POLITECNICO DI TORINO
Repository ISTITUZIONALE

Tunnel junction of helical edge states: Determining and controlling spin-preserving and spin-flipping processes through transconductance

Original
Tunnel junction of helical edge states: Determining and controlling spin-preserving and spin-flipping processes through transconductance / Sternativo, P., Dolcini, F.. - In: PHYSICAL REVIEW. B, CONDENSED MATTER AND MATERIALS PHYSICS. - ISSN 1098-0121. - STAMPA. - 89:3(2014), pp. 035415-1-035415-14. [10.1103/PhysRevB.89.035415]

Availability:
This version is available at: 11583/2531113 since:

Publisher:
American Physical Society (APS)

Published
DOI:10.1103/PhysRevB.89.035415

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

25 November 2026

Tunnel junction of helical edge states: Determining and controlling spin-preserving and spin-flipping processes through transconductance

Pietro Sternativo[1] and Fabrizio Dolcini[1,2,*]

[1]*Dipartimento di Scienza Applicata e Tecnologia del Politecnico di Torino, I-10129 Torino, Italy*

[2]*CNR-SPIN, Monte S. Angelo, via Cinthia, I-80126 Napoli, Italy*

(Received 9 August 2013; revised manuscript received 29 November 2013; published 14 January 2014)

When a constriction is realized in a 2D quantum spin Hall system, electron tunneling between helical edge states occurs via two types of channels allowed by time-reversal symmetry, namely spin-preserving (p) and spin-flipping (f) tunneling processes. Determining and controlling the effects of these two channels is crucial to the application of helical edge states in spintronics. We show that, despite that the Hamiltonian terms describing these two processes do not commute, the scattering matrix entries of the related 4-terminal setup always factorize into products of p-term and f-term contributions. Such factorization provides an operative way to determine the transmission coefficients T_p and T_f related to each of the two processes, via transconductance measurements. Furthermore, these transmission coefficients are also found to be controlled independently by a suitable combination of two gate voltages applied across the junction. This result holds for an arbitrary profile of the tunneling amplitudes, including disorder in the tunnel region, enabling us to discuss the effect of the finite length of the tunnel junction, and the space modulation of both magnitude and phase of the tunneling amplitudes.

DOI: 10.1103/PhysRevB.89.035415 PACS number(s): 73.23.−b, 72.10.−d, 73.43.Jn

I. INTRODUCTION

The discovery of topological materials [1–5] has unveiled the existence of helical edge states, i.e., linearly dispersed gapless one-dimensional modes flowing at the boundaries of the insulating bulk of a two-dimensional quantum spin Hall effect system [1–3,6,7]. Helical states are characterized by a locking of the electron group velocity to the spin orientation, so that the two counterpropagating modes flowing at a given edge exhibit opposite spin orientation. The most straightforward way to reveal the peculiar features of helical edge states is through their transport properties, in particular when the helical states of opposite edges are coupled in a tunnel junction, realized, e.g., by etching a constriction in HgTe/CdTe [2,3] or in InAs/GaSb [7] quantum wells. In such situation, time-reversal symmetry implies that two types of tunnel couplings between helical edge states exist [6,8]. The first type (p) preserves the electron spin and changes the group velocity, similarly to a backscattering term also present in conventional quantum wires. The second type (f) instead induces spin flipping by preserving the group velocity. One may expect that the coupling constant of the spin-flipping processes is smaller than that of the spin-preserving ones. However, no operative way has been conceived so far to quantitatively extract these coupling constants. In fact, the possibility of exploiting both processes is crucial for the intriguing perspective of utilizing topological edge states in spintronics [9–12], where the spin degree of freedom is used to encode information. Indeed in a spintronics nanodevice currents should be switched on-demand from one terminal to another, in a controlled way, with either preservation or flipping of the spin, depending on the specific operation to be performed.

Determining and controlling the transmission coefficients related to these two tunneling channels appears to be a challenging task, for various reasons. In the first instance the Hamiltonian terms describing the two processes do not commute, making the analysis of their interplay *a priori* nontrivial. Second, due to etching and to indirect coupling via the bulk states, a tunnel junction is a typically irregular and disordered region, implying that the tunneling amplitude of each of the two channels cannot be described by one single parameter but rather by a space-dependent profile. Furthermore, in a tunnel junction realized in a quantum well of materials such as HgTe/CdTe, where the strong spin-orbit interaction correlates momentum and spin [2,13], the breaking of the longitudinal transversal invariance originating from the finite length of the tunnel region affects spin preservation, thereby possibly modifying the relative weight of f-processes with respect to p-processes.

Most of the theoretical approaches to this problem have treated the tunnel junction as a pointlike constriction, using a δ-tunneling (DT) model, and have focused on the effect of electron-electron interaction on conductance [8,14–18] and noise [19], and on interference phenomena between two quantum point contacts [20–25]. However, while there is no clear experimental evidence that electron-electron interaction plays a significant role in helical edge state transport, it is worth noticing that the DT model, per se, completely neglects the internal structure of the junction, which may contain important physical insights in realistic implementations. Some recent works, for instance, have pointed out that the finite length of the junction plays an important role when the related Thouless energy becomes comparable to the applied voltages [9,15,17,23]. These studies were limited to specific profiles for the tunneling amplitudes and/or to the case of spin-preserving tunneling, though.

For these reasons, the question of how to operatively determine and control the spin-preserving and spin-flipping effects in a realistic tunnel junction in the presence of a disordered profile is still open. In this paper we address this problem. Focusing on the regime where electron-electron interaction is negligible, we consider a 4-terminal setup,

*fabrizio.dolcini@polito.it

 ©2014 American Physical Society

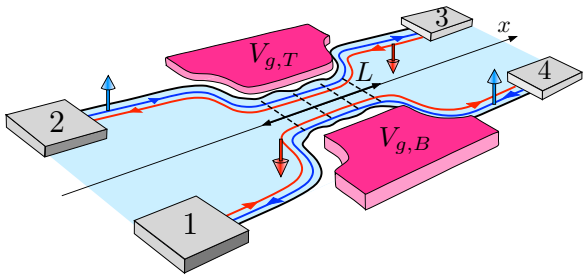

FIG. 1. (Color online) A tunnel junction coupling the helical edge states can be realized by etching a constriction in a four-terminal quantum spin Hall effect setup. Two side gates enable one to tune the potential along the junction. Two types of time-reversal symmetric tunnel processes occur at the junction, namely spin-preserving and spin-flipping processes. Determining and controlling these two processes is crucial for exploiting helical edge states in spintronics devices.

schematically depicted in Fig. 1, where helical edge states are coupled in a tunnel region characterized by a finite length and by an arbitrary profile for the tunneling amplitude of both p- and f-tunneling processes. We show that, despite that the Hamiltonian terms describing the two types of tunneling processes do not commute, the scattering matrix entries of the 4-terminal setup always factorize into two terms that depend on spin-preserving terms only and on spin-flipping terms only. This factorization provides an operative way to determine the magnitude of the transmission coefficients related to each of these terms, via transport measurements. In particular, the spin-flipping terms, although possibly quantitatively smaller in magnitude than the spin-preserving terms, induce qualitatively different features in the conductance matrix, which cannot be ascribed to p-processes. Furthermore we predict that by suitably operating with side gates, one can control independently the two transmission coefficients related to spin-preserving and spin-flipping tunneling. Importantly, the factorization property holds for an *arbitrary* profile of the tunneling amplitudes, possibly including disorder and local fluctuations effects. In particular, it also holds in the limit of short constriction. By comparing our results with the widely used DT model, we shall show that such factorization property, which is seemingly absent in the DT model, is simply hidden in a misleading parametrization of the coupling constants of that model.

The paper is organized as follows: In Sec. II, after presenting the model for the tunnel junction, we prove the factorization property of the scattering matrix for an arbitrary profile for the junction parameters. In Sec. III we discuss how such property impacts the multiterminal conductance, and show how the transmission coefficients related to the two types of tunneling can be operatively determined and controlled by gate voltages. In Sec. IV we present some explicit results for specific profiles of the junction tunneling amplitude, which enable us to discuss the role of p- and f-tunneling, the effects of the finite length of the tunnel junction, and the variation of the magnitude and the phase of the tunneling amplitude profile. We also explain why the DT model hides the factorization property. Finally, in Sec. V we discuss our results and draw some conclusions.

II. MODEL

We consider a 4-terminal setup of helical edge states, as sketched in Fig. 1. Along the top edge of the quantum well right movers are characterized by spin $\uparrow$ and left movers by spin $\downarrow$, whereas along the bottom edge the opposite spin orientations occur. A constriction is assumed to be realized in the quantum well, allowing electron tunneling between the four edge states over a region extending along the longitudinal direction x. Furthermore, we consider two gate electrodes, applied at the sides of the constriction, enabling us to shift the chemical potential of the edge states.

We describe the electron edge states through four electron field operators $\Psi_{R\uparrow}(x),\Psi_{L\uparrow}(x),\Psi_{R\downarrow}(x),\Psi_{L\downarrow}(x)$, with $\alpha = R/L = \pm$ denoting the chirality for right movers and left movers, respectively, and $\sigma = \uparrow,\downarrow$ the spin component. We then model the setup by the following low-energy Hamiltonian:

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_0 + \hat{\mathcal{H}}_{g,T} + \hat{\mathcal{H}}_{g,B} + \hat{\mathcal{H}}^p_{\rm tun} + \hat{\mathcal{H}}^f_{\rm tun}\,, \tag{1}$$

where

$$\hat{\mathcal{H}}_0 = -i\hbar v_{\rm F} \sum_{\alpha=R/L=\pm} \alpha \sum_{\sigma=\uparrow,\downarrow} \int dx \; :\Psi^\dagger_{\alpha\sigma}(x)\,\partial_x \Psi_{\alpha\sigma}(x): \tag{2}$$

describes the linear bands of the helical edge states [1,2] (: : denotes normal ordering), and

$$\hat{\mathcal{H}}_{g,T} = \int dx\, eV_{g,T}(x)[\hat{\rho}_{R\uparrow}(x) + \hat{\rho}_{L\downarrow}(x)], \tag{3}$$

$$\hat{\mathcal{H}}_{g,B} = \int dx\, eV_{g,B}(x)[\hat{\rho}_{R\downarrow}(x) + \hat{\rho}_{L\uparrow}(x)] \tag{4}$$

account for the the electric potentials applied by the side gates across the constriction. Here

$$\hat{\rho}_{\alpha\sigma}(x) =: \Psi^\dagger_{\alpha\sigma}(x)\Psi_{\alpha\sigma}(x):\tag{5}$$

is the electron chiral density. Finally, the last two terms in Eq. (1),

$$\hat{\mathcal{H}}^p_{\rm tun} = \sum_{\sigma=\uparrow,\downarrow} \int dx (\Gamma_p(x)\,\Psi^\dagger_{L\sigma}(x)\,\Psi_{R\sigma}(x) + \text{H.c.}), \tag{6}$$

$$\hat{\mathcal{H}}^f_{\rm tun} = \sum_{\alpha=R/L=\pm} \alpha \int dx\,(\Gamma_f(x)\;\Psi^\dagger_{\alpha\downarrow}(x)\,\Psi_{\alpha\uparrow}(x) + \text{H.c.}), \tag{7}$$

describe the two types of interedge tunnel coupling allowed by time-reversal symmetry at the constriction, namely the spin-preserving and the spin-flipping tunneling, respectively [6,8,9,15,20,21,23,26].

The tunneling amplitudes $\Gamma_p(x)\,,\Gamma_f(x)$ and the side gate potentials $V_{g,T}(x),V_{g,B}(x)$ characterizing the constriction region are allowed to vary along the longitudinal direction x. Their profiles can be arbitrary, with the only constraint that—sufficiently far away from the central region—they all vanish and the helical states are described by the linear dispersion band term (2) only. We can thus assume, without loss of generality, that there exist two coordinates x_1 and x_2 defining the extremal longitudinal boundaries of the constriction, such

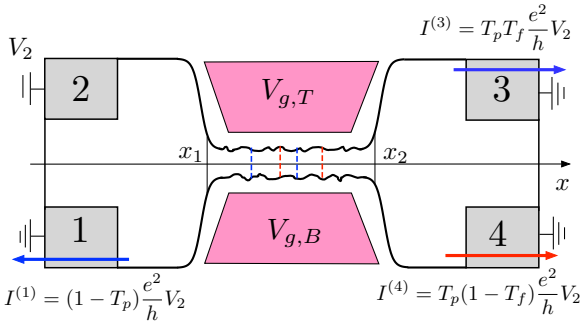

FIG. 2. (Color online) The operative way to extract the transmission coefficients T_p and T_f related to spin-preserving and spin-flipping tunneling, respectively. When a voltage bias V_2 is applied to terminal 2, the currents flowing in the other three terminals directly depend on T_p and T_f in a factorized form, so that T_p and T_f can be determined via Eqs. (46) and (47). Furthermore, T_p is controlled by the charge gate $V_{gc} = (V_{g,T} + V_{g,B})/2$, whereas T_f is controlled by the spin gate $V_{gs} = (V_{g,T} - V_{g,B})/2$.

that

$$\Gamma_p(x), \Gamma_f(x), V_{g,T}(x), V_{g,B}(x) = 0 \quad \text{for} \quad \begin{cases} x \leqslant x_1, \\ x \geqslant x_2, \end{cases} \tag{8}$$

with $L = x_2 - x_1$ denoting the tunnel junction length. The tunnel junction $x_1 \leqslant x \leqslant x_2$ can thus be regarded as the scattering region in the 4-terminal setup in Fig. 1, where the distributions of the incoming electrons are controlled by the chemical potentials μ_i $(i = 1, \ldots 4)$, and by the temperature $k_B T$ of the four reservoirs (see also Fig. 2).

It is worth emphasizing that the two types of tunneling terms (6) and (7) acting in the constriction do *not* commute,

$$\left[\hat{\mathcal{H}}^p_{\text{tun}}, \hat{\mathcal{H}}^f_{\text{tun}}\right] = 2 \int dx [\Gamma_p(x)\Gamma_f(x)\Psi^\dagger_{L\downarrow}(x)\Psi_{R\uparrow}(x) + \\ + \Gamma_p(x)\Gamma^*_f(x)\Psi^\dagger_{L\uparrow}(x)\Psi_{R\downarrow}(x) - \text{H.c.}] \neq 0. \tag{9}$$

In view of Eq. (9), the terms $\hat{\mathcal{H}}^p_{\text{tun}}$ and $\hat{\mathcal{H}}^f_{\text{tun}}$ are expected to interplay in transport measurements, so that singling out the effect of each of the two tunnel couplings looks quite difficult. This expectation seems to be confirmed by transport predictions based on the DT model, which treats the tunnel junction as a pointlike tunneling region [8,14–25]. Indeed by adopting such model the transmission coefficients are found to depend on both tunneling amplitudes in a nonfactorized way [20,21].

We shall show that such seemingly complicated dependence is an artifact of the DT model. Indeed we prove that despite the noncommutativity $\hat{\mathcal{H}}^p_{\text{tun}}$ and $\hat{\mathcal{H}}^f_{\text{tun}}$, each entry of the scattering matrix of the setup can be factorized, one related the spin-preserving and the other one to the spin-flipping tunneling. A comparison of our results with the ones of the DT model will be explicitly made in Sec. IV E.

A. Factorization of the scattering matrix entries

In order to prove the factorization of the scattering matrix entries, we first introduce a four-component electron field operator $\Psi(x) = (\Psi_{R\uparrow}(x), \Psi_{L\uparrow}(x), \Psi_{R\downarrow}(x), \Psi_{L\downarrow}(x))^T$, as well as Pauli matrices $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ acting on the spin space $(\sigma = \uparrow, \downarrow)$ and Pauli matrices $\boldsymbol{\tau} = (\tau_x, \tau_y, \tau_z)$ acting on the chirality space $(\alpha = R, L)$. Furthermore we define charge and spin gate voltages as sum and difference of the side gate voltages appearing in Eqs. (3) and (4):

$$V_{g\,c/s}(x) = [V_{g,T}(x) \pm V_{g,B}(x)]/2. \tag{10}$$

The Hamiltonian (1) is then compactly written as

$$\hat{\mathcal{H}} = \int dx \Psi^\dagger(x) \big[H_0(x) + H_{gc}(x) + H_{gs}(x) \\ + H^p_{\text{tun}}(x) + H^f_{\text{tun}}(x)\big]\Psi(x), \tag{11}$$

where the 4×4 matrices in Eq. (11) read

$$H_0(x) = -i\hbar v_F (\sigma_0 \otimes \tau_z)\, \partial_x, \tag{12}$$

$$H_{gc}(x) = eV_{gc}(x)\,(\sigma_0 \otimes \tau_0)\,, \tag{13}$$

$$H_{gs}(x) = eV_{gs}(x)\,(\sigma_z \otimes \tau_z)\,, \tag{14}$$

$$H^p_{\text{tun}}(x) = |\Gamma_p(x)|\sigma_0 \otimes [\tau_x \cos\phi_p(x) + \tau_y \sin\phi_p(x)], \tag{15}$$

$$H^f_{\text{tun}}(x) = |\Gamma_f(x)|[\sigma_x \cos\phi_f(x) + \sigma_y \sin\phi_f(x)] \otimes \tau_z, \tag{16}$$

with σ_0 and τ_0 denoting the 2×2 identity matrices in spin and chirality space, respectively.

The equation of motion $i\hbar\,\partial_t \Psi(x,t) = [\Psi(x,t), \hat{\mathcal{H}}]$ obtained from the Hamiltonian (11) implies for the stationary solutions $\Psi(x,t) = e^{-iEt/\hbar}\Psi_E(x)$ that

$$i(\sigma_0 \otimes \tau_0)\frac{\partial}{\partial x}\Psi_E(x) \\ = \{[\boldsymbol{\sigma} \cdot \mathbf{b}_f(x)] \otimes \tau_0 + \sigma_0 \otimes [\boldsymbol{\tau} \cdot \mathbf{b}_{p,E}(x)]\}\Psi_E(x), \tag{17}$$

where

$$\mathbf{b}_f(x) = (\text{Re}\Gamma_f(x), \text{Im}\Gamma_f(x), eV_{gs}(x))/\hbar v_F \tag{18}$$

is a (real) vector field that depends on the spin-flipping tunneling amplitude Γ_f and the spin gate voltage V_{gs}, whereas

$$\mathbf{b}_{p,E}(x) = (-i\,\text{Im}\Gamma_p(x)\,,\, i\,\text{Re}\Gamma_p(x)\,,\, eV_{gc}(x) - E)/\hbar v_F \tag{19}$$

is a (complex) vector field that depends on the spin-preserving tunneling amplitude Γ_p, the charge gate V_{gc}, and the energy E, defined with respect to the Dirac point level. Notice that Eq. (17) is formally equivalent to the "evolution" equation of a particle endowed with a twofold spin degree of freedom, exposed to two "time-dependent" magnetic fields $\mathbf{b}_f$ and $\mathbf{b}_{p,E}$. The role of time is played by space x, and the two magnetic fields originate from spin-preserving and spin-flipping tunneling. Outside the central scattering region, where Eq. (8) holds, the dynamics is governed by the term (12) only, and the field operator Ψ_E solving Eq. (17) acquires the simple asymptotic form

$$\Psi_E(x \leqslant x_1) = (\sigma_0 \otimes e^{i\tau_z k_E x})(a_{R\uparrow}, b_{L\uparrow}, a_{R\downarrow}, b_{L\downarrow})^T \tag{20}$$

and

$$\Psi_E(x \geqslant x_2) = (\sigma_0 \otimes e^{i\tau_z k_E x})(b_{R\uparrow}, a_{L\uparrow}, b_{R\downarrow}, a_{L\downarrow})^T, \tag{21}$$

where $a_{\alpha\sigma}$ and $b_{\alpha\sigma}$ denote operators for incoming and outgoing states, respectively, and $k_E = E/\hbar v_F$. The transfer matrix M,

connecting operators on the right of the central scattering region to the ones on the left,

$$\begin{pmatrix} b_{R\uparrow} \\ a_{L\uparrow} \\ b_{R\downarrow} \\ a_{L\downarrow} \end{pmatrix} = \mathsf{M} \begin{pmatrix} a_{R\uparrow} \\ b_{L\uparrow} \\ a_{R\downarrow} \\ b_{L\downarrow} \end{pmatrix}, \tag{22}$$

can therefore be evaluated through the relation

$$\Psi_E(x_2) = (\sigma_0 \otimes e^{i\tau_z k_E x_2})\, \mathsf{M}\, (\sigma_0 \otimes e^{-i\tau_z k_E x_1})\, \Psi_E(x_1). \tag{23}$$

In order to determine the solution of the stationary Eq. (17), we introduce the following space "evolution" operators

$$\mathbf{U}_f(x;0) = \mathrm{T}\, e^{-i\int_0^x dx' \boldsymbol{\sigma}\cdot\mathbf{b}_f(x')}, \tag{24}$$

$$\mathbf{U}_{p,E}(x;0) = \mathrm{T}\, e^{-i\int_0^x dx' \boldsymbol{\tau}\cdot\mathbf{b}_{p,E}(x')} \tag{25}$$

that appear as two continuous sets of rotations, around the local "magnetic" field $\mathbf{b}_f$ (spin space) and "pseudomagnetic" field $\mathbf{b}_{p,E}$ (chirality space) determined by the tunnel junction parameter profiles [see Eqs. (18) and (19), respectively]. Here T denotes the "time" (actually space) ordering operator, and the "evolution" is with respect to the space origin $x = 0$. Then Eqs. (24) and (25) lead to

$$i\,\partial_x \mathbf{U}_f(x;0) = [\boldsymbol{\sigma}\cdot\mathbf{b}_f(x)]\,\mathbf{U}_f(x;0), \tag{26}$$

$$i\,\partial_x \mathbf{U}_{p,E}(x;0) = [\boldsymbol{\tau}\cdot\mathbf{b}_{p,E}(x)]\,\mathbf{U}_{p,E}(x;0). \tag{27}$$

Using Eqs. (26) and (27), the solution of Eq. (17) is straightforwardly verified to be

$$\Psi_E(x) = [\mathbf{U}_f(x;0)\otimes\mathbf{U}_{p,E}(x;0)]\,\Psi(0), \tag{28}$$

where $\Psi(0)$ is a four-component field operator at the space origin. The solution (28) then implies that

$$\Psi_E(x_2) = [\mathbf{U}_f(x_2;x_1)\otimes\mathbf{U}_{p,E}(x_2;x_1)]\Psi_E(x_1). \tag{29}$$

Comparing Eqs. (23) and (29) one obtains

$$\mathsf{M} = \mathbf{m}_f \otimes \mathbf{m}_p, \tag{30}$$

where

$$\mathbf{m}_f \doteq \mathrm{T}\, e^{-i\int_{x_1}^{x_2} dx' \boldsymbol{\sigma}\cdot\mathbf{b}_f(x')}, \tag{31}$$

$$\mathbf{m}_p \doteq e^{-i\tau_z k_E x_2}(\mathrm{T}\, e^{-i\int_{x_1}^{x_2} dx' \boldsymbol{\tau}\cdot\mathbf{b}_{p,E}(x')})e^{+i\tau_z k_E x_1}. \tag{32}$$

We notice that, because $\mathbf{b}_f(x)\in\mathbb{R}$, Eq. (31) implies that $\mathbf{m}_f \in \mathrm{SU}(2)$. In contrast, because $\mathbf{b}_{p,E}(x)\in\mathbb{C}$, $\mathbf{m}_p \notin \mathrm{SU}(2)$. However, $\det\mathbf{m}_p = +1$, since $\boldsymbol{\tau}\cdot\mathbf{b}_{p,E}$ is traceless. Equation (30) shows that the transfer matrix is the tensor product of a 2×2 matrix $\mathbf{m}_f$ acting on spin space by a 2×2 matrix $\mathbf{m}_p$ acting on chirality space. This property reflects on the structure of the scattering matrix S, which expresses outgoing operators in terms of incoming operators [27],

$$\begin{pmatrix} b_{L\uparrow} \\ b_{L\downarrow} \\ b_{R\uparrow} \\ b_{R\downarrow} \end{pmatrix} = \mathsf{S} \begin{pmatrix} a_{R\downarrow} \\ a_{R\uparrow} \\ a_{L\downarrow} \\ a_{L\uparrow} \end{pmatrix}, \tag{33}$$

and which can straightforwardly be obtained from the relations (22). Exploiting Eq. (30) and the properties $\det\mathbf{m}_p = \det\mathbf{m}_f = +1$, one obtains

$$\mathsf{S} = \begin{pmatrix} 0 & r_p & t_p r_f^* & t_p t_f^* \\ r_p & 0 & t_p t_f & -t_p r_f \\ -t_p r_f^* & t_p t_f & 0 & r_p' \\ t_p t_f^* & t_p r_f & r_p' & 0 \end{pmatrix}. \tag{34}$$

In Eq. (34) the quantities

$$t_p = \frac{1}{(\mathbf{m}_p)_{22}}, \qquad r_p = -\frac{(\mathbf{m}_p)_{21}}{(\mathbf{m}_p)_{22}} = -\frac{(\mathbf{m}_p)_{12}^*}{(\mathbf{m}_p)_{22}}, \tag{35}$$

and $r_p' = -r_p^* t_p/t_p^*$ are determined by $\mathbf{m}_p$. They depend on the spin-preserving tunneling amplitude Γ_p and on the charge gate voltage V_{gc} only, besides the energy E [see Eqs. (32) and (19)]. In contrast, in Eq. (34) the quantities

$$t_f = (\mathbf{m}_f)_{11} = (\mathbf{m}_f)_{22}^*, \qquad r_f = (\mathbf{m}_f)_{21} = -(\mathbf{m}_f)_{12}^* \tag{36}$$

are determined by $\mathbf{m}_f$, and depend on the spin-flipping tunneling amplitude Γ_f and on the spin gate voltage V_{gs} only [see Eqs. (31) and (18)].

Equation (34) shows that, in a tunnel junction of helical edge states, the entries of the scattering matrix always factorize into products of two reflection and/or transmission amplitudes, one related to p-tunneling processes and the other one to f-tunneling processes. Such result holds for an *arbitrary* profile of the tunneling amplitudes $\Gamma_p(x)$ and $\Gamma_f(x)$ of spin-preserving and spin-flipping properties. Furthermore, $V_{gc}(x)$ and $V_{gs}(x)$ can be arbitrary too.

In next section we shall discuss how this factorization property enables one to operatively determine, through transport properties, the transmission coefficients

$$T_p = |t_p|^2 = \frac{1}{|(\mathbf{m}_p)_{22}|^2}, \tag{37}$$

$$T_f = |t_f|^2 = |(\mathbf{m}_f)_{22}|^2 \tag{38}$$

related to spin-preserving and spin-flipping tunneling, respectively.

The scattering matrix approach utilized here is nonperturbative, and it naturally accounts for the tunneling amplitudes $\Gamma_\nu(x)$ $(\nu = p, f)$ to all perturbative orders. However, it is maybe worth clarifying the origin of the factorization in terms of perturbative arguments as well. Neglecting for simplicity the charge and spin gates, one would derive the scattering amplitude of each multiterminal transport process by linear combinations of average values of $\hat{\rho}_{\alpha\sigma}$, performed over the time Keldysh contour K [28],

$$\langle\hat{\rho}_{\alpha\sigma}\rangle = \left\langle \hat{\rho}_{\alpha\sigma}\, e^{-\frac{i}{\hbar}\int_K dt \Psi_i^\dagger (H_{\mathrm{tun}}^p + H_{\mathrm{tun}}^f)_{ij}\Psi_j}\right\rangle_0, \tag{39}$$

where $\hat{\rho}_{\alpha\sigma}$ is the electron chiral density (5), Ψ_i is the ith component of the four-component electron field $\Psi(x)$ defined

at the beginning of Sec. II A, $H^p_{\rm tun}$, $H^f_{\rm tun}$ are the 4×4 matrices (15) and (16), and $\langle\cdots\rangle_0$ denotes the Keldysh average over H_0 [see Eq. (12)]. For simplicity of notation we have omitted space integration and space-time arguments, and we have assumed implicit summation over repeated indices. We now notice from Eqs. (15) and (16) that the lack of commutation between p- and f-processes,

$$\left[H^p_{\rm tun}\,,\,H^f_{\rm tun}\right]\neq 0, \tag{40}$$

is due to the appearance of the τ_z matrix in Eq. (16), which is necessary to ensure time-reversal symmetry of $\mathcal{H}^f_{\rm tun}$ though. Expanding perturbatively the right-hand side of Eq. (39) in powers of the tunneling amplitudes Γ_p and Γ_f, a given perturbative order is characterized by a power N_p for $H^p_{\rm tun}$ and by a power N_f for $H^f_{\rm tun}$. Because $H^p_{\rm tun}$ involves the same spin and opposite chirality, while $H^f_{\rm tun}$ involves the same chirality and opposite spin, one can realize that the only nonvanishing contributions to $\langle\hat\rho_{\alpha\sigma}\rangle$ occur when the integers N_p and N_f are both even ($N_p=2n_p$ and $N_f=2n_f$). Importantly, despite Eq. (40), one has

$$\left[\left(H^p_{\rm tun}\right)^{2n_p}\,,\left(H^f_{\rm tun}\right)^{2n_f}\right]=0. \tag{41}$$

Effectively, order by order, each nonvanishing contribution to $\langle\hat\rho_{\alpha\sigma}\rangle$ obtained from $H^p_{\rm tun}$ and $H^f_{\rm tun}$ is equal (up to a sign that counts the number of exchanges between $H^p_{\rm tun}$ and $H^f_{\rm tun}$) to the contribution one would obtain by replacing $\tau_z\to\tau_0$ in Eq. (16), i.e., by replacing $H^f_{\rm tun}$ with a matrix that commutes with $H^p_{\rm tun}$. One thus obtains factorized expressions for the nonvanishing scattering matrix entries.

We conclude this section by noticing that the scattering matrix (34) is *not* symmetric, despite that the Hamiltonian of the system is time-reversal invariant. The customary expression of Onsager relations [29] characterizing the scattering matrix entries, $S_{ij}(\mathbf{B})=S_{ji}(-\mathbf{B})$ with $\mathbf{B}$ denoting the external magnetic field, would imply that S is symmetric in the absence of a magnetic field. This is indeed the case for systems where spin is a good quantum number, which thus appears as a mere degeneracy variable. However, the spin-flipping process can occur even without breaking of time-reversal symmetry, as is the case for f-processes for helical edge states in a tunnel junction. In this case, time-reversal transformation involves a $i\sigma_y$ matrix acting on the spin sector, i.e., a sign change whenever spin-$\downarrow$ is flipped to a spin-$\uparrow$. In this case Onsager relations acquire different expressions. In particular, any entry of the scattering matrix that describes a process involving an odd number of spin flips naturally carries an additional minus sign. This is the reason for the appearance of both symmetric and antisymmetric terms in the scattering matrix (34).

III. EFFECTS OF THE FACTORIZATION ON TRANSCONDUCTANCE

We shall now present the results about transport through the setup, which are a direct consequence of the factorization (34) of the scattering matrix entries. The current operators related to the four terminals (denoted by $i=1\ldots 4$ as in Fig. 1) are defined as

$$\begin{aligned}
\hat I^{(1)}_c(x,t)&=ev_F[\hat\rho_{R\downarrow}(x,t)-\hat\rho_{L\uparrow}(x,t)],\\
\hat I^{(2)}_c(x,t)&=ev_F[\hat\rho_{R\uparrow}(x,t)-\hat\rho_{L\downarrow}(x,t)],\\
\hat I^{(3)}_c(x,t)&=ev_F[\hat\rho_{L\downarrow}(x,t)-\hat\rho_{R\uparrow}(x,t)],\\
\hat I^{(4)}_c(x,t)&=ev_F[\hat\rho_{L\uparrow}(x,t)-\hat\rho_{R\downarrow}(x,t)],
\end{aligned} \tag{42}$$

and can then be evaluated by substituting the solution (28) for a given tunnel junction into Eqs. (5) and (42), and integrating over energy E. Notice that, in defining the currents in Eq. (42), we have chosen the customary convention of multiterminal setups that the current in a terminal is considered to be positive when it is incoming from that terminal to the scattering region [27]. We recall that in multiterminal transport the average currents in the steady state are given by

$$I^{(i)}\doteq\langle\hat I^{(i)}(x,t)\rangle=\frac{1}{e}\sum_j\int_{-\infty}^{+\infty}dE\,{\rm G}_{ij}\,f_j, \tag{43}$$

where $f_i=f_i(E)=\{1+\exp[(E-\mu_i)/k_BT]\}^{-1}$ denotes the Fermi function of the ith reservoir, characterized by a temperature k_BT and a chemical potential μ_i, measured with respect to the equilibrium level E_F. In Eq. (43), ${\rm G}_{ij}$ denotes the entry of the conductance matrix G, and describes the current flowing through terminal i as a consequence of a unit voltage bias applied to terminal j. The conductance matrix entries can be expressed in terms of the scattering matrix as ${\rm G}_{ij}=(e^2/h)(\delta_{ij}-|S_{ij}|^2)$. The factorized form (34) acquired by the scattering matrix implies that the conductance matrix G reads

$${\rm G}=\frac{e^2}{h}\begin{pmatrix}
1 & -R_p & -T_pR_f & -T_pT_f\\
-R_p & 1 & -T_pT_f & -T_pR_f\\
-T_pR_f & -T_pT_f & 1 & -R_p\\
-T_pT_f & -T_pR_f & -R_p & 1
\end{pmatrix}. \tag{44}$$

At low temperatures the conductance matrix entry ${\rm G}_{ij}$ can be gained by measuring the current flowing in terminal i in response to a voltage bias applied to only terminal j (transconductance); i.e.,

$${\rm G}_{ij}=\left.\frac{\partial I^{(i)}}{\partial V_j}\right|_{V_{l\neq j}=0}. \tag{45}$$

This provides and operative way to extract the transmission coefficients T_p and T_f through Eq. (44). This is for instance carried out by applying a voltage bias V_2 to terminal 2, while keeping all other chemical potentials to the equilibrium value, as illustrated in Fig. 2. Then, the spin-preserving transmission coefficient T_p is obtained as

$$T_p=1-\frac{h}{e^2}|{\rm G}_{12}|=1-\frac{h}{e^2}\left|\frac{\partial I^{(1)}}{\partial V_2}\right|\Bigg|_{V_{l\neq 2}=0}, \tag{46}$$

whereas and spin-flipping transmission coefficient T_f is

$$T_f=\frac{|{\rm G}_{32}|}{\frac{e^2}{h}-|{\rm G}_{12}|}=\left.\frac{\left|\frac{\partial I^{(3)}}{\partial V_2}\right|}{\frac{e^2}{h}-\left|\frac{\partial I^{(1)}}{\partial V_2}\right|}\right|_{V_{l\neq 2}=0}. \tag{47}$$

Similar equivalent methods, based on biasing another terminal and measuring currents in other appropriate terminals, can be read off from the structure of the G matrix (44), and can be used as cross checks for T_p and T_f. Importantly, because T_p and T_f exhibit different dependencies

$$T_p = T_p(\Gamma_p; E_F - eV_{gc}), \tag{48}$$

$$T_f = T_f(\Gamma_f; eV_{gs}), \tag{49}$$

they are controlled independently by the charge gate voltage $V_{gc} = (V_{g,T} + V_{g,B})/2$ and the spin gate voltage $V_{gs} = (V_{g,T} - V_{g,B})/2$, respectively.

We conclude this section with a remark concerning the operative method described in Fig. 2. When a voltage bias is applied to terminal 2, the helical nature of the edge states implies that no current could be found in terminal 4 if only spin-preserving tunneling occurred in the junction. Thus, the very observation of a current in terminal 4 is a signature of the presence of spin-flipping processes. However, because f-processes interplay with p-processes too, the actual value of such current depends *also* on the latter, in an *a priori* nontrivial way. It is because of the factorization property proved here that such current appears to be simply proportional to $T_p(1 - T_f)$, thereby enabling one to extract the transmission coefficient T_f through Eq. (47).

IV. EXPLICIT RESULTS FOR SPECIAL CASES

The factorization property (34) holds for an arbitrary profile of the tunnel junction parameters. Studying the behavior of T_p and T_f with varying in all possible ways the profiles $\Gamma_p(x), \Gamma_f(x)$ of the tunneling amplitudes and of the potentials $V_{gc}(x), V_{gs}(x)$ deserves a detailed analysis that goes beyond the purpose of the present paper. Nevertheless, in order to show the potential of the result found above, in this section we explicitly discuss some effects arising from the internal structure of a tunnel junction. We shall start in Sec. IV A by considering the case where the parameters $\Gamma_p(x)$, $\Gamma_f(x)$, $V_{gc}(x)$, and $V_{gs}(x)$ have a constant profile along the length L of the junction. Such seemingly simplified model of the tunnel junction provides in fact quite useful physical insights. In the first instance it clarifies the different role of the p- and f-tunneling processes, and thereby the physical origin of the general property that T_p depends on the energy E and can be tuned by V_{gc}, whereas T_f is independent of the energy and can be tuned by V_{gs}. Second, this case allows us to account for the effects of the finite length L of the junction on both T_p and T_f, which cannot be described by the conventional DT model of the tunnel region.

Then, in Sec. IV B we show how the constant-profile case can be exploited to construct a realistic model of an actual tunnel junction with arbitrarily varying parameters. In Sec. IV C we analyze the effect of a smooth variation of the absolute value $|\Gamma_p(x)|$ and $|\Gamma_f(x)|$ of the tunneling amplitudes, whereas in Sec. IV D we analyze the effect of phase fluctuations $\phi_p(x)$ and $\phi_f(x)$. Finally, in Sec. IV E we discuss the DT limit $L \to 0$, and show why such a widely used model subtly hides the factorization properties.

A. The case of a constant profile

Let us now consider constant profiles along the tunnel junction

$$\Gamma_\nu(x) = \begin{cases} \Gamma_\nu & \text{for } x_1 < x < x_2, \\ 0 & \text{otherwise,} \end{cases} \qquad \nu = p, f, \tag{50}$$

and

$$V_{gc(s)}(x) = \begin{cases} V_{gc(s)} & \text{for } x_1 < x < x_2, \\ 0 & \text{otherwise.} \end{cases} \tag{51}$$

Under the assumption (50) and (51), the $\mathbf{m}_f$ matrix (31) becomes

$$\begin{aligned} \mathbf{m}_f &= \mathrm{T}\, e^{-i\int_{x_1}^{x_2} \boldsymbol{\sigma}\cdot\mathbf{b}_f\, dx'} = e^{-i(\boldsymbol{\sigma}\cdot\mathbf{b}_f)L} \\ &= \sigma_0 \cos(\tilde{k}_f L) - i\boldsymbol{\sigma}\cdot\mathbf{b}_f \sin(\tilde{k}_f L)/\tilde{k}_f \end{aligned} \tag{52}$$

with

$$\tilde{k}_f = \sqrt{|\Gamma_f|^2 + (eV_{gs})^2}\,/\hbar v_F, \tag{53}$$

whereas the $\mathbf{m}_p$ matrix (32) is given by

$$\begin{aligned} \mathbf{m}_p &= e^{-i\tau_z k_E x_2}\left(\mathrm{T}\, e^{-i\int_{x_1}^{x_2} \boldsymbol{\tau}\cdot\mathbf{b}_{p,E}\, dx'}\right) e^{i\tau_z k_E x_1} = e^{-i\tau_z k_E x_2} e^{-i(\boldsymbol{\tau}\cdot\mathbf{b}_{p,E})L} e^{i\tau_z k_E x_1} \\ &= \begin{cases} e^{-i\tau_z k_E x_2}(\tau_0 \cos(\tilde{k}_E L) - i\boldsymbol{\tau}\cdot\mathbf{b}_{p,E}\sin(\tilde{k}_E L)/\tilde{k}_E)\cdot e^{+i\tau_z k_E x_1} & \text{for } |E - eV_{gc}| > |\Gamma_p|, \\ e^{-i\tau_z k_E x_2}(\tau_0 \cosh(\tilde{q}_E L) - i\boldsymbol{\tau}\cdot\mathbf{b}_{p,E}\sinh(\tilde{q}_E L)/\tilde{q}_E)\cdot e^{+i\tau_z k_E x_1} & \text{for } |E - eV_{gc}| < |\Gamma_p|, \end{cases} \end{aligned} \tag{54}$$

where we have denoted

$$\tilde{k}_E = \frac{\sqrt{(E - eV_{gc})^2 - |\Gamma_p|^2}}{\hbar v_F} \quad \text{for} \quad |E - eV_{gc}| > |\Gamma_p|, \quad \tilde{q}_E = \frac{\sqrt{|\Gamma_p|^2 - (E - eV_{gc})^2}}{\hbar v_F} \quad \text{for} \quad |E - eV_{gc}| < |\Gamma_p|. \tag{55}$$

The transmission coefficient related to spin-preserving tunneling is obtained from the $\mathbf{m}_p$ matrix (54) through Eq. (37), and reads

$$T_p(E) = \begin{cases} \left(1 + \frac{|\Gamma_p|^2}{(\hbar v_F \tilde{k}_E)^2}\, \sin^2(\tilde{k}_E L)\right)^{-1} & \text{if} \quad |E - eV_{gc}| > |\Gamma_p|, \\ \left(1 + \frac{|\Gamma_p|^2}{(\hbar v_F \tilde{q}_E)^2}\, \sinh^2(\tilde{q}_E L)\right)^{-1} & \text{if} \quad |E - eV_{gc}| < |\Gamma_p|, \end{cases} \tag{56}$$

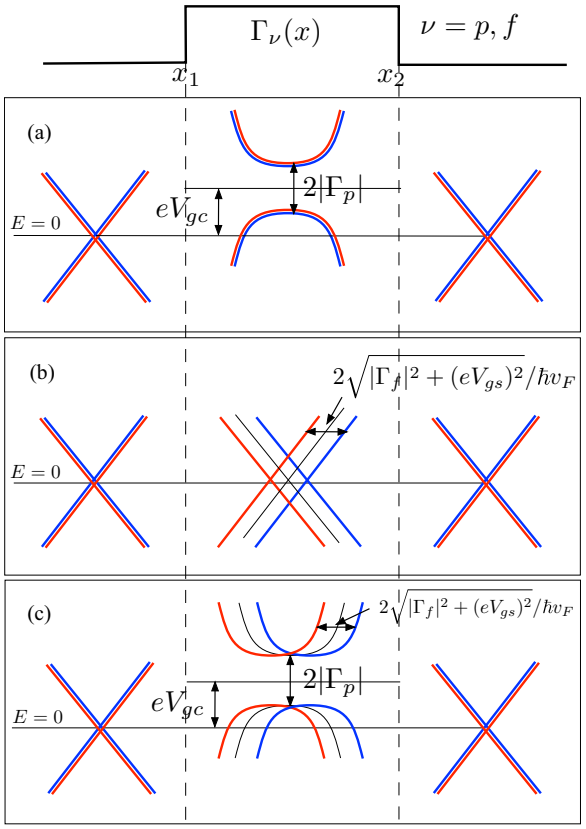

FIG. 3. (Color online) Effects of the spin-preserving and spin-flipping tunneling processes on the electronic spectrum of the edge states in the tunnel junction with profile (50) and (51). (a) The spin-preserving tunneling Γ_p tends to induce a gap in the spectrum (which in fact amounts to a reduction of the transmission, due to the finite length of the junction), whereas the charge bias V_{gc} shifts vertically the spectrum with respect to the Fermi level. (b) The spin-flipping tunneling Γ_f lifts the degeneracy of spin-↑ and spin-↓ edge modes, inducing a mixing of the eigenstate spin components in the tunnel junction as well as a mutual horizontal shift in the momenta. The spin gate V_{gs} renormalizes such shift. (c) The effect of both tunneling terms.

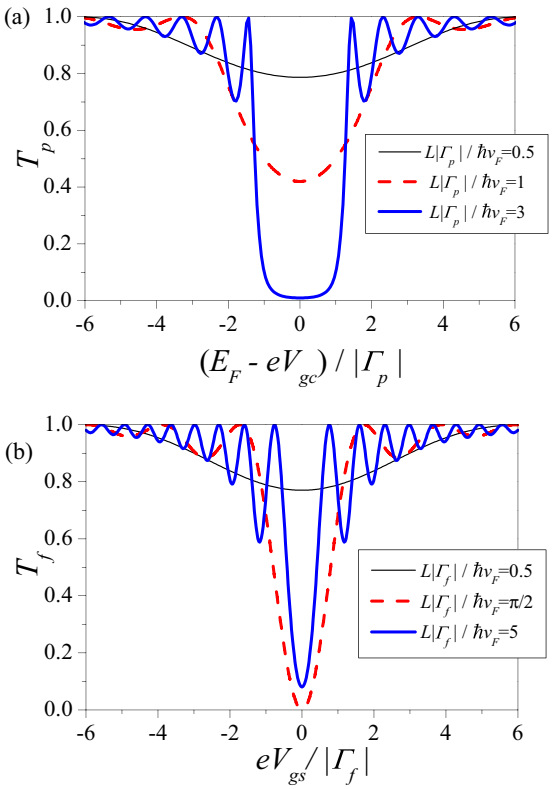

FIG. 4. (Color online) The case of a tunnel junction with a constant profile of the tunneling amplitude (50). (a) The transmission coefficient T_p due to spin-preserving tunneling (evaluated at the equilibrium Fermi level E_F) is plotted as a function of $E_F - eV_{gc}$ (with V_{gc} denoting the charge gate voltage applied in the tunnel region), for different values of the length of the junction. In the energy range $|E_F - eV_{gc}| < |\Gamma_p|$, T_p exhibits a minimum, which gradually becomes a "gap" with increasing length of the junction [see also Fig. 2(a)], whereas in the energy range $|E_F - eV_{gc}| > |\Gamma_p|$ one can observe oscillations due to the interference of electronic waves backward scattered at the ends of the tunnel junction. The amplitude and frequency of the oscillations depend on the spin-preserving strength of the tunnel junction $a_p = |\Gamma_p|L/\hbar v_F$: the oscillations increase in depth and frequency with increasing a_p. (b) The transmission coefficient T_f due to spin-flipping tunneling is plotted as a function of the spin gate voltage V_{gs} for various values of the length of the junction. In this case no "gapped" energy range is observed [see also Fig. 2(b)]. The oscillations are due to forward scattering interference and increase in amplitude and frequency with increasing $a_f = |\Gamma_f|L/\hbar v_F$.

whereas the transmission coefficient related to spin-flipping tunneling is easily obtained from the $\mathbf{m}_f$ matrix (52) through Eq. (38), and reads

$$T_f = 1 - \frac{|\Gamma_f|^2}{(\hbar v_F \tilde{k}_f)^2} \sin^2[\tilde{k}_f L]. \tag{57}$$

Equations (56) and (57) directly involve Γ_p, V_{gc} and Γ_f, V_{gs} respectively, and allow one to identify the physical effect of these terms.

We start by analyzing the role of the spin-preserving tunneling. These processes would tend—per se—to create a gap in the electronic spectrum [see Fig. 3(a)] [9]. In fact, an actual gap would be present only for an infinitely long tunnel junction ($L \to \infty$), whereas in a realistic tunnel junction with a finite length L two energy regimes can be identified. For $|E - eV_{g,c}| < |\Gamma_p|$ the electronic states in the tunnel junction region consist of evanescent waves in the longitudinal direction x, which decay as $\sim \exp[-\tilde{q}_E|x|]$, where $\tilde{q}_E$ is given in Eq. (55). In contrast, for $|E - eV_{g,c}| > |\Gamma_p|$ one has propagating waves, where the dispersion relation characterized by $\tilde{k}_E$ [see Eq. (55)] is not linear though, due to the interedge coupling. This explains why T_p depends on the energy E. Notice that the charge gate voltage V_{gc} produces a vertical shift of the dispersion relation by changing the position of the Fermi level with respect to the Dirac point, so that T_p is controlled by V_{gc}. These effects determine the behavior of T_p, which is plotted in Fig. 4(a) at the equilibrium Fermi energy $E = E_F$ as a function of $E_F - eV_{gc}$, for different values of the tunnel junction length L. The minimum of the transmission coefficient at the Dirac point $E_F = eV_{gc}$ corresponds to the highest value $\tilde{q}_{E_F} L = |\Gamma_p|L/\hbar v_F$ the evanescent

wave decay rate along the whole junction takes. While for short junction $|\Gamma_p|L/\hbar v_F < 1$ the value of the minimum is finite, by increasing the length L of the junction and/or the tunnel coupling Γ_p, one observes a strong suppression of the T_p minimum, which becomes a minigap as soon as $|\Gamma_p|L/\hbar v_F > 1$.

In contrast, the spin-flipping tunneling Γ_f lifts the degeneracy of spin-↑ and spin-↓ energy bands, by introducing an equal and opposite horizontal shift by $\pm\tilde{k}_f$ in the momenta of the dispersion relation, where $\tilde{k}_f$ is given by Eq. (53). Such shift is *independent* of the energy E of the incoming electron [see Fig. 3(b)], which explains why T_f is energy independent. Nevertheless, $\tilde{k}_f$ depends on V_{gs}, so that T_f can be controlled by the spin gate. Importantly, although $|\Gamma_f|$ and V_{gs} have the same effect on the dispersion relation, they have a quite different effect on the eigenstates. Indeed, while V_{gs} lifts the degeneracy by preserving the eigenstates, the coupling Γ_f mixes spin-↑ and spin-↓ states, and in the tunnel junction the eigenstates are not characterized by a unique spin orientation. Due to the factorization property these features hold also when both spin-preserving and spin-flipping tunneling are present [see Fig. 3(c)]. In Fig. 4(b), the spin-flipping transmission coefficient T_f [see Eq. (57)] is plotted as a function of the spin gate voltage V_{gs}, for different values of the junction length.

We notice that both T_p and T_f exhibit oscillations, which are both a signature of electron quantum interference, although the origin is different. The oscillations of T_p [see Figs. 4(a) and 5(a)] originate from the spin-preserving tunneling that changes the group velocity. It is therefore an interference induced by backward scattering at the two ends of the tunnel junctions, where the phase difference of the interfering waves is controlled by the charge gate voltage V_{gc}. The amplitude and frequency of these oscillations depend on the spin-preserving "strength" of the tunnel junction, i.e., on the dimensionless junction parameter $a_p = |\Gamma_p|L/\hbar v_F$, combining length L and tunneling amplitude $|\Gamma_p|$. Indeed the minima occur at energies $|E_F - eV_{gc}| \simeq |\Gamma_p|\sqrt{1+(m+1/2)^2\,(\pi/a_p)^2}$, and their related values are approximately $[1+(a_p/\pi)^2/(m+1/2)^2]^{-1}$, with $m \in \mathbb{Z}$. This is in agreement with the results of Ref. [23], where only spin-preserving tunneling inside the tunnel junction was considered. In practice, the oscillations increase in depth and frequency with increasing strength a_p. This implies that while in the regime $|E_F - eV_{g,c}| < |\Gamma_p|$ the transmission coefficient T_p decreases as a function of the tunnel junction strength a_p, in the regime $|E_F - eV_{g,c}| > |\Gamma_p|$ it oscillates with L, as shown in Fig. 5(a), where T_p at $E_F = 0$ is plotted as a function of a_p. The two different dependencies can be accessed by tuning the charge gate voltage V_{gc}.

In contrast, the oscillations of T_f [see Figs. 4(b) and 5(b)] originate from f-tunneling processes, which couple electronic waves with the same chirality. To illustrate this effect, let us imagine that a right-moving electron wave with spin-↑ is injected from terminal 2. Due to f-tunneling process, at the left end of the tunnel junction the electronic wave is split into two components, both propagating rightwards but with opposite spin orientation. At the right end of the junction another spin-flipping tunneling process may flip the spin-↓ component back to spin-↑, inducing interference with the transmitted wave. It is therefore a forward-scattering interference, where

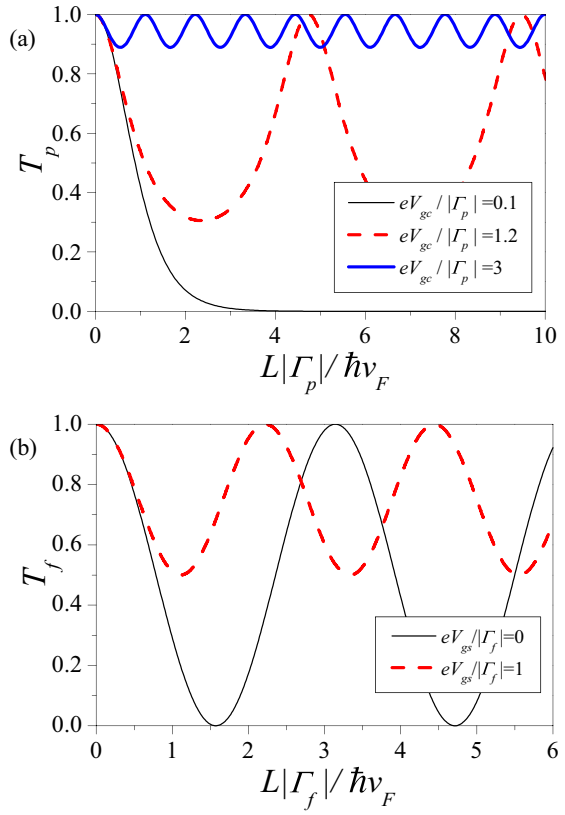

FIG. 5. (Color online) The behavior of the transmission coefficients T_p and T_f as a function of the length L of the tunnel junction is typically nonmonotonic. (a) The spin-preserving transmission coefficient T_p (evaluated at the energy $E_F = 0$) is plotted as a function of $L|\Gamma_p|/\hbar v_F$, for various charge gate voltages. While for $|eV_{g,c}| > |\Gamma_p|$ it oscillates with L, in the regime $|eV_{g,c}| < |\Gamma_p|$ it exponentially decreases. (b) T_f is plotted as a function of the length of the junction for various voltages V_{gs}. The behavior is always oscillatory. For particular values of the junction length $L|\Gamma_f|/\hbar v_F = (m+1/2)\pi$ (at $eV_{gs} = 0$) the coefficient T_f vanishes, leading to a complete spin flip.

the phase difference $\tilde{k}_f L$ accumulated along the junction is determined by the spin-gate V_{gs}. The amplitude and frequency of these oscillations depend on the spin-flipping "strength" of the tunnel junction, i.e., on the dimensionless parameter $a_f = |\Gamma_f|L/\hbar v_F$. Indeed the minima occur at spin gate voltage values $eV_{gs} \simeq |\Gamma_f|\sqrt{\pi^2(m+1/2)^2/a_f^2 - 1}$, and approximately take values $1-(a_f/\pi)^2/(m+1/2)^2$, with $m \in \mathbb{Z}$. In turn, this interference effect also implies a nonmonotonic dependence of T_f on the length of the junction, as shown in Fig. 5(b). With varying the value of V_{gs}, one can thus control the percentage of the transmitted current that flows to terminal 4 with spin-↓ with respect to the current flowing to terminal 3 with a spin-↑. Notice that for particular values $a_f = (m+1/2)\pi$ of the tunnel junction spin-flipping "strength" and for $eV_{gs} = 0$, the transmission T_f can even vanish, leading to a complete spin flip. This proves that spin-flipping tunneling processes have a dramatic impact on transport, where the helical nature of the edge states can be exploited to realize tunable spin polarizers for spintronics applications.

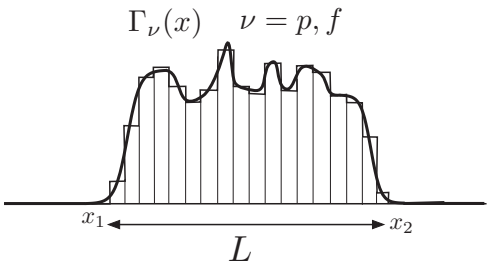

FIG. 6. Sketch of the tunneling amplitude profile $\Gamma_\nu(x)$ in the tunnel junction region $x_1 \leqslant x \leqslant x_2$ ($\nu = p$ for spin-preserving and $\nu = f$ for spin-flipping tunneling processes, respectively). An arbitrary profile can be approximated with a staircase profile, and the transfer matrix is easily obtained as a product of the constant profile transfer matrices [see Eq. (58)].

In Sec. IV C we shall discuss how the oscillations of T_p and T_f change when the sharp transition from vanishing tunneling amplitudes Γ_p and Γ_f outside the junction to a constant tunneling amplitude inside the junction [see Eq. (50)] is replaced by a smoother profile.

B. Generalization to an arbitrary profile

For an arbitrary profile of tunneling couplings $\Gamma_p(x)$ and $\Gamma_f(x)$ (see thick line in Fig. 6), the factorization property (30) of the transfer matrix implies that one can compute $\mathbf{m}_p$ and $\mathbf{m}_f$ separately, and that the transmission coefficients T_p and T_f are then evaluated through Eqs. (37) and (38). In order to compute $\mathbf{m}_p$ and $\mathbf{m}_f$, we notice that the profile can be approximated with the desired accuracy by an N-step staircase profile (thin line of Fig. 6). Then, exploiting a general property of transfer matrices [30], $\mathbf{m}_p$ is straightforwardly obtained as the product of the $\mathbf{m}_p^{(n)}$ matrices related to each individual nth stair step, characterized by a locally constant profile (here $n = 1, \ldots N$, from right to left). Similarly for $\mathbf{m}_f$, obtaining

$$\mathbf{m}_\nu = \prod_{n=1}^{N} \mathbf{m}_\nu^{(n)}, \qquad \nu = p, f. \tag{58}$$

One can thus use the case of a constant profile investigated in Sec. IV A as a building block to model a tunnel junction with arbitrary profiles for $\Gamma_p(x)$, $\Gamma_f(x)$, $V_{gc}(x)$, and $V_{gs}(x)$. In the following sections we apply this general method to investigate the effects of smoothing length and phase fluctuations in the tunneling amplitudes $\Gamma_\nu(x)$. For simplicity we shall restrict the gate profiles $V_{gc}(x)$ and $V_{gs}(x)$ to a constant gate profile (51).

C. Effects of a finite smoothing length

We investigate here the case where the tunneling amplitudes $\Gamma_\nu(x)$ ($\nu = p, f$) change from a vanishing value (outside the tunnel junction) to a "bulk" value Γ_ν^0 over a *finite* smoothing length λ, as shown in the profile depicted in Fig. 7(a). For simplicity, we shall consider the variation of the absolute value $|\Gamma_\nu(x)|$, and assume that the phase $\phi_\nu(x)$ remains constant. For a given value of the smoothing length λ, the actual profile is approximated by a staircase profile with N_λ steps in the smoothing region λ, as described in Sec. IV B. The value of N_λ is increased until convergence in the transmission coefficients is reached. In practice, it turns out that a number $N_\lambda \sim 20\lambda/L$ of steps is sufficient to reach a convergence in the results, so

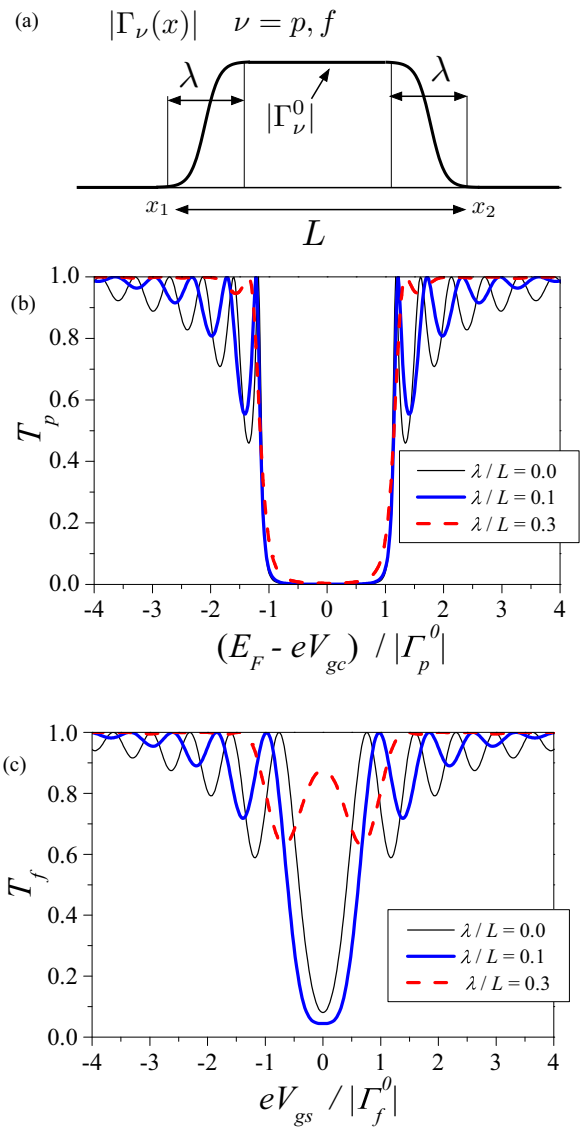

FIG. 7. (Color online) (a) Sketch of a tunnel amplitude profile $\Gamma_\nu(x)$, where the transition from a vanishing tunneling coupling to a bulk value Γ_ν^0 occurs over a smoothing length scale λ (here $\nu = p, f$ for spin-preserving and spin-flipping processes, respectively), where L is the total length of the junction. (b) The effect of a smoothing length λ on the spin-preserving transmission coefficient: T_p (evaluated at the equilibrium Fermi level E_F) is plotted as a function of $E_F - eV_{gc}$, for different values of the ratio λ/L, for the case $L|\Gamma_p^0|/\hbar v_F = 5$. While the minigap region is essentially unaffected by λ, the amplitude and frequency of the oscillations are reduced as λ/L increases. They are still visible as long as one can define a longitudinal "bulk" of the tunnel junction with a constant value Γ_0, i.e., for $\lambda/L < 0.25$. (c) The effect of a smoothing length λ on the spin-flipping transmission coefficient: T_f is plotted as a function of $eV_{g,s}$, for the case $L|\Gamma_f^0|/\hbar v_F = 5$. With increasing λ, the minimum at $V_{gs} = 0$ exhibits a nonmonotonic behavior, decreasing for $\lambda/L = 0.1$, and then increasing and even turning into a local maximum for $\lambda/L = 0.3$. Similarly to the oscillations of T_p, also for T_f the amplitude of the oscillations is suppressed when the tunnel junction profile becomes very smooth.

that, e.g., for $\lambda/L = 0.2$ the correct result is obtained by using only $N_\lambda = 4$. This proves that the method can be implemented with ordinary numerical routines in a treatable and fast way, thereby proving its usefulness and flexibility in handling an arbitrary tunnel junction.

The results are plotted in Fig. 7. In particular, in Fig. 7(b) the spin-preserving transmission coefficient T_p (evaluated at the equilibrium Fermi energy E_F) is plotted as a function of $E_F - eV_{gc}$, for different values of the smoothing length λ, with $\lambda = 0$ corresponding to the case of the constant profile discussed in Sec. IV A. With increasing λ, the suppression of the transmission coefficient in the "subgap" region $|E_F - eV_{gc}| < |\Gamma_p^0|$ remains unaffected, whereas the visibility of the oscillations appearing in the "supragap" regime $|E_F - eV_{gc}| > |\Gamma_p^0|$ tends to be suppressed. The spin-flipping transmission coefficient T_f, plotted in Fig. 7(c), exhibits a nonmonotonic behavior of the $V_{gs} = 0$ minimum by increasing λ: with respect to the case of constant profile (thin solid black line) it decreases for $\lambda/L = 0.1$, whereas it increases and even turns into a local maximum for $\lambda/L = 0.3$. Similarly to the oscillations of T_p, also for T_f the amplitude of the oscillations is suppressed when the tunnel junction profile becomes very smooth ($\lambda \sim L/2$), like in a quantum point contact [31]. This is due to an effective averaging over various lengths of the backscattering processes causing the interference behavior. In contrast, the oscillations are fairly visible as long as a "bulk" of the junction can be identified (i.e., for $\lambda < 0.25$). By combining etching and lithographic techniques this can easily be realized in QSHE systems based on HgTe/CdTe quantum wells [32]. In this case, because $|\Gamma_\nu(x)|$ depends exponentially on the transversal width of the tunnel junction [33], it is reasonable that $|\Gamma_\nu(x)| \simeq$ constant inside the junction, and rapidly vanishing $|\Gamma_\nu(x)|$ at the ends of the tunnel region.

D. Effect of phase fluctuations

The tunneling amplitude is in general a complex function $\Gamma_\nu(x) = |\Gamma_\nu(x)| \exp[i\phi_\nu(x)]$, characterized by an absolute value and a phase. While the effect of variation of the absolute value $|\Gamma_\nu(x)|$ has been considered in Secs. IV A and IV C, here we would like to focus on the role of the phase profile $\phi_\nu(x)$. To begin with, we observe that for the case of a constant profile $\Gamma_\nu(x) \equiv \Gamma_\nu \exp[i\phi_\nu]$ [see Eq. (50)] the specific value of the phase ϕ_ν is irrelevant, and only the absolute value $|\Gamma_\nu|$ matters in determining T_p and T_f [see Eqs. (56) and (57)]. While the assumption $|\Gamma(x)| \simeq$ constant seems to be fairly reasonable in the central tunnel region, the situation may be different for the phase $\phi_\nu(x)$. One can expect that, especially in the presence of disorder, local potential fluctuations at each side of the tunnel junction may lead to random changes in the local Fermi wave vector k_F, affecting the electron phase $\sim \exp(ik_F x)$. At a given longitudinal position x, the transversal overlap integral determining the tunneling amplitude $\Gamma_\nu(x)$ may thus acquire phase fluctuations. To discuss these effects we shall consider a profile where the absolute value $|\Gamma_\nu(x)|$ remains constant, and the phase $\phi_\nu(x)$ fluctuates along the junction over a typical length scale λ_{ϕ_ν}, as shown in Fig. 8(a). The effect on the transmission coefficients T_p and T_f are shown in panels (b) and (c). In particular, in panel (b) T_p is plotted as a function of $E_F - eV_{gc}$. As compared to the case of a constant

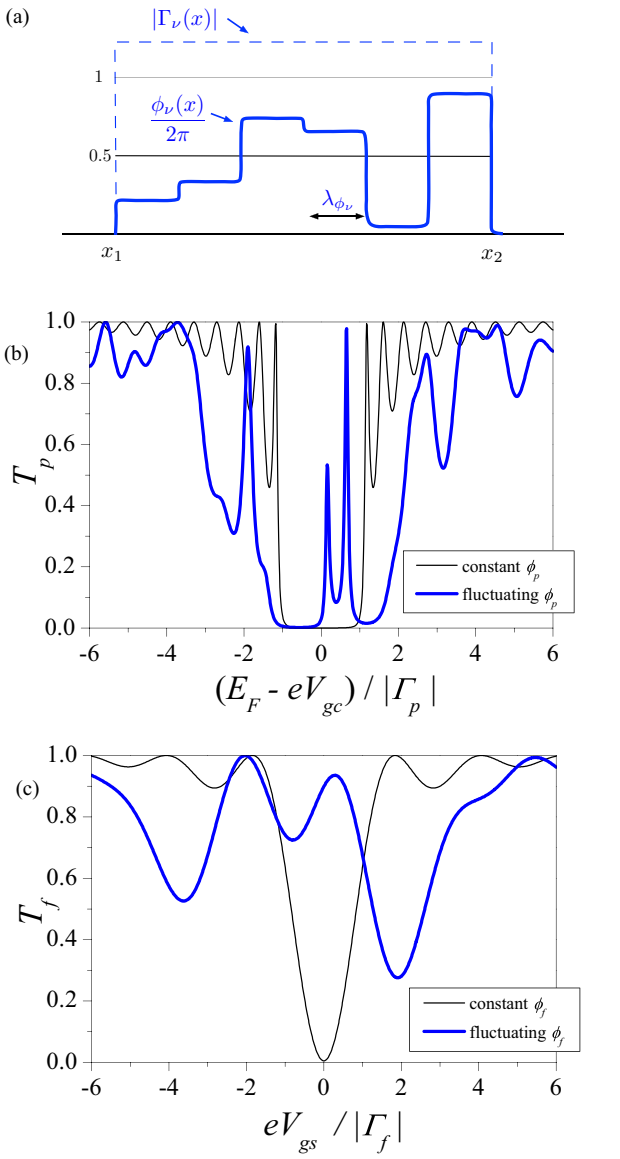

FIG. 8. (Color online) (a) The phase ϕ_ν of the tunneling amplitude $\Gamma_\nu(x) = |\Gamma_\nu(x)| \exp[i\phi_\nu(x)]$ is assumed to fluctuate inside the junction over a typical length scale λ_{ϕ_ν}, whereas the absolute value $|\Gamma_\nu(x)|$ is assumed to be constant ($\nu = p$ for spin-preserving and $\nu = f$ for spin-flipping tunneling). (b) The spin-preserving transmission coefficient T_p (evaluated at the equilibrium Fermi energy E_F) as a function of $E_F - eV_{gc}$, for the case $L|\Gamma_p|/\hbar v_F = 5$. With respect to the case of constant ϕ_p (black thin line), the fluctuations of ϕ_p induce resonance maxima which appear inside the the "gap" region $|E_F - eV_{gc}| < |\Gamma_p|$, which in turn also broadens, whereas the oscillations in the "supragap" region $|E_F - eV_{gc}| > |\Gamma_p|$ are enhanced (blue thick line). The curve becomes also asymmetric with respect to $E_F - eV_{gc}$. (c) The spin-flipping transmission coefficient T_f is plotted as a function of $eV_{g,c}$, for the case $L|\Gamma_f|/\hbar v_F = 1.5$: the fluctuations of the phase ϕ_f modify the oscillatory pattern with respect to the case of constant ϕ_f (black thin line), by increasing the amplitude and making the plot asymmetric with respect to $V_{gs} = 0$ (blue thick line).

phase ϕ_ν (black thin line), the curve determined by phase fluctuations (blue thick line) exhibits various features: (i) the appearance of some resonance maxima inside the "gapped"

region $|E_F - eV_{gc}| < |\Gamma_p|$, whose location depends on λ_{ϕ_ν} and on the typical deviation $\Delta\phi_p$ of the fluctuations around the average phase $\langle\phi_\nu\rangle$; (ii) the broadening of the "subgap" region $|E_F - eV_{gc}| < |\Gamma_p|$; (iii) an enhancement of the amplitude of the oscillations in the "supragap" region $|E_F - eV_{gc}| > |\Gamma_p|$; (iv) the symmetry of T_p with respect to $E_F - eV_{gc}$ is lost. The last two effects are particularly striking in the behavior of T_f [see panel (c)], where the amplitude of the oscillatory pattern is increased and T_f is asymmetric in the spin gate bias V_{gs}.

E. The δ-tunneling (DT) limit

We now want to compare the case of finite length tunnel junction with the widely used [8,14–25] δ-tunneling model of a pointlike constriction. Such model amounts to adopting sharply peaked profiles for $\Gamma_p(x)$ and $\Gamma_f(x)$ in Eqs. (6) and (7); i.e.,

$$\Gamma_\nu(x) = 2\hbar v_F\, \gamma_\nu^{\rm DT}\, \delta(x), \quad \nu = p, f, \tag{59}$$

where $\gamma_{p/f}^{\rm DT}$ denote dimensionless δ-tunneling amplitude parameters. Solving the field equation of motion (17) for the δ-profile (59), one can determine the conductance matrix $\mathrm{G}^{\rm DT}$ describing the transmission coefficients between the four terminals, obtaining [20]

$$\begin{aligned}
\mathrm{G}_{12}^{\rm DT} = \mathrm{G}_{34}^{\rm DT} &= -\frac{e^2}{h}\frac{4\left|\gamma_p^{\rm DT}\right|^2}{\left(1+\left|\gamma_p^{\rm DT}\right|^2+\left|\gamma_f^{\rm DT}\right|^2\right)^2},\\
\mathrm{G}_{31}^{\rm DT} = \mathrm{G}_{42}^{\rm DT} &= -\frac{e^2}{h}\frac{4\left|\gamma_f^{\rm DT}\right|^2}{\left(1+\left|\gamma_p^{\rm DT}\right|^2+\left|\gamma_f^{\rm DT}\right|^2\right)^2},\\
\mathrm{G}_{41}^{\rm DT} = \mathrm{G}_{32}^{\rm DT} &= -\frac{e^2}{h}\frac{\left(1-\left|\gamma_p^{\rm DT}\right|^2-\left|\gamma_f^{\rm DT}\right|^2\right)^2}{\left(1+\left|\gamma_p^{\rm DT}\right|^2+\left|\gamma_f^{\rm DT}\right|^2\right)^2}.
\end{aligned} \tag{60}$$

Importantly, the coefficients (60) are *not* factorized into a spin-preserving and a spin-flipping contribution. This lack of factorization seems to contradict the result found above for an arbitrary tunneling profile, since the DT model should be recovered from the finite length junction as the limit of $L \to 0$ of short length. To solve this seeming paradox one can proceed as follows. Instead of adopting the mathematical pointlike tunneling profile (59), one can follow a physically more correct procedure starting from a model where the tunnel junction has a finite length, and taking the limit of vanishing length. This is for instance accomplished by assuming the constant profile described in Sec. IV A, where $\Gamma_\nu(x) = \Gamma_\nu\,\theta(x - x_1)\theta(x_2 - x)$, with $\Gamma_\nu = |\Gamma_\nu|e^{i\phi_\nu}$. Taking the limit of short tunnel junction $L = x_2 - x_1 \to 0$ and $|\Gamma_\nu| \to \infty$, with keeping the tunnel junction strengths $L|\Gamma_\nu|/\hbar v_F =$ constant, one can operatively identify the expressions for the coefficients $\gamma_{p/f}^{\rm DT}$ in terms of $L|\Gamma_p|/\hbar v_F$ and $L|\Gamma_f|/\hbar v_F$. A lengthy but straightforward calculation leads to

$$\gamma_p^{\rm DT} = \frac{\sinh\left(\frac{L|\Gamma_p|}{\hbar v_F}\right)e^{i\phi_p}}{\cosh\left(\frac{L|\Gamma_p|}{\hbar v_F}\right)+\cos\left(\frac{L|\Gamma_f|}{\hbar v_F}\right)}, \tag{61}$$

$$\gamma_f^{\rm DT} = \frac{\sin\left(\frac{L|\Gamma_f|}{\hbar v_F}\right)e^{i\phi_f}}{\cosh\left(\frac{L|\Gamma_p|}{\hbar v_F}\right)+\cos\left(\frac{L|\Gamma_f|}{\hbar v_F}\right)}. \tag{62}$$

Equations (61) and (62) show that the bare tunneling amplitudes $\gamma_{p/f}^{\rm DT}$ used in the mathematical δ-like profile (59) actually depend on *both* the physical spin-preserving and spin-flipping tunneling amplitudes Γ_p and Γ_f of the more realistic (i.e., narrow but finite) constriction model.

Physically, this seemingly surprising result can be understood as follows. Let us focus, for instance, on the spin-flipping channel: A spin-flipping tunneling event can be either direct, i.e., resulting from one single f-process, or indirect, i.e., "dressed" by additional tunneling events occurring along the junction. In particular, also an even number of p-processes can contribute to the spin-flipping tunneling, with a weight determined by the strength $a_p = |\Gamma_p|L/\hbar v_F$ of the spin-preserving tunnel coupling, which combines the local tunneling amplitude $|\Gamma_p|$ and the length L of the junction. At first, one is tempted to think that the DT limit, where the length L of the tunnel junction vanishes, only involves direct tunneling events. However, because $|\Gamma_p| \to \infty$ and $|\Gamma_p|L$ is kept constant, dressing p-processes do matter if $|\Gamma_p|L/\hbar v_F \sim 1$, so that the parameter $\gamma_f^{\rm DT}$ appearing in Eq. (59) in fact describes the overall result of both the direct f-tunneling and all the dressing p-tunneling events. Similarly for the other channel. At mathematical level, such effect originates from the fact that, when the size of the tunneling region becomes small, the wave function inside the constriction stretches and eventually becomes discontinuous in the limit $L \to 0$. The discontinuity is given by the integral $\int dx\,\Psi(x)$ over the tunneling region $x_1 < x < x_2$. Thus, although the space "evolution" for $\Psi(x)$ is factorized into a product of p- and f-contributions [see Eq. (28)], its integral is not, $\int dx\,[\mathbf{U}_f(x)\otimes\mathbf{U}_{p,E}(x)] \neq [\int dx\mathbf{U}_f(x)]\otimes[\int dx\,\mathbf{U}_{p,E}(x)]$, so that $\int dx\,\Psi(x)$ depends on both Γ_p and Γ_f in a nontrivial way. This integral is precisely what determines the parameters $\gamma_p^{\rm DT}$ and $\gamma_f^{\rm DT}$ of the DT model (59). Such singular behavior of the Dirac equation in the presence of a δ-like potential or tunneling profiles has long been known [34], and similar technical subtleties arise also in other physical situations, such as the transport of chiral electrons in graphene through gapped regions [35] and tunneling through Majorana states appearing at the edges of superconductors [36].

Importantly, using Eqs. (61) and (62) to reexpress the mathematical tunneling amplitude $\gamma_{p/f}^{\rm DT}$ appearing in the DT model in terms of the physical parameters $L\Gamma_p/\hbar v_F$ and $L\Gamma_f/\hbar v_F$, the conductance matrix entries (60) do acquire the factorized form (44), with spin-preserving and spin-flipping transmission coefficients

$$T_p^{\rm DT} = \cosh^{-2}(L|\Gamma_p|/\hbar v_F), \tag{63}$$

$$T_f^{\rm DT} = \cos^2(L|\Gamma_f|/\hbar v_F), \tag{64}$$

consistently with the limit $L \to 0$ and $|\Gamma_{p/f}| \to \infty$ in Eqs. (56) and (57). This proves that in the DT model the factorization is only seemingly lacking, and is hidden in the physically misleading parametrization in terms of $\gamma_{p/f}^{\rm DT}$.

This comparison allows one to identify the range of validity of the DT model (59). The DT model is applicable when (i) both the tunnel junction strength parameters are small (i.e., $L|\Gamma_\nu|/\hbar v_F \ll 1$, with $\nu = p, f$) and (ii) when one is probing energy ranges smaller than $|\Gamma_\nu|$, i.e., when $|E_F - eV_{gc}| \ll |\Gamma_p|$ and $|V_{gs}| \ll |\Gamma_f|$. If the first condition is not fulfilled,

then for each channel $\nu = p, f$ the related bare DT parameter $\gamma_\nu^{\rm DT}$ should actually be "dressed" by contributions arising from higher order processes of the other channel [see Eqs. (61) and (62)]. If the second condition is not fulfilled, the DT model cannot reproduce the features arising from the internal structure of the tunnel junction, such as the oscillatory pattern shown in Fig. 4.

Within the validity regime of the DT model, one can directly determine the tunneling parameters $|\Gamma_p|$ and $|\Gamma_f|$ to the measured transmission coefficients. Indeed by performing transconductance measurements, one can extract [Eqs. (46) and (47)]

$$T_p^{\rm DT} = 1 - \frac{h}{e^2}\left|{\rm G}_{12}^{\rm DT}\right|, \tag{65}$$

$$T_f^{\rm DT} = \frac{\left|{\rm G}_{32}^{\rm DT}\right|}{\frac{e^2}{h} - \left|{\rm G}_{12}^{\rm DT}\right|}. \tag{66}$$

Then, by using Eqs. (63) and (64), one obtains

$$|\Gamma_p| = \frac{\hbar v_F}{L}{\rm arctanh}\left(\sqrt{1 - T_p^{\rm DT}}\right), \tag{67}$$

$$|\Gamma_f| = \frac{\hbar v_F}{L}\arcsin\left(\sqrt{1 - T_f^{\rm DT}}\right). \tag{68}$$

V. DISCUSSION AND CONCLUSIONS

We have investigated a tunnel junction coupling the helical states flowing at the two edges of a 2D quantum spin Hall system (see Fig. 1), which can be realized by etching a constriction in a HgTe/CdTe or InAs/GaSb quantum well. In such situation, electron tunneling occurs through two types of time-reversal symmetric channels, namely spin-preserving (p) and spin-flipping (f) processes, making such system a bench test for possible applications of helical edge states in spintronics. Indeed, due to the helical nature of the edge states, currents in the setup can in principle be switched from a terminal to another with either preserving or flipping the spin orientation. To this purpose, the crucial issue is to determine and control the transmission coefficients related to the two types of tunneling processes. This challenging task involves various difficulties, arising from the fact that the Hamiltonian terms describing these two processes do not commute, and that a tunnel region has a finite length and a typically irregular and disordered profile, so that the tunneling amplitude of each of the two processes cannot be described by one single parameter but rather by a space-dependent profile.

We have demonstrated that there exists an operative way to separately extract the transmission coefficients T_p and T_f, related to p- and f-processes, respectively. Indeed, despite the noncommutativity of the two tunneling terms, the analysis of the scattering matrix of the 4-terminal setup has revealed that its entries are always factorized into two terms, one depending on the p-processes only and another one depending on the f-processes only. This factorization of the scattering matrix entries directly leads to the factorization of the conductance matrix entries ${\rm G}_{ij}$, which determine the current flowing in terminal i when a voltage bias is applied to terminal j. It is thus possible to extract, via transconductance measurements, the transmission coefficients T_p and T_f related to these two processes (see Fig. 2). Furthermore, by considering the presence of two electric gates across the junction, characterized by gate voltages $V_{g,T}$ and $V_{g,B}$, we have shown that T_p is controlled by the charge gate $V_{gc} = (V_{g,T} + V_{g,B})/2$ only, whereas T_f is controlled by the spin gate $V_{gs} = (V_{g,T} - V_{g,B})/2$ only.

The factorization of the scattering matrix entries is seemingly lacking in the customary DT model for tunnel junction, and we have shown that it is in fact subtly hidden in a misleading parametrization of the coupling constants of that model. In fact, we have proved that the factorization property holds for an arbitrary profile of the tunneling amplitudes $\Gamma_\nu(x)$ $(\nu = p, f)$. This enables one to go beyond the DT model, and to investigate also the effects arising from the internal structure of the tunnel junction on the transmission coefficients T_p and T_f. We have first considered the effects of the finite length L of the tunnel junction. In particular we have shown that L determines the existence of two energy regimes on T_p. In the range $|E_F - eV_{gc}| < |\Gamma_p|$ the coefficient T_p exhibits a minimum, which is considerably suppressed and tends to acquire a gaplike feature when the length of the tunnel junction is increased [see Fig. 4(a)], whereas in the range $|E_F - eV_{gc}| > |\Gamma_p|$ the coefficient T_p exhibits an oscillatory behavior, with a period related to the length of the junction through the spin-preserving strength $a_p = L|\Gamma_p|/\hbar v_F$. The effect of the finite length on the spin-flipping transmission coefficient T_f is different. This is due to the fact that Γ_f breaks the spin degeneracy of the helical states, without the tendency to create a gap. Thus, T_f exhibits a minimum as a function of the spin-gate potential eV_{gs}, which never becomes a flat dip [see Fig. 4(b)]. Oscillatory behavior is still present, similarly to the "supragap" region of T_p. The coefficients T_p and T_f exhibit a nonmonotonic behavior as a function of the length L of the tunnel junction (see Fig. 5).

We have then investigated how the scenario changes when the profile of the tunneling amplitudes varies from a vanishing value outside the junction up to a constant value Γ_ν^0 inside the junction over a smoothing length λ [see Fig. 7(a)]. As far as T_p is concerned, the presence of the smoothing length λ has a minor effect in the "subgap" region $|E - eV_{gc}| < |\Gamma_p|$, whereas the amplitude of the oscillations occurring in the "supragap" region tends to be suppressed as λ/L is increased [see Fig. 7(b)]. In contrast, as far as T_f is concerned, the minimum at $V_{gs} = 0$ exhibits a nonmonotonic behavior as a function of λ and, for $\lambda/L > 0.1$, it turns into a maximum [see Fig. 7(c)]. As a whole, the amplitude of the oscillations is suppressed as λ/L increases. However, the oscillations are still visible as long as a longitudinal "bulk" with a roughly constant $|\Gamma_\nu^0|$ can be identified, i.e., for $\lambda/L < 0.25$.

Then, we have investigated the role of fluctuations of the phase ϕ_ν of the tunneling amplitude [see Fig. 8(a)] $\Gamma_\nu(x) = |\Gamma_\nu(x)|\exp[i\phi_\nu(x)]$, arising from disorder in the tunnel junction. We have seen that in the presence of random fluctuations of the spin-preserving tunneling amplitude phase ϕ_p, resonances appear in the "subgap" region $|E_F - eV_{gc}| < |\Gamma_p|$ of T_p, whereas the amplitude of the oscillations in the "supragap" region $|E_F - eV_{gc}| > |\Gamma_p|$ is also enhanced [see

Fig. 8(b)]. Furthermore, T_p is no longer symmetric with respect to $E_F - eV_{gc}$ in the presence of such fluctuations. On the other hand, the fluctuations of the phase ϕ_f of the spin-flipping tunneling amplitude lead to an enhancement of the amplitude and location of the oscillations of T_f as a function of the spin gate voltage V_{gs} [see Fig. 8(c)].

Experimental conditions. Let us now briefly discuss the experimental conditions to realize the setup. It is well known that, similarly to graphene, the helical edge states of QSHE cannot be confined simply by electrical gating, due to the linear Dirac-like spectrum and the Klein tunneling. Tunnel junctions in QSHE are thus typically realized by lateral etching of the quantum well, and lithographic techniques can be exploited to tailor arbitrary shapes. Once the etched constriction induces electron tunneling, lateral gates can be used to control it, as described above. For a typical tunnel region of a width $W \sim 100$ nm, one can estimate $|\Gamma_p| \sim 1.3$ meV and $|\Gamma_f| \sim 0.3$ meV [9,23,33]. Notice that $|\Gamma_f| \sim |\Gamma_p|/4$; i.e., $|\Gamma_f|$ is smaller, but not negligible with respect to $|\Gamma_p|$. These values, together with the length L of the junction, determine the variation range for the gate voltages V_{gc} and V_{gs} to vary T_p and T_f by a significant amount (see, e.g., Fig. 4). For a $L \sim 1$ μm long junction, this range is a few meV. These values are well below the bulk gap and are consistent with the typical experimental conditions [2], so that T_p and T_f should be tunable in these regimes and display the internal structure of the junction.

Applications in spintronics. Our results suggest the possibility to exploit the setup in Fig. 1 as a building block for spintronics nanodevices. The underlying idea is the following: due to the factorization property shown here, for any given device in Fig. 1 one can first determine T_p and T_f through transconductance measurements as described above. In particular, one can extract the dependence of T_p and T_f on the related gate voltage V_{gc} and V_{gs}, respectively. Such "spectrum" of T_p and T_f depends on the specific features of the tunnel junction (and possibly on the presence of disorder) and represents a sort of fingerprints of the tunnel junction. Then, by tuning the values of V_{gc} and V_{gs} according to the obtained fingerprints, one can operate electrically on T_p and T_f, independently, thereby realizing a device where spin-polarized currents can be steered and redistributed in the four terminals.

Effects of electron-electron interaction. Although the analysis of the interacting case is beyond the purpose of the present paper, we would like to briefly discuss this aspect, addressing the main underlying issues. Theoretical predictions show that, in the presence of electron-electron interaction, a pair of helical edge states realize a helical Luttinger liquid (LL) where, besides a repulsively interacting charge sector characterized by a Luttinger parameter $g < 1$, the spin sector is also interacting with an attractive strength $1/g$ [6,8,14–16] ($g = 1$ corresponds to the noninteracting limit). This unconventional Luttinger liquid is thus particularly interesting and, despite that no clear experimental evidence of interaction effects on helical edge states has been observed so far, the search for conditions where these effects can be disguised is a fascinating problem. When a tunnel junction is realized, electron-electron interaction interplays with tunneling in a nontrivial way, leading to qualitatively different features as compared to the noninteracting case. In the limit of a short junction, and for $1/2 < g < 1$ (a range that includes the experimentally plausible regime $g \lesssim 1$ of weak interaction) the analysis based on the the DT model [8,14,16] shows that p- and f-tunneling terms are both irrelevant operators, with the same scaling dimension, due to the $g \leftrightarrow 1/g$ relation between charge and spin sector. Corrections to the ideal conductance are thus due to the finite bias voltage and/or temperature, and appear as a power law with a g-dependent exponent. However, when the finite length L of the junction is taken into account, the problem becomes intrinsically more complicated for various reasons. In the first instance, besides tunneling terms, also interedge forward interaction terms arise along the junction region, similarly to a spinful LL, breaking the $g \leftrightarrow 1/g$ relation that holds away from the junction [15,37]. As a consequence, p- and f-tunneling terms acquire different scaling dimensions and, even to lowest order in tunneling, qualitative modifications in the bias voltage dependence of the conductance are expected as compared to the interacting DT model. In the case of stronger tunneling $|\Gamma_\nu L|/\hbar v_F \sim 1$, these modifications may even be more significant because of the "dressing" of each DT tunneling amplitude by higher order contributions stemming from the other tunneling channel, similarly to the noninteracting case Eq. (62). Furthermore, interedge interaction also involves two types of 2-particle backward scattering, which preserve and flip spin, respectively [8,15,37]. Finally, while along a noninteracting edge Rashba spin-orbit coupling cannot induce single-particle backscattering, in the presence of electron-electron interaction such coupling can lead to an effective 2-particle backscattering along the edge [38,39]. In a tunnel junction, such intraedge effect is expected to interplay with interedge tunneling and interaction. The whole problem can thus be formulated in terms of two coupled Luttinger liquids (for charge and spin sectors) with inhomogeneous interaction parameters $g_c(x)$ and $g_s(x)$, and with inhomogenous nonlinear coupling arising from both tunneling and interaction terms, over the junction length. This highly nontrivial problem does not have an exact solution is general, and deserves a specific analysis. On the basis of results obtained in some specific cases and on formal similarities with inhomogeneous LLs in the presence of impurities [17,40], we can formulate some expectations for the case of weak tunneling $|\Gamma_\nu L|/\hbar v_F \ll 1$. In this case the conductances are modified with respect to the DT model by a modulation factor f, which depends in a nonmonotonic way on the the ratio eV/E_L between the bias voltage V and the energy scale $E_L \propto v_F/L$ associated with L. Also, the period of the oscillations shown, e.g., in Fig. 4 should be modified by an interaction-dependent factor. Whether the factorization persists in the presence of interaction is a challenging question.

ACKNOWLEDGMENTS

The authors greatly acknowledge the German-Italian Vigoni Program for financial support, and B. Trauzettel and H. Buhmann for fruitful discussions. F.D. also acknowledges financial support from FIRB 2012 project HybridNanoDev (Grant No. RBFR1236VV).

[1] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005); **95**, 226801 (2005); B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, Science **314**, 1757 (2006); B. A. Bernevig and S.-C. Zhang, Phys. Rev. Lett. **96**, 106802 (2006).

[2] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang, Science **318**, 766 (2007); 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); A. Roth, C. Brüne, H. Buhmann, L. W. Molenkamp, J. Maciejko, X.-L. Qi, and S.-C. Zhang, Science **325**, 294 (2009); 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).

[3] Y. Ma *et al*., arXiv:1212.6441; K. C. Nowack *et al*., Nat. Mater. **12**, 787 (2013).

[4] N. Nagaosa, Science **318**, 758 (2007); L. Fu and C. L. Kane, Phys. Rev. B **76**, 045302 (2007); D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature (London) **452**, 970 (2008); X.-L. Qi, T. L. Hughes, and S.-C. Zhang, Phys. Rev. B **78**, 195424 (2008); H. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, Nat. Phys. **5**, 438 (2009); Y. Xia *et al*., *ibid*. **5**, 398 (2009); D. Hsieh *et al*., Science **323**, 919 (2009); Y. L. Chen *et al*., *ibid*. **325**, 178 (2009).

[5] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010); X.- L- Qi and S. C. Zhang, *ibid*. **83**, 1057 (2011).

[6] C. Wu, B. A. Bernevig, and S.-C. Zhang, Phys. Rev. Lett. **96**, 106401 (2006).

[7] C. Liu, T. L. Hughes, X.-L. Qi, K. Wang, and S.-C. Zhang, Phys. Rev. Lett. **100**, 236601 (2008); I. Knez, R.-R. Du, and G. Sullivan, *ibid*. **107**, 136603 (2011); L. Du, I. Knez, G. Sullivan, and R.-R. Du, arXiv:1306.1925.

[8] J. C. Y. Teo and C. L. Kane, Phys. Rev. B **79**, 235321 (2009).

[9] V. Krueckl and K. Richter, Phys. Rev. Lett. **107**, 086803 (2011).

[10] G. Dolcetto, F. Cavaliere, D. Ferraro, and M. Sassetti, Phys. Rev. B **87**, 085425 (2013).

[11] J. M. Kikkawa and D. D. Awschalom, Nature (London) **397**, 139 (1999); S. A. Wolf, D. D. Awschalom, R. A. Buhrman, J. M. Daughton, S. von Molnár, M. L. Roukes, A. Y. Chtchelkanova, and D. M. Treger, Science **294**, 1488 (2001); S. Murakami, N. Nagaosa, and S.-C. Zhang, *ibid*. **301**, 1348 (2003); I. Žutić, J. Fabian, and S. Das Sarma, Rev. Mod. Phys. **76**, 323 (2004); J. H. Stotz, R. Hey, P. V. Santos, and K. H. Ploog, Nat. Mater. **4**, 585 (2005); D. D. Awschalom and M. E. Flatté, Nat. Phys. **3**, 153 (2007).

[12] C. Day, Phys. Today **61**(1), 19 (2008); J. E. Moore, Nature (London) **464**, 194 (2010); X.-L. Qi and S.-C. Zhang, Phys. Today **63**(1), 33 (2010); T. Yokoyama and S. Murakami, Physica E **55**, 1 (2014).

[13] Y. S. Gui, C. R. Becker, N. Dai, J. Liu, Z. J. Qiu, E. G. Novik, M. Schäfer, X. Z. Shu, J. H. Chu, H. Buhmann, and L. W. Molenkamp, Phys. Rev. B **70**, 115328 (2004).

[14] C.-Y. Hou, E.-A. Kim, and C. Chamon, Phys. Rev. Lett. **102**, 076602 (2009).

[15] C.-X. Liu, J. C. Budich, P. Recher, and B. Trauzettel, Phys. Rev. B **83**, 035407 (2011).

[16] A. Ström and H. Johannesson, Phys. Rev. Lett. **102**, 096806 (2009).

[17] G. Dolcetto, S. Barbarino, D. Ferraro, N. Magnoli, and M. Sassetti, Phys. Rev. B **85**, 195138 (2012).

[18] F. Dolcini, Phys. Rev. B **85**, 033306 (2012).

[19] T. L. Schmidt, Phys. Rev. Lett. **107**, 096602 (2011); J.-R. Souquet and P. Simon, Phys. Rev. B **86**, 161410(R) (2012); Y.-W. Lee, Y.-L. Lee, and C.-H. Chung, *ibid*. **86**, 235121 (2012).

[20] F. Dolcini, Phys. Rev. B **83**, 165304 (2011).

[21] R. Citro, F. Romeo, and N. Andrei, Phys. Rev. B **84**, 161301(R) (2011).

[22] J. Nilsson and A. R. Akhmerov, Phys. Rev. B **81**, 205110 (2010).

[23] F. Romeo, R. Citro, D. Ferraro, and M. Sassetti, Phys. Rev. B **86**, 165418 (2012).

[24] P. Virtanen and P. Recher, Phys. Rev. B **83**, 115332 (2011).

[25] B. Rizzo, L. Arrachea, and M. Moskalets, Phys. Rev. B **88**, 155433 (2013).

[26] F. Dolcini and L. Dell'Anna, Phys. Rev. B **78**, 024518 (2008).

[27] M. Büttiker, Phys. Rev. B **46**, 12485 (1992); H. Föster, P. Samuelsson, and M. M. Büttiker, New J. Phys. **9**, 117 (2007).

[28] L. V. Keldysh, Zh. Eksp. Teor. Fiz. **47**, 1515 (1965) [Sov. Phys. JETP **20**, 1018 (1965)]; H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics* (World Scientific, Singapore, 1995).

[29] L. Onsager, Phys. Rev. **38**, 2265 (1931); H. B. G. Casimir, Rev. Mod. Phys. **17**, 343 (1945); M. Büttiker, Phys. Rev. Lett. **57**, 1761 (1986).

[30] T. Ihn, *Semiconductor Nanostructures, Quantum States, and Electronic Transport* (Oxford University Press, Oxford, 2010).

[31] F. Martins, S. Faniel, B. Rosenow, H. Sellier, S. Huant, M. G. Pala, L. Desplanque, X. Wallart, V. Bayot, and B. Hackens, Sci. Rep. **3**, 1416 (2013).

[32] H. Buhmann (private communication).

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

[34] B. Sutherland and D. C. Mattis, Phys. Rev. A **24**, 1194 (1981); Bruce H. J. McKellar and G. J. Stephenson Jr., *ibid*. **36**, 2566 (1987); Phys. Rev. C **35**, 2262 (1987); C. L. Roy, Phys. Rev. A **47**, 3417 (1993).

[35] J. Viana Gomes and N. M. R. Peres, J. Phys.: Condens. Matter **20**, 325221 (2008).

[36] C.-Y. Hou, K. Shtengel, and G. Refael, Phys. Rev. B **88**, 075304 (2013); K. T. Law, P. A. Lee, and T. K. Ng, Phys. Rev. Lett. **103**, 237001 (2009).

[37] Y. Tanaka and N. Nagaosa, Phys. Rev. Lett. **103**, 166403 (2009).

[38] A. Ström, H. Johannesson, and G. I. Japaridze, Phys. Rev. Lett. **104**, 256804 (2010).

[39] F. Crépin, J. C. Budich, F. Dolcini, P. Recher, and B. Trauzettel, Phys. Rev. B **86**, 121106 (2012).

[40] F. Dolcini, H. Grabert, I. Safi, and B. Trauzettel, Phys. Rev. Lett. **91**, 266402 (2003).