESs in the green regions, which coincide with the peaks in the projection seen in Fig. 6. The parameters for HgTe in Table I are used together with $R_0 = A$ and $W = 200$ nm.

$\Psi_{k_x,s\tau}$ are clearly present on both edges, but in a very particular way: the spinor localized on one edge, $\varphi_a \propto (is2b, -s\zeta_{s\tau})^T$, is always orthogonal to the spinor $\varphi_b \propto (i\zeta_{s\tau}, 2b)^T$ localized on the opposite edge, since $\varphi_a^\dagger\varphi_b = 0$. In other words, the squared projections on the basis states on opposite edges are pairwise identical, i.e., $|\langle\psi^+_{k_x\uparrow}|\Psi_{k_x,s\tau}\rangle|^2 = |\langle\psi^+_{k_x\downarrow}|\Psi_{k_x,s\tau}\rangle|^2$ and $|\langle\psi^-_{k_x\downarrow}|\Psi_{k_x,s\tau}\rangle|^2 = |\langle\psi^-_{k_x\uparrow}|\Psi_{k_x,s\tau}\rangle|^2$. Thus the states always have half of the weight on each edge, i.e., $|\langle\psi^+_{k_x\uparrow}|\Psi_{k_x,s\tau}\rangle|^2 + |\langle\psi^-_{k_x\downarrow}|\Psi_{k_x,s\tau}\rangle|^2 = 1/2$ independently of k_x. Moreover, the Kramers partner of $\Psi_{k_x,\pm\pm}$ is $\Psi_{-k_x,\mp\pm}$, which can be seen by using that $E^\pm_{k_x}$ and b are even in k_x and $d_\pm$ is odd such that $\zeta_{s\tau}(-k_x) = \zeta_{-s\tau}(k_x)$, see Appendix C 2. Furthermore, the dispersions $E^{\rm RSOC}_{k_x,s\tau}$ (27) and eigenstates (28) depend on both $E^+_{k_x}$ and $E^-_{k_x}$, and therefore only well-defined for momenta k_x, where both $E^-_{k_x}$ and $E^+_{k_x}$ are well-defined, see Figs. 2 and 9.

Now, we will argue that the eigenstates $\Psi_{k_x,s\tau}$ are GHESs, since their spin orientation on a single edge depends on k_x. Due to the structure of $\Psi_{k_x,s\tau}$ discussed above, we observe that the two edges of the ribbon suffer the same—but opposite—spin rotation. Due to the coupling between the two edges and the RSOC, the dispersions have two avoided crossings. These avoided crossings induce some particular characteristics of the spin rotation. Figure 6 shows the analytical results for the projection onto the spin-down subspace, $P = |\langle\psi^+_{k_x\downarrow}|\Psi_{k_x,s=1\tau=-1}\rangle|^2$, as a function of k_x for one of the edge states, which asymptotically is more than 99% spin up for different values of W on lower edge ($y = -W/2$) and $k_x > 0$. We can see that the projection reaches a relatively high value, higher for larger widths, but in a very narrow range of k_x, smaller for larger widths. The peaks of the projections are located close to the position of the gap $k_{\rm gap}$. For clarity, we only show a range of W from 140 to 300 nm, but the trend goes on indefinitely.

The avoided crossings of the ribbon dispersions and the associated spin structure of the GHESs are illustrated in Fig. 7. As discussed above, the GHESs $\Psi_{k_x,s\tau}$, Eq. (28), are always equally present on the lower ($y = -W/2$) and the upper ($y = W/2$) edge. Figure 7 only shows the spin structure of the lower edge. We illustrate how to understand this by using the state $\Psi_{k_x,+-}$ as an example. Away from the avoided crossing (the green region), we find

$$\Psi_{k_x,+-} = \frac{1}{\sqrt{2}}\left(-\psi^-_{k_x\downarrow} + i\psi^-_{k_x\uparrow}\right) \tag{30}$$

with more than 99% accuracy. For $k_x > 0$, $\psi^-_{k_x\uparrow}$ ($\psi^-_{k_x\downarrow}$) is localized near the lower (upper) edge and vice versa for $k_x < 0$, see Eq. (22). Thus, in this sense, $\Psi_{k_x,+-}$ is spin ↑ for $k_x > 0$ (blue region) and spin ↓ for $k_x < 0$ (yellow region) *on the lower edge*, while the upper edge has the opposite spin structure. In between these regions of almost pure spin ↑ or ↓, the states become genuine GHESs with sizable amounts of both spin ↑ and ↓ present on each edge (the green region). These regions are quantified by the peaks in the projections shown in Fig. 6. Noticeably, the weight of each spin component in the almost pure spin regions (blue/yellow) of the GHESs is only 1/2, see, e.g., Eq. (30). Thus the entire weight of one spin component on one edge is carried by two different dispersion curves. This is vastly different from the two simple linear dispersions found for an isolated edge. Therefore it is now clear that our states $\Psi_{k_x,s\tau}$ indeed are GHESs with k_x-dependent spin orientation. A remarkable difference to the isolated edge case is that the spin-orientation change is enhanced a great deal by the finite size.

In Fig. 8, we show that the total spin rotation T_s of $\Psi_{k_x,+-}$ scales with R_0^2 for not too large values of R_0. The total spin rotation is essentially the integral of the projections in Fig. 6 due to our choice of k_1 in Eq. (5). We only show the numerical

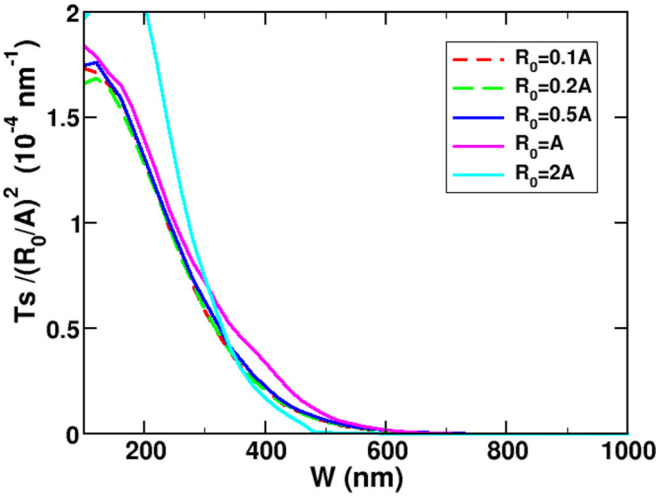

FIG. 8. Numerical value of the total spin rotation T_s of the GHES $\Psi_{k_x,+-}$ rescaled with R_0^2 as a function of the ribbon width for different values of R_0.

results as the analytical states (28) are only available in the small k_x range, where both $E^+_{k_x}$ and $E^-_{k_x}$ are well-defined. Nevertheless, using the analytical states to find T_s very similar results are obtained for analytically feasible values of R_0 and W. Although the maximum of the spin projection in Fig. 6 increases, the total value of the integral is reduced, when the ribbon widens to the single-edge limit. The scaling with R_0^2 works perfectly well for $R_0 \lesssim 0.5A$ except for very small values of the width. (Note that we only show $W > 100$ nm in Fig. 8.) For larger values of R_0, the scaled total spin rotation T_s/R_0^2 increases compared to the values of $R_0 \leqslant 0.5A$. However, for very large spin-orbit couplings (like $R_0 = 2A$ in Fig. 8) and large widths, we obtain smaller T_s/R_0^2 probably due to the reduced bulk gap of the system.

V. SUMMARY

We have analyzed the spin structure of the generic helical edge states appearing at the boundary of 2D TIs without axial spin symmetry. For the usual helical edge states in a 2D TI, the spin and propagation direction are locked in such a way that the spin orientation is energy independent. However, for the GHESs, the spin orientation varies with energy or equivalently momentum k. This is possible in systems without axial spin symmetry, broken for instance by spin-orbit coupling. Importantly, time reversal symmetry still ensures counterpropagating states to be Kramers partners with orthogonal spins, but the spin orientations of neighboring states with different energies are not identical. This opens the possibility of inelastic scattering and thereby deviations from quantized conductance [21].

Our study is focused on the GHESs produced by Rashba spin-orbit coupling within the BHZ model. We use HgTe QWs and InAs/GaSb double QWs as concrete examples. We analyze two situations: (i) a pair of GHESs at an isolated edge and (ii) the two pairs of GHESs in a ribbon with two parallel boundaries. In both cases, we employ an analytical approach, where the GHESs *with* RSOC are found within a reduced basis consisting of the HESs *without* RSOC. This is a good approximation, since the bulk and edge states are usually well separated spatially—especially for small k within the bulk energy gap. We also use a numerical tight-binding regularization of the BHZ model including RSOC to verify the analytical approach and, moreover, obtain independent valuable information.

For an isolated boundary, our analytical approach gives rise to a 2×2 Hamiltonian (12), which is formally equivalent to a simple 1D model of a pair of HESs with a phenomenological spin-orbit coupling. From this analogy, we discover that GHESs are produced, when the effective spin-orbit coupling term is nonlinear in the momentum. In contrast, no GHESs appear for a linear effective spin-orbit coupling term within our framework. Moreover, we find the effective RSOC α_{k_y} in terms of the BHZ parameters. We also obtain the pair of GHESs in Eq. (14b), where the velocity has been renormalized. Using our insights into linear versus nonlinear effective RSOC terms, we are able to provide an explicit expression for the so-called spin-structure parameter k_0, which measures the amount of spin-orientation variation for small k. The spin-structure parameter k_0 was originally deduced by symmetry arguments [21] and it is interesting to have an expression in a concrete case. For instance, it shows that k_0 depends on the RSOC strength R_0 and the Dirac mass M_0 in nontrivial ways. Moreover, $1/k_0^2$ vanishes when the particle-hole symmetry parameter D of the BHZ model is zero. This statement is in fact more general: the effective RSOC term becomes exactly linear for $D = 0$ such that only ordinary HESs appear in this case. For realistic HgTe and InAs/GaSb TIs, we observe that the spin orientation of the edge states are quite robust against even large RSOC strengths R_0 for the single-edge case. Nevertheless, the spin orientation does change slightly with energy. Moreover, we find good agreement between the numerical and analytical approaches.

Now we turn to the case of a ribbon, where the change in the spin orientation of the GHESs is enhanced substantially for realistic HgTe TIs. The new physical element of the ribbon, compared to the isolated edge, is the coupling of the GHESs on opposite edges. This finite size effect—even without RSOC—produces a gap in the HES spectrum [43]. Now combining the finite width and the RSOC, two gaps and two associated avoided crossings arise in the GHESs spectrum symmetrically around $k = 0$ as shown in Fig. 7. Our analytical approach shows that the interedge RSOC is responsible for the avoided crossings to take place at finite momenta, which is evident from the dispersions in Eq. (27). Moreover, we find the position in momentum of these gaps and their size Δ_{edge} versus the ribbon width. The analytical and numerical results for these quantities compare well, except at certain widths where the full numerical calculation reveals an interesting destructive interference effect. From our analytical approach, we find the GHESs including the RSOC in Eq. (28). Remarkably, they consist of two orthogonal spinors, one on each side of the ribbon. Thus, the states are equally distributed on the two parallel edges. The states become true GHESs with a sizable variation in the spin orientation close to the two avoided crossings in the GHES spectrum, where all the states are close in energy. We show in Fig. 6 that the region in k space of sizable spin-orientation variation

becomes wider, if the ribbon becomes narrower. On the other hand, widening the ribbon increases the maximal value of the projection, which measures the change in spin orientation. To quantify this further, we find the total spin rotation T_s, Eq. (5), which is related to the integral of the spin-orientation variation over the entire region of k space. The numerical calculations show that the total spin rotation decreases with the ribbon width and, moreover, that $T_s \propto R_0^2$ for values of $R_0 \lesssim 0.5A$.

Our analytical GHESs for both the isolated edge and the ribbon open the possibility to study other effects in the presence of RSOC. For instance, scattering of magnetic impurities or the nuclear spins in the crystal could be studied. Furthermore, it would be interesting to explore the transport properties of a ribbon, since we found a significant spin-orientation change.

ACKNOWLEDGMENTS

This work has been funded by Spanish Government projects: FIS2012-33152, FIS2012-34479, MAT2014-58241-P, FIS2015-63770-P and CAM research consortium QUITEMAD+ S2013/ICE-2801. AML acknowledges A. Fernandez Romero and financial support from the Carlsberg Foundation.

APPENDIX A: THE TIME-REVERSAL OPERATOR WITHIN THE BHZ FRAMEWORK

This Appendix provides the time-reversal operator Θ for a wave function expanded in terms of the BHZ basis states, $|E\pm\rangle$ and $|H\pm\rangle$. The time-reversal operator is only defined up to a phase factor and works differently in different bases, so it is important to keep the basis fixed throughout a calculation [56]. Here we use $\Theta = -i\sigma_y K$, where σ_y is a Pauli matrix in spin space and K is the operator for complex conjugation. With this definition of Θ and by writing the BHZ basis states within the envelope function approximation, one obtains

$$\Theta|E\pm\rangle = \mp|E\mp\rangle, \tag{A1a}$$

$$\Theta|H\pm\rangle = \mp|H\mp\rangle, \tag{A1b}$$

see Appendix A of Ref. [34] for a derivation. Therefore the Kramers partner of some wave function $\varphi(x,y)$ written in the BHZ basis $\{|E+\rangle,|H+\rangle,|E-\rangle,|H-\rangle\}$ is

$$\Theta\varphi(x,y) = \Theta\begin{pmatrix}\varphi_{E+}(x,y)\\ \varphi_{H+}(x,y)\\ \varphi_{E-}(x,y)\\ \varphi_{H-}(x,y)\end{pmatrix} = \begin{pmatrix}\varphi^*_{E-}(x,y)\\ \varphi^*_{H-}(x,y)\\ -\varphi^*_{E+}(x,y)\\ -\varphi^*_{H+}(x,y)\end{pmatrix} \tag{A2}$$

and we get $\Theta^2\varphi(x,y) = -\varphi(x,y)$ as expected.

APPENDIX B: THE HELICAL EDGE STATES WITHOUT RASHBA SPIN-ORBIT INTERACTION

In this Appendix, we provide (i) details on the method used to obtain the HESs without the RSOC within the BHZ model and (ii) the HESs obtained for an isolated edge and a finite-width ribbon.

1. On the derivation of the helical edge states

Various methods have been used to study the HESs at the boundary of a TI [43,57,58]. Here we follow Zhou *et al.* [43] and simply set the wave function to zero at the boundary of the TI, which is possible despite the Dirac-like nature of the BHZ model due to the second-order derivatives [59].

Now we provide the overall steps of the derivation in Ref. [43]. The block diagonal form of the BHZ hamiltonian (7) allows one to solve the two blocks separately. Mathematically, each block leads to a homogeneous system of two coupled linear ordinary differential equations with spatially independent coefficients. The upper block gives the following system of differential equations:

$$\begin{pmatrix} M_0 - B_+\left(-\partial_x^2 + k_y^2\right) & A(-i\partial_x + ik_y) \\ A(-i\partial_x - ik_y) & -\left[M_0 - B_-\left(-\partial_x^2 + k_y^2\right)\right]\end{pmatrix}\begin{pmatrix}\varphi_{E+,E}(x,k_y)\\ \varphi_{H+,E}(x,k_y)\end{pmatrix} = E\begin{pmatrix}\varphi_{E+,E}(x,k_y)\\ \varphi_{H+,E}(x,k_y)\end{pmatrix}, \tag{B1}$$

where $B_\pm = B \pm D$. For simplicity, we assume translational symmetry along the y axis such that k_y is a good quantum number. In contrast, we use broken translational symmetry along the x axis, so $k_x = -i\partial_x$ by the Peierls substitution. In other words, Eq. (B1) is for one or more edges parallel to the y axis. This can be varied at will to study the HESs of any geometric structure [60]. Here, the real-space wave function is

$$\varphi_E(x,y) = e^{ik_y y}\varphi_E(x,k_y) = e^{ik_y y}\begin{pmatrix}\varphi_{E+,E}(x,k_y)\\ \varphi_{H+,E}(x,k_y)\end{pmatrix},$$

where the zeros in the two last components of the entire four-vector are implicit, i.e., $\psi_{\uparrow,E} = [\varphi_E,0,0]^T$. The coupling of the two blocks of H_0 is in fact the difficulty that appears by trying to include the RSOC exactly, since four coupled linear differential equations appear.

The mathematical method to solve this kind of system of differential equations (B1) is to substitute $\varphi_E(x,k_y)$ by the ansatz $e^{\lambda x}\phi_\lambda$ and find all possible values of λ. Importantly, the vector ϕ_λ is independent of x. Since the system of differential equations (B1) is linear, the general solution is a linear combination of all possible ansatz solutions, $e^{\lambda_i x}\phi_{\lambda_i}$, weighted by c_i, i.e.,

$$\varphi_E(x,k_y) = \sum_i c_i e^{\lambda_i x}\phi_{\lambda_i}. \tag{B2}$$

To find all possible λ_i, the weights c_i and the dispersion relations E, we use the boundary condition(s) and the

normalization of the wave functions. The wave function is set to zero at the boundaries [43]. Thus the boundary condition for the isolated edge of the half-plane $x > 0$ is

$$\varphi_E(x=0,k_y) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{B3a}$$

and boundary conditions for the ribbon of width W are

$$\varphi_E(x=\pm W/2,k_y) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{B3b}$$

Moreover, the HESs are by definition not extended into the bulk, so we require them to be bounded and normalized in the direction perpendicular to the edge(s), i.e.,

$$\int dx|\varphi_E(x,k_y)|^2 = 1, \tag{B4}$$

where the integral goes from 0 to ∞ in the case of an isolated edge and from $-W/2$ to $W/2$ for a ribbon. Therefore we now have the necessary equations to find all possible λ_i and c_i and the corresponding energy dispersions for the HESs at an isolated boundary and for a ribbon.

2. HESs at an isolated boundary

Applying the method presented above, one can find the pair of HESs without RSOC located at the boundary of the half-plane $x > 0$ to be

$$\psi_{k_y\uparrow}(x,y) = \frac{1}{\sqrt{L}} e^{ik_y y} g_{k_y}(x)\hat{\phi}_\uparrow, \tag{B5a}$$

$$\psi_{k_y\downarrow}(x,y) = \frac{1}{\sqrt{L}} e^{ik_y y} g_{-k_y}(x)\hat{\phi}_\downarrow, \tag{B5b}$$

as presented in Eq. (10) of the main text, however, without the complete specification given below. The energy dispersion relations for the HESs (B5) are

$$E_{\uparrow k_y} = E_0 + \hbar v k_y \text{ and } E_{\downarrow k_y} = E_0 - \hbar v k_y, \tag{B6}$$

as seen in Fig. 9. Both the velocity $v = -\sqrt{B^2 - D^2}\frac{|A|}{\hbar B}$ and the energy shift $E_0 = -\frac{M_0 D}{B}$ are positive for the parameters in Table I. Interestingly, the dispersions are exactly linear for an isolated edge [46]. The k_y-independent spinors $\hat{\phi}_\sigma$ are

$$\hat{\phi}_\uparrow = \mathfrak{n}\begin{pmatrix} -i\frac{A}{|A|} \\ \frac{\sqrt{B_+B_-}}{B_-} \\ 0 \\ 0 \end{pmatrix}, \quad \hat{\phi}_\downarrow = \mathfrak{n}\begin{pmatrix} 0 \\ 0 \\ +i\frac{A}{|A|} \\ \frac{\sqrt{B_+B_-}}{B_-} \end{pmatrix}, \tag{B7}$$

where $B_\pm = B \pm D$ and $\mathfrak{n} = \sqrt{B_-/(2B)}$. The real and normalized transverse wave function $g_{k_y}(x)$ is

$$g_{k_y}(x) = \sqrt{\frac{2\lambda_1\lambda_2(\lambda_1+\lambda_2)}{(\lambda_1-\lambda_2)^2}}(e^{-\lambda_1 x} - e^{-\lambda_2 x}), \tag{B8}$$

where the length scale $1/\lambda_2$ is the penetration length of the HES into the bulk of the TI. Moreover, the k_y dependence of

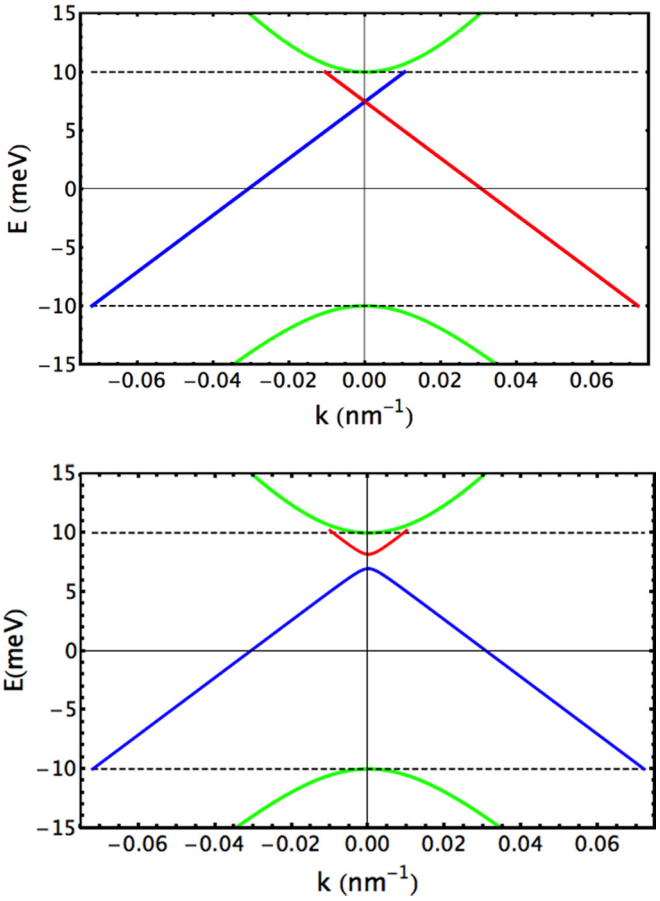

FIG. 9. The HES dispersions (red/blue) and bulk bands (green) without the RSOC for an isolated edge (top panel) and for a 200-nm wide ribbon (bottom panel). The ribbon dispersions $E^+_{k_x}$ (red) and $E^-_{k_x}$ (blue) are spin-degenerate, since they include the HESs on both sides of the ribbon. The BHZ parameters for HgTe in Table I have been used.

g_{k_y} is in λ_1 and λ_2 as

$$\lambda_1 = \frac{1}{\sqrt{B_-B_+}}\left(\frac{|A|}{2} + \sqrt{Z_{k_y}}\right), \tag{B9a}$$

$$\lambda_2 = \frac{1}{\sqrt{B_-B_+}}\left(\frac{|A|}{2} - \sqrt{Z_{k_y}}\right), \tag{B9b}$$

where we defined

$$Z_{k_y} = \left(\frac{A^2}{4} - \frac{M_0}{B}B_+B_-\right) + \frac{D|A|\sqrt{B_+B_-}}{B}k_y + B_+B_-k_y^2. \tag{B10}$$

The BHZ model only hosts HESs in the TI regime where $M_0/B > 0$. Moreover, the explicit forms of the HESs with real $\lambda_{1,2}$ presented here are found under the assumption that $0 \leqslant M_0/B \leqslant A^2/(4B^2)$, which is fulfilled for the parameters in Table I. Furthermore, the HESs are well-localized at the boundary within the bulk energy gap with a fairly small penetration length $1/\lambda_2$ on the order of tens of nm. However, above the upper bulk band gap edge, there is a region of coexistence of edge and bulk states before the penetration length diverges as seen in Fig. 10. Coexistence of bulk and

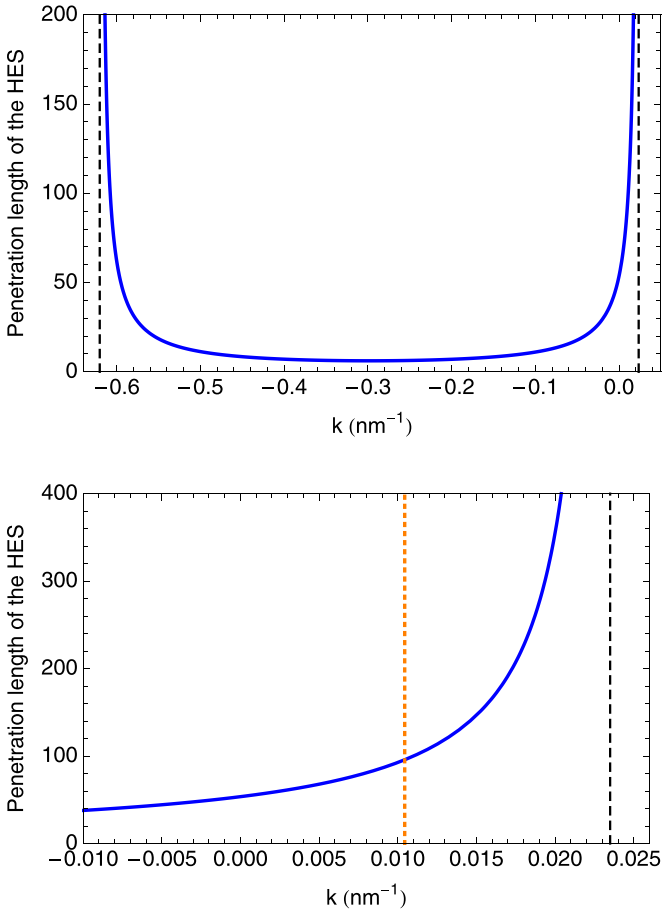

FIG. 10. The penetration length $1/\lambda_2$, Eq. (B9b), of the HES $\psi_{k_y\uparrow}$ into the bulk of the TI. (The penetration length for the other HES $\psi_{k_y\downarrow}$ is found by the replacement $k_y \to -k_y$.) The upper figure shows the large region of finite penetration length and the points where it diverges, $\lambda_2 = 0$, indicated by vertical dashed black lines. The lower figure focuses on the region close to the upper edge of the bulk band gap shown by the vertical orange dotted line. A region of coexistence of edge and bulk bands is clearly seen (between the vertical orange dotted and black dashed lines), even though the HES widens and gradually looses the localization characteristic of an *edge* state. Nevertheless, this facilitates the use of our analytical theory in Sec. IV A beyond the boundaries of the bulk band gap, since it requires the presence of both HESs without RSOC. The parameters for HgTe in Table I are used.

edge states has recently been studied [61]. Finally, we remark that $\psi_{k_y\uparrow}$ and $\psi_{-k_y\downarrow}$ are Kramers partners, since $\Theta\psi_{k_y\uparrow}(x,y) = -\psi_{-k_y\downarrow}(x,y)$ by using Eq. (A2).

In passing, we remark that the HESs along the perpendicular direction are different in a nontrivial way from the ones presented in Eq. (B5). If we consider the HESs running along the x axis instead of the y axis at the boundary of the half-plane $y > 0$, then we find $\psi_{k_x\sigma}(x,y) = \frac{1}{\sqrt{L}} e^{ik_x x} g_{-\mathfrak{s}k_x}(y)\tilde{\phi}_\sigma$ and $E_{\sigma k_x} = E_0 - \mathfrak{s}\hbar v k_x$, where $\mathfrak{s} = +(-)$ for $\sigma = \uparrow(\downarrow)$. The exchange of k_y by $-k_x$ is natural in order for the states to be connected correctly in, e.g., the corner of the TI [34]. A more interesting fact is that the imaginary unit i disappears from the spinors, i.e.,

$$\tilde{\phi}_\uparrow = \mathfrak{n}\begin{pmatrix} \frac{A}{|A|} \\ \frac{\sqrt{B_+B_-}}{B_-} \\ 0 \\ 0 \end{pmatrix}, \quad \tilde{\phi}_\downarrow = \mathfrak{n}\begin{pmatrix} 0 \\ 0 \\ \frac{A}{|A|} \\ \frac{\sqrt{B_+B_-}}{B_-} \end{pmatrix}, \tag{B11}$$

where $\mathfrak{n} = \sqrt{B_-/(2B)}$ as in Eq. (B7).

3. HESs for a finite width ribbon

Now we turn to the HESs for a ribbon with edges at $y = \pm W/2$ [i.e., parallel to the x axis in contrast to the case in Eq. (B5)] as considered by Zhou *et al.* [43]. By use of the ansatz function $e^{\lambda y}\phi_\lambda$ and the boundary conditions, the energy dispersions $E^{e=\pm}_{k_x}$ discussed in Sec. IV B become the solutions to the following implicit equation:

$$\frac{\tanh\left(\frac{\lambda_1 W}{2}\right)}{\tanh\left(\frac{\lambda_2 W}{2}\right)} + \frac{\tanh\left(\frac{\lambda_2 W}{2}\right)}{\tanh\left(\frac{\lambda_1 W}{2}\right)} = \frac{\alpha_1^2\lambda_2^2 + \alpha_2^2\lambda_1^2 - k_x^2(\alpha_1 - \alpha_2)^2}{\alpha_1\alpha_2\lambda_1\lambda_2}, \tag{B12}$$

where we have introduced

$$\lambda_1^2 = k_x^2 + F + \sqrt{F^2 - \frac{M_0^2 - E^2}{B^2 - D^2}}, \tag{B13a}$$

$$\lambda_2^2 = k_x^2 + F - \sqrt{F^2 - \frac{M_0^2 - E^2}{B^2 - D^2}}, \tag{B13b}$$

$$\alpha_j = E - M_0 + B_+\left(k_x^2 - \lambda_j^2\right) \quad \text{for} \quad j = 1,2, \tag{B13c}$$

and $F = [A^2 - 2(M_0 B + ED)]/(2B_+B_-)$. As in the previous section, we assume that $\lambda_{1,2}$ are real and in particular define $\lambda_{1,2}$ to be the positive root in Eq. (B13), i.e., $\lambda_{1,2} > 0$. Here $\lambda_{1,2}$ are not identical to the ones for an isolated edge in Eq. (B9), since the energy dispersions differ in the two cases. Moreover, the dispersions for an isolated edge Eq. (B6) come out correctly in the limit $W \to \infty$, where the left-hand side of Eq. (B12) is equal to 2. The dispersions $E^{\pm}_{k_x}$ for a ribbon are seen in Fig. 9, where $E^{+}_{k_x}$ is the upper dispersion ($E^{+}_{k_x} > E^{-}_{k_x}$).

The four HESs $\psi^e_{k_x\sigma}$ for the ribbon are all proportional to a plane-wave running along the edges, i.e., $\psi^e_{k_x\sigma}(x,y) \propto e^{ik_x x}$. Moreover, the spinor and the transverse wave function do not factorize in contrast to the case of an isolated edge in Eq. (B5). For a ribbon, the HESs are

$$\psi^+_{k_x\uparrow}(x,y) = \tilde{c}_+ \frac{e^{ik_x x}}{\sqrt{L}} \begin{pmatrix} f_+ - \gamma^+_{k_x} f_- \\ \gamma^+_{k_x}\eta_2^+ f_+ - \eta_1^+ f_- \\ 0 \\ 0 \end{pmatrix}, \tag{B14a}$$

$$\psi^-_{k_x\uparrow}(x,y) = \tilde{c}_- \frac{e^{ik_x x}}{\sqrt{L}} \begin{pmatrix} -\gamma^-_{k_x} f_+ + f_- \\ -\eta_2^- f_+ + \gamma^-_{k_x}\eta_1^- f_- \\ 0 \\ 0 \end{pmatrix}, \tag{B14b}$$

$$\psi^+_{k_x\downarrow}(x,y) = -\tilde{c}_+ \frac{e^{ik_x x}}{\sqrt{L}} \begin{pmatrix} 0 \\ 0 \\ f_+ + \gamma^+_{k_x} f_- \\ -\gamma^+_{k_x}\eta_2^+ f_+ - \eta_1^+ f_- \end{pmatrix}, \tag{B14c}$$

$$\psi^{-}_{k_x\downarrow}(x,y) = -\tilde{c}_- \frac{e^{ik_x x}}{\sqrt{L}} \begin{pmatrix} 0 \\ 0 \\ \gamma^-_{k_x} f_+ + f_- \\ -\eta^-_2 f_+ - \gamma^-_{k_x}\eta^-_1 f_- \end{pmatrix}. \quad \text{(B14d)}$$

All the spatial dependence of $\psi^e_{k_x\sigma}(x,y)$ are in the functions $f_\pm \equiv f_\pm(y,k_x,E)$, where the subscript denotes the parity, i.e., $f_\pm(-y,k_x,E) = \pm f_\pm(y,k_x,E)$. These are

$$f_+(y,k_x,E) = \left(\frac{\cosh(\lambda_1 y)}{\cosh\left(\frac{\lambda_1 W}{2}\right)} - \frac{\cosh(\lambda_2 y)}{\cosh\left(\frac{\lambda_2 W}{2}\right)} \right), \quad \text{(B15a)}$$

$$f_-(y,k_x,E) = \left(\frac{\sinh(\lambda_1 y)}{\sinh\left(\frac{\lambda_1 W}{2}\right)} - \frac{\sinh(\lambda_2 y)}{\sinh\left(\frac{\lambda_2 W}{2}\right)} \right), \quad \text{(B15b)}$$

which vanish on the boundaries $y = \pm W/2$. Here both k_x and E are explicitly written as variables in $f_\pm$ in order to keep track of which dispersion is used, $E^+_{k_x}$ or $E^-_{k_x}$. The HESs (B14) also include the following space-independent quantities:

$$\eta^\pm_1 = \left. \frac{\alpha_2 - \alpha_1}{A\left[\lambda_1 \coth\left(\frac{\lambda_1 W}{2}\right) - \lambda_2 \coth\left(\frac{\lambda_2 W}{2}\right)\right]} \right|_{E=E^\pm_{k_x}},$$

$$\eta^\pm_2 = \left. \frac{\alpha_2 - \alpha_1}{A\left[\lambda_1 \tanh\left(\frac{\lambda_1 W}{2}\right) - \lambda_2 \tanh\left(\frac{\lambda_2 W}{2}\right)\right]} \right|_{E=E^\pm_{k_x}},$$

$$\gamma^+_{k_x} = \left. \frac{(\alpha_2 - \alpha_1)k_x}{\alpha_2\lambda_1 \tanh\left(\frac{\lambda_1 W}{2}\right) - \alpha_1\lambda_2 \tanh\left(\frac{\lambda_2 W}{2}\right)} \frac{\eta^+_1}{\eta^+_2} \right|_{E=E^+_{k_x}},$$

$$\gamma^-_{k_x} = \left. \frac{(\alpha_2 - \alpha_1)k_x}{\alpha_2\lambda_1 \coth\left(\frac{\lambda_1 W}{2}\right) - \alpha_1\lambda_2 \coth\left(\frac{\lambda_2 W}{2}\right)} \frac{\eta^-_2}{\eta^-_1} \right|_{E=E^-_{k_x}},$$

which all depend on k_x and the energy dispersions $E = E^\pm_{k_x}$. We remark that some of the signs in the HESs presented in Eq. (B14) are not identical to the ones found in Ref. [43]. The reason is that Zhou *et al.* [43] have the opposite sign in front of k_y in the BHZ Hamiltonian (7) compared to the one used here and in, e.g., Ref. [35]. (In fact, the sign of k_y in the BHZ Hamiltonian varies throughout the literature.)

Finally, we find the normalization constants to be

$$\tilde{c}_+ = \frac{1}{\sqrt{\Gamma^+_{++}\left[1 + \left(\gamma^+_{k_x}\right)^2(\eta^+_2)^2\right] + \Gamma^-_{++}\left[\left(\gamma^+_{k_x}\right)^2 + (\eta^+_1)^2\right]}},$$

$$\tilde{c}_- = \frac{1}{\sqrt{\Gamma^+_{--}\left[\left(\gamma^-_{k_x}\right)^2 + (\eta^-_2)^2\right] + \Gamma^-_{--}\left[1 + \left(\gamma^-_{k_x}\right)^2(\eta^-_1)^2\right]}},$$

where we introduced

$$\Gamma^\tau_{ee'} = \int_{-W/2}^{W/2} dy\ f_\tau\left(y,k_x,E^e_{k_x}\right) f_\tau\left(y,k_x,E^{e'}_{k_x}\right). \quad \text{(B16)}$$

The expressions for $\Gamma^+_{\pm\pm}$ and $\Gamma^-_{\pm\pm}$ are given in Eqs. (C10) and (C11), respectively. The integral over two functions of opposite parity is zero, $\int dy f_+ f_- = 0$, so there is no need to include this possibility in the definition of $\Gamma^\tau_{ee'}$.

For the HESs of a ribbon, the Kramers partner of $\psi^\pm_{k_x\uparrow}$ is $\psi^\pm_{-k_x\downarrow}$, since $\Theta\psi^\pm_{k_x\uparrow}(x,y) = \psi^\pm_{-k_x\downarrow}(x,y)$ by the help of Eq. (A2). To find this result, we use that the energy dispersion is even in k_x, $E^\pm_{-k_x} = E^\pm_{k_x}$, as seen from Eq. (B12). This in terms leads to $\lambda_{1,2}$ in Eq. (B13) and $\eta^\pm_{1,2}$ being even in k_x and finally that $\gamma^\pm_{-k_x} = -\gamma^\pm_{k_x}$.

APPENDIX C: DETAILS OF THE CALCULATION WITH RSOC

In this paper, we find the analytical forms of the GHESs in the presence of RSOC by assuming that the GHESs can be written as combinations of the HESs without RSOC. Therefore we diagonalized the RSOC $\mathcal{H}_R$ in two bases, namely $\{|\psi_{k_x\uparrow}\rangle, |\psi_{k_x\downarrow}\rangle\}$ for an isolated edge and $\{|\psi^+_{k_x\uparrow}\rangle, |\psi^-_{k_x\downarrow}\rangle, |\psi^-_{k_x\uparrow}\rangle, |\psi^+_{k_x\downarrow}\rangle\}$ for a ribbon. Here we provide various technical details for these calculations left out in the main text.

1. Details for the case of an isolated edge

In Eq. (13) in Sec. IV A, we only give the effective RSOC α_{k_y} up to second order in k_y. However, the exact result can easily be found to be

$$\begin{aligned} \alpha_{k_y} &= \frac{\langle\psi_{k_y\uparrow}|H_R|\psi_{k_y\downarrow}\rangle}{k_y} \\ &= R_0 \frac{B-D}{2Bk_y} \int_0^\infty dx g_{k_y}(x)[\partial_x + k_y] g_{-k_y}(x) \\ &= R_0 \frac{B-D}{2Bk_y}\left(\nu_{k_y} + k_y \xi_{k_y}\right), \end{aligned} \quad \text{(C1)}$$

where we use the transverse wave functions $g_{\pm k_y}(x)$ for an isolated edge Eq. (B8) and introduce

$$\xi_{k_y} \equiv \int_0^\infty dx g_{k_y}(x) g_{-k_y}(x), \quad \text{(C2a)}$$

$$\nu_{k_y} \equiv \int_0^\infty dx g_{k_y}(x) \partial_x g_{-k_y}(x). \quad \text{(C2b)}$$

Here it is evident that $\alpha_{-k_y} = \alpha_{k_y}$, since $\xi_{-k_y} = \xi_{k_y}$ and $\nu_{-k_y} = -\nu_{k_y}$ by using partial integration. Moreover, we find that $\xi_{k_y=0} = 1$ due to the normalization of g_{k_y}. The full expressions for ξ_{k_y} and ν_{k_y} are

$$\xi_{k_y} = h(k_y)\omega(k_y) \quad \text{and} \quad \nu_{k_y} = h(k_y)\beta(k_y), \quad \text{(C3)}$$

where

$$\begin{aligned} h(k_y) &= 2\sqrt{\frac{\lambda^-_1\lambda^-_2(\lambda^-_1 + \lambda^-_2)}{(\lambda^-_1 - \lambda^-_2)^2}} \sqrt{\frac{\lambda^+_1\lambda^+_2(\lambda^+_1 + \lambda^+_2)}{(\lambda^+_1 - \lambda^+_2)^2}}, \\ \omega(k_y) &= \frac{1}{\lambda^-_1 + \lambda^+_1} - \frac{1}{\lambda^-_2 + \lambda^+_1} - \frac{1}{\lambda^-_1 + \lambda^+_2} + \frac{1}{\lambda^-_2 + \lambda^+_2}, \\ \beta(k_y) &= \lambda^-_1\left(\frac{1}{\lambda^-_1 + \lambda^+_2} - \frac{1}{\lambda^-_1 + \lambda^+_1}\right) \\ &\quad + \lambda^-_2\left(\frac{1}{\lambda^+_1 + \lambda^-_2} - \frac{1}{\lambda^-_2 + \lambda^+_2}\right). \end{aligned} \quad \text{(C4)}$$

Here we used the shorthand notation $\lambda^\pm_i = \lambda_i(\pm k_y)$, where $\lambda_{1,2}$ are given in Eq. (B9). Note that $\lambda_{1,2}$ are not even functions of k_y for an isolated edge in contrast to $\lambda_{1,2}$ for a ribbon. The reason is that the dispersions are not even in k_y for an isolated edge as they are for a ribbon. For the parameters in Table I,

the exact result for α_{k_y}, Eq. (C1), is very well approximated by the second-order expansion given in Eq. (13) in the main text.

In the limit of a particle-hole symmetric BHZ Hamiltonian, i.e., $D = 0$, we have $g_{k_y}(x) = g_{-k_y}(x)$, since $Z_{k_y}^{(D=0)}$, Eq. (B10), and thereby also $\lambda_{1,2}^{(D=0)}$, Eq. (B9), become even. This means that $\xi_{k_y}^{D=0} = 1$ for all k_y by normalization of g_{k_y}. Moreover, $g_{k_y}(x) = g_{-k_y}(x)$ forces $\nu_{k_y}^{(D=0)}$ to be even, but we already found ν_{k_y} to be generally odd, so we have to conclude that $\nu_{k_y}^{(D=0)} = 0$. Therefore we are left with the exact result

$$\alpha_{k_y}^{(D=0)} = R_0/2. \tag{C5}$$

As discussed in the main text [below Eq. (19)], such an k_y-independent α_{k_y} means that the spin orientation remains fixed for $R_0 \neq 0$, so no GHESs appear for $D = 0$.

2. Details for the ribbon calculation

In this section, we provide the matrix elements of the RSOC H_R between the HESs $\psi_{k_x\sigma}^e$ for a ribbon, which were used—but not given explicitly—in Sec. IV B. Using the HESs $\psi_{k_x\sigma}^e$, Eq. (B14), we find the matrix elements to be diagonal in k_x due to the plane waves running along the edges, $\psi_{k_x\sigma}^e \propto e^{ik_x x}$. Moreover, H_R, Eq. (8), only couples opposite spins, i.e.,

$$\langle \psi_{k_x\sigma}^e | H_R | \psi_{k_x'\sigma'}^{e'} \rangle = \int_{-W/2}^{W/2} dy \int_{-L/2}^{L/2} dx \left[\psi_{k_x\sigma}^e(x,y)\right]^\dagger H_R \psi_{k_x'\sigma'}^{e'}(x,y) \propto \delta_{k_x,k_x'}\delta_{\bar{\sigma}\sigma'},$$

where $\bar{\sigma}$ denote the opposite spin of σ. Therefore out of the ten possible matrix elements of H_R in Eq. (24) (accounting for the hermiticity of H_R), we are now left with only four different nonzero matrix elements. Moreover, we find below that $\langle \psi_{k_x\downarrow}^- | H_R | \psi_{k_x\uparrow}^+ \rangle = -\langle \psi_{k_x\downarrow}^+ | H_R | \psi_{k_x\uparrow}^- \rangle$ such that only three different nonzero matrix elements of H_R remain in Eq. (24).

The interedge matrix elements in Eq. (25) involving HESs localized on opposite edges are

$$id_+ = \langle \psi_{k_x\downarrow}^+ | H_R | \psi_{k_x\uparrow}^+ \rangle = -iR_0(\tilde{c}_+)^2\left[k_x\left[\Gamma_{++}^+ - \left(\gamma_{k_x}^+\right)^2\Gamma_{++}^-\right] - 2\gamma_{k_x}^+\Omega_{++}^{+-}\right], \tag{C6a}$$

$$id_- = \langle \psi_{k_x\uparrow}^- | H_R | \psi_{k_x\downarrow}^- \rangle = +iR_0(\tilde{c}_-)^2\left[k_x\left[\Gamma_{--}^- - \left(\gamma_{k_x}^-\right)^2\Gamma_{--}^+\right] + 2\gamma_{k_x}^-\Omega_{--}^{+-}\right], \tag{C6b}$$

while the intraedge matrix element (26) between HESs on the same edge is

$$ib = \langle \psi_{k_x\downarrow}^- | H_R | \psi_{k_x\uparrow}^+ \rangle = iR_0\tilde{c}_+\tilde{c}_-\left[k_x\left(\gamma_{k_x}^+\Gamma_{-+}^- - \gamma_{k_x}^-\Gamma_{-+}^+\right) - \Omega_{-+}^{-+} + \gamma_{k_x}^-\gamma_{k_x}^+\Omega_{-+}^{+-}\right]. \tag{C7}$$

These elements are written in terms of the quantities appearing in the ribbon HESs given in Sec. B 3 and integrals involving $f_\pm$ in Eq. (B15), namely, $\Gamma_{ee'}^\tau$ in Eq. (B16) and

$$\Omega_{ee'}^{\tau\tau'} = \int_{-W/2}^{W/2} dy\ f_\tau\left(y,k_x,E_{k_x}^e\right)\partial_y f_{\tau'}\left(y,k_x,E_{k_x}^{e'}\right). \tag{C8}$$

The parity of $f_\pm$ leads to $\Omega_{ee'}^{++} = \Omega_{ee'}^{--} = 0$ and a partial integration (with zero boundary term) gives $\Omega_{ee'}^{\tau\tau'} = -\Omega_{e'e}^{\tau'\tau}$.

Now we turn to the intraedge matrix element between the HESs localized on the opposite edge compared to the HESs in $ib = \langle \psi_{k_x\downarrow}^- | H_R | \psi_{k_x\uparrow}^+ \rangle$ and show that we find the same result. As discussed before, see Eq. (22), the HESs $\{\psi_{k_x\downarrow}^+, \psi_{k_x\uparrow}^-\}$ are localized on the opposite edge of the HESs $\{\psi_{k_x\downarrow}^-, \psi_{k_x\uparrow}^+\}$, which are used in ib. Direct calculation gives

$$\langle \psi_{k_x\downarrow}^+ | H_R | \psi_{k_x\uparrow}^- \rangle = iR_0\tilde{c}_+\tilde{c}_-\left[k_x\left(\gamma_{k_x}^-\Gamma_{+-}^+ - \gamma_{k_x}^+\Gamma_{+-}^-\right) - \Omega_{+-}^{+-} + \gamma_{k_x}^-\gamma_{k_x}^+\Omega_{+-}^{-+}\right].$$

Therefore, by comparing to ib in Eq. (C7) and using that $\Omega_{ee'}^{\tau\tau'} = -\Omega_{e'e}^{\tau'\tau}$ and $\Gamma_{ee'}^\tau = \Gamma_{e'e}^\tau$, we find

$$\langle \psi_{k_x\downarrow}^+ | H_R | \psi_{k_x\uparrow}^- \rangle = -\langle \psi_{k_x\downarrow}^- | H_R | \psi_{k_x\uparrow}^+ \rangle = -ib. \tag{C9}$$

This is a physically sound result, since the two edges are physically equivalent.

Finally, we give the integrals $\Gamma_{ee'}^\tau$ in Eq. (B16) and $\Omega_{ee'}^{\tau\tau'}$ in Eq. (C8). There are six nonzero $\Gamma_{ee'}^\tau$. These are

$$\Gamma_{\pm\pm}^+ = \frac{W}{2}\left[\frac{1}{\cosh^2\left(\frac{W\lambda_1^\pm}{2}\right)} + \frac{1}{\cosh^2\left(\frac{W\lambda_2^\pm}{2}\right)}\right] + \frac{\lambda_1^\pm[3(\lambda_2^\pm)^2 + (\lambda_1^\pm)^2]\tanh\left(\frac{W\lambda_2^\pm}{2}\right) - \lambda_2^\pm[3(\lambda_1^\pm)^2 + (\lambda_2^\pm)^2]\tanh\left(\frac{W\lambda_1^\pm}{2}\right)}{\lambda_1^\pm\lambda_2^\pm[(\lambda_1^\pm)^2 - (\lambda_2^\pm)^2]}, \tag{C10}$$

where $\lambda_{1,2}^\pm = \lambda_{1,2}(E_{k_x}^\pm)$ in Eq. (B13). Note that the notation $\lambda_{1,2}^\pm$ is not identical to the one used in Sec. C 1 for an isolated edge. Moreover, we obtain

$$\Gamma_{\pm\pm}^- = -\frac{W}{2}\left[\frac{1}{\sinh^2\left(\frac{W\lambda_1^\pm}{2}\right)} + \frac{1}{\sinh^2\left(\frac{W\lambda_2^\pm}{2}\right)}\right] + \frac{\lambda_1^\pm[3(\lambda_2^\pm)^2 + (\lambda_1^\pm)^2]\coth\left(\frac{W\lambda_2^\pm}{2}\right) - \lambda_2^\pm[3(\lambda_1^\pm)^2 + (\lambda_2^\pm)^2]\coth\left(\frac{W\lambda_1^\pm}{2}\right)}{\lambda_1^\pm\lambda_2^\pm[(\lambda_1^\pm)^2 - (\lambda_2^\pm)^2]}, \tag{C11}$$

which is related to $\Gamma^{+}_{\pm\pm}$ by exchanging $\cosh^2(W\lambda_i^\tau/2) \to -\sinh^2(W\lambda_i^\tau/2)$ and $\tanh(W\lambda_i^\tau/2) \to \coth(W\lambda_i^\tau/2) = 1/\tanh(W\lambda_i^\tau/2)$. We remark that $\Gamma^{\tau}_{\pm\pm}$ are invariant under interchange of $\lambda_1^\pm$ and $\lambda_2^\pm$. The $\Gamma^{\tau}_{ee'}$ with $e \neq e'$ are

$$\Gamma^{+}_{-+} = \Gamma^{+}_{+-} = \frac{2\lambda_1^- \tanh\left(\frac{W\lambda_1^-}{2}\right) - 2\lambda_1^+ \tanh\left(\frac{W\lambda_1^+}{2}\right)}{(\lambda_1^-)^2 - (\lambda_1^+)^2} + \frac{2\lambda_2^- \tanh\left(\frac{W\lambda_2^-}{2}\right) - 2\lambda_2^+ \tanh\left(\frac{W\lambda_2^+}{2}\right)}{(\lambda_2^-)^2 - (\lambda_2^+)^2}$$
$$- \frac{2\lambda_1^+ \tanh\left(\frac{W\lambda_1^+}{2}\right) - 2\lambda_2^- \tanh\left(\frac{W\lambda_2^-}{2}\right)}{(\lambda_1^+)^2 - (\lambda_2^-)^2} - \frac{2\lambda_1^- \tanh\left(\frac{W\lambda_1^-}{2}\right) - 2\lambda_2^+ \tanh\left(\frac{W\lambda_2^+}{2}\right)}{(\lambda_1^-)^2 - (\lambda_2^+)^2} \tag{C12}$$

and

$$\Gamma^{-}_{-+} = \Gamma^{-}_{+-} = \frac{2\lambda_1^- \coth\left(\frac{W\lambda_1^-}{2}\right) - 2\lambda_1^+ \coth\left(\frac{W\lambda_1^+}{2}\right)}{(\lambda_1^-)^2 - (\lambda_1^+)^2} + \frac{2\lambda_2^- \coth\left(\frac{W\lambda_2^-}{2}\right) - 2\lambda_2^+ \coth\left(\frac{W\lambda_2^+}{2}\right)}{(\lambda_2^-)^2 - (\lambda_2^+)^2}$$
$$- \frac{2\lambda_1^+ \coth\left(\frac{W\lambda_1^+}{2}\right) - 2\lambda_2^- \coth\left(\frac{W\lambda_2^-}{2}\right)}{(\lambda_1^+)^2 - (\lambda_2^-)^2} - \frac{2\lambda_1^- \coth\left(\frac{W\lambda_1^-}{2}\right) - 2\lambda_2^+ \coth\left(\frac{W\lambda_2^+}{2}\right)}{(\lambda_1^-)^2 - (\lambda_2^+)^2}, \tag{C13}$$

where $\Gamma^{\tau}_{ee'} = \Gamma^{\tau}_{e'e}$ follows from the definition of $\Gamma^{\tau}_{ee'}$ in Eq. (B16). Here we observe that $\Gamma^{-}_{\mp\pm}$ is related to $\Gamma^{+}_{\mp\pm}$ by interchanging $\tanh(W\lambda_i^\tau/2)$ and $\coth(W\lambda_i^\tau/2)$.

Furthermore, there are four different nonzero $\Omega^{\tau\tau'}_{ee'}$ (remembering that $\Omega^{\tau\tau'}_{ee'} = -\Omega^{\tau'\tau}_{e'e}$ and $\Omega^{\pm\pm}_{ee'} = 0$). These are

$$\Omega^{+-}_{\pm\pm} = W\left[\frac{\lambda_1^\pm}{\sinh(W\lambda_1^\pm)} + \frac{\lambda_2^\pm}{\sinh(W\lambda_2^\pm)}\right] + \frac{2\lambda_1^\pm\lambda_2^\pm}{(\lambda_1^\pm)^2 - (\lambda_2^\pm)^2}\left[\frac{\tanh\left(\frac{W\lambda_2^\pm}{2}\right)}{\tanh\left(\frac{W\lambda_1^\pm}{2}\right)} - \frac{\tanh\left(\frac{W\lambda_1^\pm}{2}\right)}{\tanh\left(\frac{W\lambda_2^\pm}{2}\right)}\right], \tag{C14}$$

$$\Omega^{-+}_{-+} = -2\Bigg[-\frac{(\lambda_1^+)^2}{(\lambda_1^+)^2 - (\lambda_1^-)^2} + \frac{(\lambda_1^+)^2}{(\lambda_1^+)^2 - (\lambda_2^-)^2} - \frac{(\lambda_1^-)^2}{(\lambda_1^-)^2 - (\lambda_2^+)^2} + \frac{(\lambda_2^-)^2}{(\lambda_2^-)^2 - (\lambda_2^+)^2}$$
$$- \frac{\lambda_1^-\lambda_1^+}{(\lambda_1^-)^2 - (\lambda_1^+)^2}\frac{\tanh\left(\frac{\lambda_1^+ W}{2}\right)}{\tanh\left(\frac{\lambda_1^- W}{2}\right)} + \frac{\lambda_1^-\lambda_2^+}{(\lambda_1^-)^2 - (\lambda_2^+)^2}\frac{\tanh\left(\frac{\lambda_2^+ W}{2}\right)}{\tanh\left(\frac{\lambda_1^- W}{2}\right)}$$
$$+ \frac{\lambda_1^+\lambda_2^-}{(\lambda_2^-)^2 - (\lambda_1^+)^2}\frac{\tanh\left(\frac{\lambda_1^+ W}{2}\right)}{\tanh\left(\frac{\lambda_2^- W}{2}\right)} - \frac{\lambda_2^-\lambda_2^+}{(\lambda_2^-)^2 - (\lambda_2^+)^2}\frac{\tanh\left(\frac{\lambda_2^+ W}{2}\right)}{\tanh\left(\frac{\lambda_2^- W}{2}\right)}\Bigg], \tag{C15}$$

and

$$\Omega^{+-}_{-+} = +2\Bigg[-\frac{(\lambda_1^-)^2}{(\lambda_1^-)^2 - (\lambda_1^+)^2} + \frac{(\lambda_1^-)^2}{(\lambda_1^-)^2 - (\lambda_2^+)^2} - \frac{(\lambda_1^+)^2}{(\lambda_1^+)^2 - (\lambda_2^-)^2} + \frac{(\lambda_2^+)^2}{(\lambda_2^+)^2 - (\lambda_2^-)^2}$$
$$- \frac{\lambda_1^+\lambda_1^-}{(\lambda_1^+)^2 - (\lambda_1^-)^2}\frac{\tanh\left(\frac{\lambda_1^- W}{2}\right)}{\tanh\left(\frac{\lambda_1^+ W}{2}\right)} + \frac{\lambda_1^+\lambda_2^-}{(\lambda_1^+)^2 - (\lambda_2^-)^2}\frac{\tanh\left(\frac{\lambda_2^- W}{2}\right)}{\tanh\left(\frac{\lambda_1^+ W}{2}\right)}$$
$$+ \frac{\lambda_1^-\lambda_2^+}{(\lambda_2^+)^2 - (\lambda_1^-)^2}\frac{\tanh\left(\frac{\lambda_1^- W}{2}\right)}{\tanh\left(\frac{\lambda_2^+ W}{2}\right)} - \frac{\lambda_2^+\lambda_2^-}{(\lambda_2^+)^2 - (\lambda_2^-)^2}\frac{\tanh\left(\frac{\lambda_2^- W}{2}\right)}{\tanh\left(\frac{\lambda_2^+ W}{2}\right)}\Bigg]. \tag{C16}$$

We note that $\Omega^{+-}_{-+} = -\Omega^{-+}_{+-}$ and that Ω^{-+}_{+-} can be found by interchanging $E^+_{k_x}$ and $E^-_{k_x}$ in Ω^{-+}_{-+} in Eq. (C15) (i.e., interchanging λ_i^+ and λ_i^- for $i = 1,2$). We also remark that all the integrals $\Gamma^{\tau}_{ee'}$ and $\Omega^{\tau\tau'}_{ee'}$ are even in k_x, since $\lambda^\pm_{1,2} = \lambda_{1,2}(E^\pm_{k_x})$ is even in k_x. We now have all the integrals $\Gamma^{\tau}_{ee'}$ and $\Omega^{\tau\tau'}_{ee'}$ appearing in the matrix elements $id_\pm$ Eq. (C6) and ib Eq. (C7).

APPENDIX D: ON THE NUMERICAL TIGHT-BINDING FORMULATION

In this Appendix, we briefly discuss the lattice regularization of the BHZ model and its formulation for the ribbon geometry. In order to map a continuous model to a tight-binding model, we use the standard tight-binding regularization procedure. For a 2D square lattice, this consists in making the replacements

$$k_i \to \frac{1}{a}\sin k_i a, \tag{D1}$$

$$k_i^2 \to \frac{2}{a^2}(1 - \cos k_i a), \tag{D2}$$

with a being the lattice constant and $i = x, y$. The two quantities are equal only in the long wavelength limit, $k_i a \to 0$. This tight-binding regularization has been extensively used in the literature to study the BHZ Hamiltonian [45,62]. The fermion doubling problem that usually occurs when discretizing massless Dirac particles does not directly affect our calculations as the RSOC term breaks

chiral symmetry, which is one of the conditions for the fullfillment of the no-go theorem by Nielsen and Ninomiya [63,64]. Moreover, we have checked that the topological properties of the Hamiltonian are unchanged in the tight-binding version for the parametric regimes that we have explored.

To transfer the Hamiltonian from momentum space onto a real-space lattice, we perform a Fourier transformation. The chosen form of the tight-binding regularization implies that the hopping terms in the lattice model exist only between nearest-neighbor sites. For the calculations presented in the paper, we have a ribbon of finite width in one direction (y) and periodic boundary conditions in the orthogonal direction (x). Thus we perform a Fourier transformation only in the direction of finite width (y) and obtain a k_x-dependent Hamiltonian:

$$\mathcal{H}(k_x) = \sum_j \mathcal{H}_{jj} c_j^\dagger c_j + (\mathcal{H}_{jj+1} c_j^\dagger c_{j+1} + \text{H.c.}), \tag{D3}$$

$$\mathcal{H}_{jj} = \begin{pmatrix} M - 2B_+(2-\cos k_x) & A\sin k_x & -iR_0 \sin k_x & 0 \\ A\sin k_x & -M + 2B_-(2-\cos k_x) & 0 & 0 \\ iR_0 \sin k_x & 0 & M - 2B_+(2-\cos k_x) & -A\sin k_x \\ 0 & 0 & -A\sin k_x & -M + 2B_-(2-\cos k_x) \end{pmatrix}, \tag{D4}$$

$$\mathcal{H}_{jj+1} = \frac{1}{2}\begin{pmatrix} 2B_+ & +A & +iR_0 & 0 \\ -A & -2B_- & 0 & 0 \\ +iR_0 & 0 & 2B_+ & +A \\ 0 & 0 & -A & -2B_- \end{pmatrix}. \tag{D5}$$

The number of sites in the simulation varied between 200 and 2000. We set the value of the lattice spacing a such that we obtain the required width for the particular case under study. Changing the number of sites and a help us to make sure that the results for the relevant values of k did not depend on the details of the tight-binding regularization.

APPENDIX E: HIGHER-ORDER RASHBA SPIN-ORBIT COUPLINGS

Rothe *et al.* [35] derived the RSOC in the BHZ basis up to third order in the momentum. The calculations in the main text only include the first-order term as seen in Eq. (8). Here we discuss the effects of the second and third-order terms on a pair of HESs at an isolated boundary.

The RSOC Hamiltonian in the BHZ basis to third order is [35]

$$\begin{aligned} H_R &= H_R^{(1)} + H_R^{(2)} + H_R^{(3)} \\ &= \begin{pmatrix} 0 & 0 & -iR_0 k_- & -S_0 k_-^2 \\ 0 & 0 & S_0 k_-^2 & iT_0 k_-^3 \\ iR_0 k_+ & S_0 k_+^2 & 0 & 0 \\ -S_0 k_+^2 & -iT_0 k_+^3 & 0 & 0 \end{pmatrix}, \end{aligned} \tag{E1}$$

where $k_\pm = k_x \pm ik_y$ and the superscript n on $H_R^{(n)}$ indicate the order of the momentum that it represents, i.e., the main text only use $H_R = H_R^{(1)}$ in Eq. (8). Here each order in momentum has its own constant prefactor, namely R_0, S_0, and T_0.

To incorporate the higher-order RSOC terms $H_R^{(2)}$ and $H_R^{(3)}$ into the 2×2 Hamiltonian (12) for a pair of HESs at an isolated edge, we need to include the matrix elements $\langle \psi_{k_y\uparrow} | H_R^{(2)} | \psi_{k_y\downarrow} \rangle$ and $\langle \psi_{k_y\uparrow} | H_R^{(3)} | \psi_{k_y\downarrow} \rangle$ into the effective RSOC α_{k_y}. Here $\psi_{k_y\sigma}$ for $\sigma = \downarrow, \uparrow$ are the HESs in Eq. (B5).

We begin by noticing that

$$\langle \psi_{k_y\uparrow} | H_R^{(2)} | \psi_{k_y\downarrow} \rangle = 0, \tag{E2}$$

i.e., the second-order RSOC term $H_R^{(2)}$ does not contribute to the RSOC for a pair of HESs at an isolated edge within our analytical approach. The same kind of cancellation was found by considering the BIA term in Eq. (20). In fact, both cancellations stem from the alternating signs in the antidiagonal of H_{BIA} and $H_R^{(2)}$, respectively. Hence the cancellation is independent of the details of the transversal wave function of the HESs.

Now we turn to the third-order RSOC $H_R^{(3)}$, where we will see how our analytical approach actually fails to give the correct answer. A straightforward calculation gives

$$\begin{aligned} &\langle \psi_{k_y\uparrow} | H_R^{(3)} | \psi_{k_y\downarrow} \rangle \\ &\quad = iT_0 \frac{B_+}{2B} \int_0^\infty dx\, g_{k_y}(x) (k_-)^3 g_{-k_y}(x) \\ &\quad = -T_0 \frac{B_+}{2B} \left[a_{k_y}^{(3)} + 3k_y a_{k_y}^{(2)} + 3k_y^2 a_{k_y}^{(1)} + k_y^3 a_{k_y}^{(0)} \right], \end{aligned} \tag{E3}$$

where $k_- = -i(\partial_x + k_y)$ and we introduced

$$\begin{aligned} a_{k_y}^{(n)} &= \int_0^\infty dx\; g_{k_y}(x) \partial_x^n g_{-k_y}(x) \\ &= h(k_y) \left[(-1)^n (\lambda_1^-)^n \left(\frac{1}{\lambda_1^+ + \lambda_1^-} - \frac{1}{\lambda_1^- + \lambda_2^+} \right) \right. \\ &\quad \left. + (-1)^n (\lambda_2^-)^n \left(\frac{1}{\lambda_2^+ + \lambda_2^-} - \frac{1}{\lambda_1^+ + \lambda_2^-} \right) \right]. \end{aligned} \tag{E4}$$

Here we use $h(k_y)$ from Eq. (C4) and the shorthand notation $\lambda_i^\pm = \lambda_i(\pm k_y)$ for the two inverse length scales in Eq. (B9) that appear in the transversal wave functions of the HESs $g_{\pm k_y}(x)$ in Eq. (B8). We emphasize that even though the

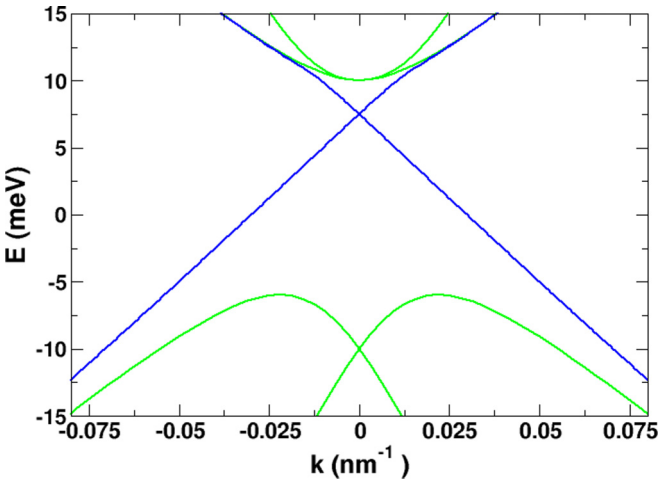

FIG. 11. The edge-state dispersions including both the first- and third-order RSOC terms, $H_R^{(1)}$ and $H_R^{(3)}$, found by a numerical tight-binding calculation. The parameters are the same as in Fig. 2 (top). Thus we observe that the third-order term only introduces minor changes in the edge-state spectrum. The prefactor T_0 of the third-order RSOC term is chosen such that $T_0 = 0.57R_0$ according to the values given by Rothe *et al.* [35].

calculus leading to this result is correct, the result itself is not correct. For instance, it leads to a nonzero matrix element at $k_y = 0$, which in turn introduces a gap in the edge-state spectrum of the Hamiltonian (12). This is obviously not correct since the RSOC is invariant under time reversal symmetry and therefore *no* gap should be opened by $H_R^{(3)}$. To emphasize this point, we have performed a numerical tight-binding calculation including both $H_R^{(1)}$ and $H_R^{(3)}$, which indeed shows that no gap appears in the edge-state spectrum, see Fig. 11. Now, to pinpoint the origin of the problem further, we consider the matrix element (E3) at $k_y = 0$, which erroneously was found to be nonzero. By using integration by parts repeatedly, we obtain

$$\left\langle \psi_{k_y=0\uparrow} \middle| H_R^{(3)} \middle| \psi_{k_y=0\downarrow} \right\rangle = -\frac{B_+}{2B} T_0 \int_0^\infty dx g_0(x) \partial_x^3 g_0(x)$$
$$= -\frac{B_+}{4B} T_0 [\partial_x g_0(x=0)]^2. \quad \text{(E5)}$$

Hence, the matrix element is proportional to the square of the derivative of the transversal wave function of the HES, $[\partial_x g_0(x)]^2$, at the boundary $x = 0$. We found the transverse wave function $g_{k_y}(x)$ analytically using the simple hard-wall boundary condition that $g_{k_y}(x = 0) = 0$ (see Appendix B 2 and Refs. [43,46]). This produces an artificial discontinuity in the derivative of $g_{k_y}(x)$ at the boundary $x = 0$—just as for the textbook example of an infinitely deep square well. It is this discontinuity that gives the incorrect nonzero matrix element at $k_y = 0$. For any smooth boundary potential (or a finite step potential), $\partial_x g_0(x)$ would be zero at the boundary of the integral (not necessarily at $x = 0$) and thereby give the correct result. Unfortunately, it is hard to obtain analytical wave functions for these potentials. In other words, we seem to get a nonzero result due to our crude approximation for the transversal wave function. However, for integrals involving only the first-order derivatives of g_{k_y} as in the main text, we can still use the HESs in Eq. (B5) in our analytical approach.

A similar situation is found in the use of $\mathbf{k} \cdot \mathbf{p}$ theory to describe confined structures by the envelope function approximation. Here, hard-wall boundaries are often used to describe structures in accordance with experimental observations, even though the envelope function approximation, in principle, requires smooth potentials. This has been justified in some cases, but remains a problematic issue for other cases (see Sec. 4.1 in Ref. [65] for a discussion).

[1] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).
[2] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).
[3] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).
[4] Y.Ando, J. Phys. Soc. Jpn. **82**, 102001 (2013).
[5] C. Xu and J. E. Moore, Phys. Rev. B **73**, 045322 (2006).
[6] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang, Science **318**, 766 (2007).
[7] A. Roth, C. Brüne, H. Buhmann, L. W. Molenkamp, J. Maciejko, X.-L. Qi, and S.-C. Zhang, Science **325**, 294 (2009).
[8] C. Brüne, A. Roth, H. Buhmann, E. M. Hankiewicz, L. W. Molenkamp, J. Maciejko, X.-L. Qi, and S.-C. Zhang, Nat. Phys. **8**, 485 (2012).
[9] M. König, M. Baenninger, A. G. F. Garcia, N. Harjee, B. L. Pruitt, C. Ames, P. Leubner, C. Brüne, H. Buhmann, L. W. Molenkamp, and D. Goldhaber-Gordon, Phys. Rev. X **3**, 021003 (2013).
[10] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, Science **314**, 1757 (2006).
[11] C. Liu, T. L. Hughes, X.-L. Qi, K. Wang, and S.-C. Zhang, Phys. Rev. Lett. **100**, 236601 (2008).
[12] I. Knez, R.-R. Du, and G. Sullivan, Phys. Rev. Lett. **107**, 136603 (2011).
[13] K. Suzuki, Y. Harada, K. Onomitsu, and K. Muraki, Phys. Rev. B **87**, 235311 (2013).
[14] I. Knez, C. T. Rettner, S.-H. Yang, S. S. P. Parkin, L. Du, R.-R. Du, and G. Sullivan, Phys. Rev. Lett. **112**, 026602 (2014).
[15] L. Du, I. Knez, G. Sullivan, and R.-R. Du, Phys. Rev. Lett. **114**, 096802 (2015).
[16] F. Nichele, H. J. Suominen, M. Kjaergaard, C. M. Marcus, E. Sajadi, J. A. Folk, F. Qu, A. J. Beukman, F. K. de Vries, J. van Veen, S. Nadj-Perge, L. P. Kouwenhoven, B.-M. Nguyen, A. A. Kiselev, M. S. Wei Yi, M. J. Manfra, E. M. Spanton, and K. A. Moler, arXiv:1511.01728 (2015).
[17] G. M. Gusev, Z. D. Kvon, O. A. Shegai, N. N. Mikhailov, S. A. Dvoretsky, and J. C. Portal, Phys. Rev. B **84**, 121302 (2011).
[18] G. Grabecki, J. Wróbel, M. Czapkiewicz, L. Cywiński, S. Gierałtowska, E. Guziewicz, M. Zholudev, V. Gavrilenko,

N. N. Mikhailov, S. A. Dvoretski, F. Teppe, W. Knap, and T. Dietl, Phys. Rev. B **88**, 165309 (2013).
[19] G. M. Gusev, Z. D. Kvon, E. B. Olshanetsky, A. D. Levin, Y. Krupko, J. C. Portal, N. N. Mikhailov, and S. A. Dvoretsky, Phys. Rev. B **89**, 125305 (2014).
[20] E. M. Spanton, K. C. Nowack, L. Du, G. Sullivan, R.-R. Du, and K. A. Moler, Phys. Rev. Lett. **113**, 026804 (2014).
[21] T. L. Schmidt, S. Rachel, F. von Oppen, and L. I. Glazman, Phys. Rev. Lett. **108**, 156402 (2012).
[22] A. Ström, H. Johannesson, and G. I. Japaridze, Phys. Rev. Lett. **104**, 256804 (2010).
[23] J. C. Budich, F. Dolcini, P. Recher, and B. Trauzettel, Phys. Rev. Lett. **108**, 086602 (2012).
[24] F. Crépin, J. C. Budich, F. Dolcini, P. Recher, and B. Trauzettel, Phys. Rev. B **86**, 121106 (2012).
[25] N. Lezmy, Y. Oreg, and M. Berkooz, Phys. Rev. B **85**, 235304 (2012).
[26] F. Geissler, F. Crépin, and B. Trauzettel, Phys. Rev. B **89**, 235136 (2014).
[27] N. Kainaris, I. V. Gornyi, S. T. Carr, and A. D. Mirlin, Phys. Rev. B **90**, 075118 (2014).
[28] J. I. Väyrynen, M. Goldstein, and L. I. Glazman, Phys. Rev. Lett. **110**, 216402 (2013).
[29] Y. Tanaka, A. Furusaki, and K. A. Matveev, Phys. Rev. Lett. **106**, 236402 (2011).
[30] A. M. Lunde and G. Platero, Phys. Rev. B **86**, 035112 (2012).
[31] E. Eriksson, A. Ström, G. Sharma, and H. Johannesson, Phys. Rev. B **86**, 161103 (2012).
[32] E. Eriksson, Phys. Rev. B **87**, 235414 (2013).
[33] B. Probst, P. Virtanen, and P. Recher, Phys. Rev. B **92**, 045430 (2012).
[34] A. M. Lunde and G. Platero, Phys. Rev. B **88**, 115411 (2013).
[35] 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).
[36] P. Virtanen and P. Recher, Phys. Rev. B **85**, 035310 (2012).
[37] D. G. Rothe and E. M. Hankiewicz, Phys. Rev. B **89**, 035418 (2014).
[38] 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).
[39] P. M. Ostrovsky, I. V. Gornyi, and A. D. Mirlin, Phys. Rev. B **86**, 125323 (2012).
[40] C. P. Orth, G. Strübi, and T. L. Schmidt, Phys. Rev. B **88**, 165315 (2013).
[41] C. P. Orth, R. P. Tiwari, T. Meng, and T. L. Schmidt, Phys. Rev. B **91**, 081406 (2015).
[42] A. Rod, T. L. Schmidt, and S. Rachel, Phys. Rev. B **91**, 245112 (2015).
[43] B. Zhou, H.-Z. Lu, R.-L. Chu, S.-Q. Shen, and Q. Niu, Phys. Rev. Lett. **101**, 246807 (2008).
[44] C. Liu and S.-C. Zhang, *Models and Materials for Topological Insulators*, edited by M. Franz and L. Molenkamp, Topological Insulators Vol. 7 (Springer, Vancouver, 2013).
[45] S.-Q. Shen, *Topological Insulators* (Springer-Verlag, Hong Kong, 2011).
[46] M. Wada, S. Murakami, F. Freimuth, and G. Bihlmayer, Phys. Rev. B **83**, 121310 (2011).
[47] B. A. Bernevig and with T. L. Hughes, *Topological Insulators and Topological Superconductors* (Princeton University Press, New York, 2013).
[48] P. Michetti, P. H. Penteado, J. C. Egues, and P. Recher, Semicond. Sci. Technol. **27**, 124007 (2012).
[49] V. Krueckl and K. Richter, Phys. Rev. Lett. **107**, 086803 (2011).
[50] W.-Y. Shan, H.-Z. Lu, and S.-Q. Shen, New J. Phys. **12**, 043048 (2010).
[51] Y. Takagaki, Phys. Rev. B **90**, 165305 (2014).
[52] A. Medhi and V. B. Shenoy, J. Phys.: Condens. Matter **24**, 355001 (2012).
[53] J. Linder, T. Yokoyama, and A. Sudbø, Phys. Rev. B **80**, 205401 (2009).
[54] C.-X. Liu, H. J. Zhang, B. Yan, X.-L. Qi, T. Frauenheim, X. Dai, Z. Fang, and S.-C. Zhang, Phys. Rev. B **81**, 041307 (2010).
[55] H.-Z. Lu, W.-Y. Shan, W. Yao, Q. Niu, and S.-Q. Shen, Phys. Rev. B **81**, 115407 (2010).
[56] J. J. Sakurai, *Modern Quantum Mechanics* (Addison Wesley, Geneva, 1993).
[57] S. Murakami, S. Iso, Y. Avishai, M. Onoda, and N. Nagaosa, Phys. Rev. B **76**, 205304 (2007).
[58] M. Berry and R. J. Mondragon, Proc. R. Soc. London A **412**, 53 (1987).
[59] J. A. Fürst, J. G. Pedersen, C. Flindt, N. A. Mortensen, M. Brandbyge, T. G. Pedersen, and A.-P. Jauho, New J. Phys. **11**, 095020 (2009).
[60] P. Michetti and P. Recher, Phys. Rev. B **83**, 125420 (2011).
[61] Y. Baum, T. Posske, I. C. Fulga, B. Trauzettel, and A. Stern, Phys. Rev. Lett. **114**, 136801 (2015).
[62] K.-I. Imura, A. Yamakage, S. Mao, A. Hotta, and Y. Kuramoto, Phys. Rev. B **82**, 085118 (2010).
[63] H. Nielsen and M. Ninomiya, Nucl. Phys. B **185**, 20 (1981).
[64] H. Nielsen and M. Ninomiya, Nucl. Phys. B **193**, 173 (1981).
[65] R. Winkler, *Spin-orbit Coupling Effects in Two-Dimensional Electron and Hole Systems*, Springer Tracts in Modern Physics (Springer, Aarhus, 2003).