Interaction and disorder effects in three-dimensional topological insulator thin films

E. J. König,[1,2] P. M. Ostrovsky,[3,4] I. V. Protopopov,[1,4,5] I. V. Gornyi,[5,6] I. S. Burmistrov,[4] and A. D. Mirlin[1,2,5,7]

[1]*Institut für Theorie der kondensierten Materie, Karlsruhe Institute of Technology, 76128 Karlsruhe, Germany*

[2]*DFG Center for Functional Nanostructures, Karlsruhe Institute of Technology, 76128 Karlsruhe, Germany*

[3]*Max-Planck-Institute for Solid State Research, D-70569 Stuttgart, Germany*

[4]*L. D. Landau Institute for Theoretical Physics RAS, 119334 Moscow, Russia*

[5]*Institut für Nanotechnologie, Karlsruhe Institute of Technology, 76021 Karlsruhe, Germany*

[6]*A. F. Ioffe Physico-Technical Institute, 194021 St. Petersburg, Russia*

[7]*Petersburg Nuclear Physics Institute, 188300 St. Petersburg, Russia*

(Received 13 May 2013; revised manuscript received 5 June 2013; published 8 July 2013)

A theory of combined interference and interaction effects on the diffusive transport properties of 3D topological insulator surface states is developed. We focus on a slab geometry (characteristic for most experiments) and show that interactions between the top and bottom surfaces are important at not too high temperatures. We treat the general case of different surfaces (different carrier densities, uncorrelated disorder, arbitrary dielectric environment, etc.). In order to access the low-energy behavior of the system, we renormalize the interacting diffusive σ model in the one loop approximation. It is shown that intersurface interaction is relevant in the renormalization group (RG) sense and the case of decoupled surfaces is therefore unstable. An analysis of the emerging RG flow yields a rather rich behavior. We discuss realistic experimental scenarios and predict a characteristic nonmonotonic temperature dependence of the conductivity. In the infrared (low-temperature) limit, the system flows into a metallic fixed point. At this point, even initially different surfaces have the same transport properties. Investigating topological effects, we present a local expression of the $\mathbb{Z}_2$ theta term in the sigma model by first deriving the Wess-Zumino-Witten theory for class DIII by means of non-Abelian bosonization and then breaking the symmetry down to AII. This allows us to study a response of the system to an external electromagnetic field. Further, we discuss the difference between the system of Dirac fermions on the top and bottom surfaces of a topological insulator slab and its nontopological counterpart in a double-well structure with strong spin-orbit interaction.

DOI: 10.1103/PhysRevB.88.035106 PACS number(s): 73.23.−b, 73.43.Nq, 73.20.Fz

I. INTRODUCTION

Topological states of matter have recently attracted immense scientific interest, which was, in particular, boosted by the theoretical prediction[1–5] and, subsequent, experimental discovery[6,7] of two-dimensional (2D) and three-dimensional (3D) time-reversal invariant topological insulators.

In their bulk, topological insulators[8–11] (TI) are electronic band insulators characterized by a topological invariant that accounts for the nontrivial structure of the Bloch states. In contrast, the interface between two topologically different phases (e.g., TI-vacuum) hosts gapless, extended boundary states. Their appearance is topologically protected via the bulk-boundary correspondence.[12] In retrospect, we understand that the quantum Hall effect (QHE)[13] at given quantized transverse conductance was the first example of a topological insulator: the Landau levels provide the bulk band gap, which is accompanied by the topological Thouless-Kohmoto-Nightingale-den Nijs (TKNN)[14] number and the protected chiral edge states.

In contrast to the QHE, the newly discovered 2D and 3D topological insulators require the absence of magnetic field and rely on strong spin-orbit interaction. Further, their topological invariant takes only values in $\mathbb{Z}_2$ (contrary to the TKNN integer). The 2D TI phase (also known as quantum spin Hall state) was experimentally identified by the characteristic quantized value $2e^2/h$ of the two-point conductance in HgTe quantum wells.[6] The discriminating feature of all 3D TI is the massless Dirac states on the 2D boundary, which were first spectroscopically detected in BiSb[7] alloys and, subsequently, in many other materials.[8] To present date, various experimental groups confirmed predominant surface-state transport (for a review, see Ref. 15), in particular, elucidating ambipolar field effect[16–20] and the typical QHE steps of Dirac electrons,[21–25] Aharonov-Bohm oscillations[26–28] as well as weak antilocalization (WAL) corrections in the magnetoconductivity data.[29–31] Moreover, several transport experiments reveal the importance of electron-electron interactions in 3D TI materials.[31–33]

Inspired by recent experimental advances, we present here a detailed analysis of interference and interaction corrections to conductivity in the most conventional setup for transport experiments: the slab geometry, in which the 3D TI films are rather thin (down to $\sim$10 nm) although still thick enough to support well separated surface states. As we will explain in more detail, the long-range Coulomb interaction between the two major surfaces plays an important role. We derive the quantum corrections to conductivity in the diffusive regime by taking into consideration the WAL effect as well as corrections of Altshuler-Aronov type[34] induced by inter and intrasurface interactions. We consider the general situation of different surfaces subjected to different random potentials, mismatch in carrier densities, and unequal dielectric environment.

The present paper constitutes a natural extension of the previous work[35] by three of the authors in which a single 3D TI surface was analyzed. It was found that the interplay of topological protection and interaction- and interference-induced conductivity corrections drives the system into a novel

critical state with longitudinal conductance of the order of e^2/h. As we show below, the intersurface interaction in a thin TI slab makes the overall picture much more complex and crucially affects the ultimate infrared behavior.

In another recent paper,[36] two of us were involved in the theoretical investigation of inter and intrawell interaction effects in double quantum well heterostructures studied experimentally in Ref. 37. Let us point out key differences between the present paper and that work. First, in Ref. 36, only equal carrier densities were considered. Second, disorder was assumed to be the same in both quantum wells (and thus completely correlated). This affects the soft-mode content of the low-energy theory. Third, quantum wells host electrons with spin degeneracy, which can be lifted by a magnetic field. As a consequence, (i) electrons in double quantum wells fall into a symmetry class different from that of 3D TI and (ii) more interaction channels have to be included. These subtleties affect in a crucial way the renormalization group (RG) flow: according to the analysis of Ref. 36, the interwell interaction becomes irrelevant at low energies, which is opposite to what we find in the two-surface TI model in the present paper. Finally, the TI problem shows topology-related effects that were absent in the double quantum well structure.

As in the two preceding works, we here use the interacting, diffusive nonlinear sigma model (NLσM) approach to capture the diffusive low-energy physics. Quantum corrections to the longitudinal conductivity σ are obtained by renormalization of this effective action in the one-loop approximation (i.e., perturbatively in $1/\sigma$ but exactly in interaction amplitudes). The interacting NLσM was originally developed by Finkel'stein in the 1980's[38,39] (for review articles, see Refs. 40–42). In addition to perturbative RG treatment (which can also be performed diagrammatically[43]), it also allows one to incorporate topological effects and was thus a fundamental tool for understanding the interplay of disorder and interactions in a variety of physical problems, including the superconducting transition in dirty films,[44,45] the integer QHE,[46,47] and the metal-insulator transition in Si MOSFETs.[48]

Analyzing the RG equations for the thin 3D TI film, we find that (in contrast to the previous work on the double quantum well structure) the intersurface interaction is relevant in the RG sense. The system flows towards a metallic fixed point at which even two originally different surfaces are characterized by the same conductivities. As we discuss in detail below, the hallmark of the intersurface interaction in 3D TI transport experiments is a characteristic nonmonotonic temperature dependence of the conductivity. In contrast to the case of decoupled surfaces, due to the intersurface interaction, quantum corrections to the conductivity depend on the carrier densities.

The paper is structured as follows. In Sec. II, we expose in detail the theoretical implications of a typical experimental slab geometry setup, demonstrate the relevance of intersurface interaction and introduce the microscopic fermionic Hamiltonian. Subsequently (see Sec. III), we use the non-Abelian bosonization technique to map the fermionic theory on the [**U**(1)-] gauged, interacting NLσM with a $\mathbb{Z}_2$ topological term. Here, we also discuss the Fermi liquid treatment of generally strong electron-electron interactions. Next, we renormalize the NLσM in Sec. IV. Sections III and IV contain both pedagogical explanations and important details for experts. Readers purely interested in the results can jump to Sec. V, where the RG flow and the implied phase diagram are analyzed. Detailed predictions for typical experiments can be found in Sec. VI. We close the paper by summarizing our results and discussing prospects for future work in Sec. VII.

II. TOPOLOGICAL INSULATOR SLABS: EXPERIMENTAL SETUP AND THEORETICAL MODEL

A. Setup

In this work, we analyze the effect of interaction on transport properties of strong 3D topological insulator thin films in the diffusive regime. While we mainly focus on the theoretically most interesting case of purely surface transport, we also show that our theory can easily be extended to a case when only a part of the sample is in the topological phase, i.e., one has a conduction through a topologically protected surface spatially separated from a thick (bulk) conducting region.

A typical experimental setup is shown in Fig. 1. Our analysis is valid in the regime where the penetration depth of surface states a is small with respect to the film thickness d. We therefore neglect intersurface tunneling (which would destroy the topological protection). Further, we assume the disorder correlation length (depicted by the range of the impurity potentials) to be small $\xi \ll d$. We treat a generic case when the vicinity to the coat or, respectively, to the substrate may induce a different degree of disorder on the top and bottom surfaces. We thus consider the corresponding mean-free paths l_1 and l_2 as two independent parameters. Moreover, we also allow the chemical potentials μ_1 and μ_2 on the two surfaces to be different. (By convention, we set $\mu_s = 0$ at the Dirac point. Here and below, $s = 1,2$ denotes the surface index.) The chemical potentials may be experimentally controlled by means of electrostatic gates. As has been stated above, we mostly focus on the situation where both μ_1 and μ_2 lie well within the bulk gap Δ_{bulk}. The extension of our results to the experimentally important regime when only one of chemical potentials is located within the bulk gap, $|\mu_1| \ll \Delta_{\text{bulk}} \lesssim |\mu_2|$, can be found in Sec. V B1.

If the electrostatic gates are present and too close[121] to the sample, Coulomb interaction is externally screened and the electron-electron interaction is purely short range. However, such an experimental scenario is a rare exception from the rule. Therefore, in the main text, we assume sufficiently distant

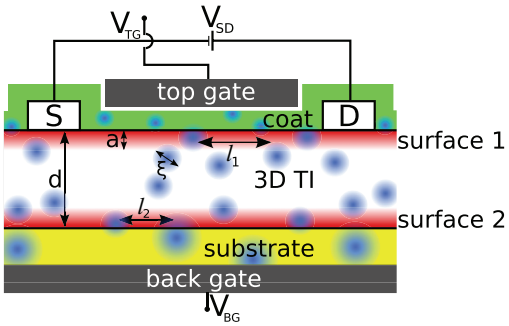

FIG. 1. (Color online) Scheme of a typical experimental setup. The hierarchy of length scales is explained in the main text.

gates and concentrate on the limit of long-range Coulomb interaction. In addition, we derive general RG equations (see Appendix D), which allow us to explore the crossover from the long-range case to the short-range one, see Appendix F. Qualitatively, the RG flow for a sufficiently strong short-range interaction in the case of externally screened surfaces turns out to be similar to the flow in the absence of external screening. Since we assume that the thickness d of the sample is much smaller than its other linear dimensions, we neglect contributions of four side faces of the slab (whose area is proportional to d).

The goal of the present analysis is to study conduction properties of thin 3D TI films in the diffusive regime, i.e., at energy scales E far below the elastic scattering rates $1/\tau_s$ of both surfaces,

$$E \ll \min_{s=1,2} \hbar/\tau_s \,. \tag{1}$$

In turn, the elastic scattering rates are assumed to be small compared to the chemical potentials

$$\hbar/\tau_s \ll |\mu_s|. \tag{2}$$

In experiment, E is set by the ac frequency ($E = \hbar\omega$) or by temperature ($E = k_B T$), whichever of the two is larger. Equation (1) is equivalent to the hierarchy of length scales

$$l \ll L_E \,, \tag{3}$$

where we have introduced the maximal mean-free path $l = \max_{s=1,2} l_s$ and the length scale $L_E = \min_{s=1,2}(\hbar D_s/E)^{1/2}$, with D_s being the diffusion coefficients for the two surfaces.

B. Interaction

Can Coulomb interaction between the top and bottom surface states play an important role in the experiment? To answer this question, we compare the sample thickness with all natural length scales of the system: the screening length $l_{\rm scr}$, the (maximal) mean-free path l and the experimentally tunable scale L_E.

The Coulomb interaction is (throughout the paper underlined symbols denote 2×2 matrices in the surface space)

$$\underline{U}_0(\boldsymbol{r}) = \frac{e^2}{\epsilon} \begin{pmatrix} \frac{1}{r} & \frac{1}{\sqrt{r^2+d^2}} \\ \frac{1}{\sqrt{r^2+d^2}} & \frac{1}{r} \end{pmatrix}. \tag{4}$$

The two-dimensional vector $\boldsymbol{r}$ connects the two-dimensional positions of the particles, $r = |\mathbf{r}|$, e is the charge of the electrons, and ϵ denotes the effective dielectric constant.

Fourier transformation and RPA-screening leads to[36,49–51] [$U \equiv U(\boldsymbol{q}) \equiv 2\pi e^2/\epsilon q$]

$$\underline{U}_{\rm scr}(\boldsymbol{q}) = \frac{\underline{U}}{1 - (\Pi_1 + \Pi_2)\,U + U^2 \Pi_1 \Pi_2 (1 - e^{-2dq})} \tag{5}$$

with

$$\underline{U} = U \begin{pmatrix} 1 - \Pi_2 U (1 - e^{-2dq}) & e^{-dq} \\ e^{-dq} & 1 - \Pi_1 U (1 - e^{-2dq}) \end{pmatrix}.$$

Here, Π_s is the polarization operator of the surface states.

In the present section, we will concentrate on the statically screened interaction potential. In this limit, the polarization operator is determined by the thermodynamic density of states: $\Pi_s(\omega = 0, \boldsymbol{q}) = -\nu_s$.

In the diffusive regime defined by the condition (3), the wave vector q satisfies the inequality $1/L_E \ll q \ll 1/l$. Therefore, in a sample of thickness $d \gg L_E$, we always have $dq \gg 1$ and the two surfaces decouple,

$$\underline{U}_{\rm scr} \overset{d \gg L_E}{=} 2\pi \frac{e^2}{\epsilon} \begin{pmatrix} \frac{1}{q+\kappa_1} & 0 \\ 0 & \frac{1}{q+\kappa_2} \end{pmatrix}, \tag{6}$$

where $\kappa_s = 2\pi e^2 \nu_s/\epsilon$ is the inverse Thomas-Fermi screening length for a single surface s. A universal form of the Altshuler-Aronov correction to conductivity induced by the Coulomb interaction[34,35] arises in the unitary limit when one can neglect q as compared with κ_s in Eq. (6). The unitary limit is achieved if $\kappa_s^{-1} \ll l$ (the meaning of this condition as well as the complementary case are discussed in Sec. III F 3).

In the opposite limit of a small interlayer distance, $d \ll l$, we can approximate $e^{-dq} \approx 1$ in the whole diffusive regime. This implies

$$\underline{U}_{\rm scr} \overset{d \ll l}{=} \frac{2\pi e^2}{\epsilon} \frac{1}{q + \kappa_1 + \kappa_2 + 2d\kappa_1\kappa_2\,(1 - qd)} \times \begin{pmatrix} 1 + 2\kappa_2 d & 1 \\ 1 & 1 + 2\kappa_1 d \end{pmatrix}. \tag{7}$$

At the first glance, it looks as if also negative interaction potential was possible. However, this is not the case as shall be explained in what follows. Depending on the hierarchy of the length scales κ_1^{-1},κ_2^{-1} and d the following scenarios are conceivable.

First, consider $\kappa_s d \ll 1$ for both $s = 1$ and 2. In this case, the q dependence of the interaction potential implies the definition of the coupled layer screening length $l_{\rm scr}$:

$$(\underline{U}_{\rm scr})_{ss'}(\boldsymbol{q}) \sim \frac{1}{q + \kappa_1 + \kappa_2} \;\Rightarrow\; l_{\rm scr} = \frac{1}{\kappa_1 + \kappa_2}. \tag{8}$$

If in addition the condition $l_{\rm scr} \ll l$ is fulfilled, the Coulomb interaction potential (7) becomes "overscreened" (q-independent) for all diffusive momenta $q \ll l^{-1}$.

Second, assume that $\kappa_s d \gg 1$ for at least one surface. Then the q dependence of $\underline{U}_{\rm scr}$ is always negligible and thus the notion of coupled layer screening length is meaningless. It is worthwhile to remark that, as expected, the potential (7) reduces to the decoupled form (6) in the limit when $\kappa_s^{-1} \ll d$ for both surfaces (which also implies that $\kappa_s^{-1} \ll l$).

In this paper, we derive the conductivity corrections in the unitarity limit of q-independent interaction, see Eq. (94). As expected, in the limit of decoupled surfaces, $\kappa_s^{-1} \ll d$, they reproduce the previous result,[35] while whenever $d \ll \kappa_1^{-1}$ or $d \ll \kappa_2^{-1}$ novel conductivity corrections induced by intersurface electron-electron interaction emerge.

Finally, in the intermediate regime $l \ll d \ll L_E$, the scale-dependent conductivity can be obtained by the following two-step RG analysis. First, one integrates the single-surface RG equations starting from the shortest scale l up to the intersurface distance d. After this, one uses the running coupling constants at scale d as starting values for the coupled-surface RG flow and integrates these RG equations up to the scale l_E.

Different regimes discussed above are shown schematically in Fig. 2 in the parameter plane d–κ^{-1}. For simplicity, we

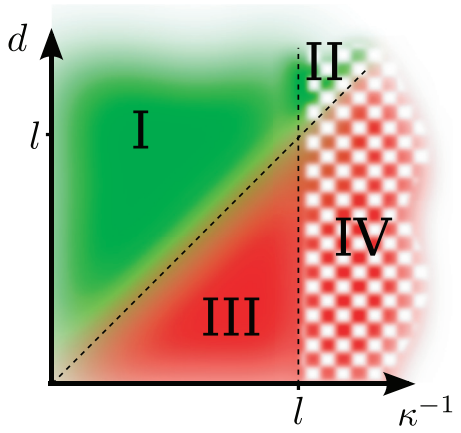

FIG. 2. (Color online) Sketch of the regimes discussed in the main text for the case of comparable screening lengths, $\kappa_1^{-1} \sim \kappa_2^{-1}$ (denoted by κ^{-1}). The regimes I and II correspond to effectively decoupled surfaces (studied in Ref. 35), while in regimes III and IV, intersurface interaction is important. The conductivity corrections in I and III are due to "overscreened" Coulomb interaction. In contrast, in II and IV, this type of corrections sets in only in the low-energy regime where the running length scale (i.e., the typical scale L_E) exceeds the screening length.

assume there the two surfaces have comparable screening lengths: $\kappa_1^{-1} \sim \kappa_2^{-1}$.

In the end of the paper, Sec. VI, we analyze in detail the regions and limits of applicability of our theory with respect to representative experimental setups. In particular, we show that the hierarchy of scales $d \ll l \ll L_E$ is realistic.

In order to illustrate the importance of intersurface interaction (i.e., the relevance of the inequality $d \lesssim \kappa_s^{-1}$) under realistic conditions, we show in Fig. 3 a dependence of the screening length on the Fermi momentum.

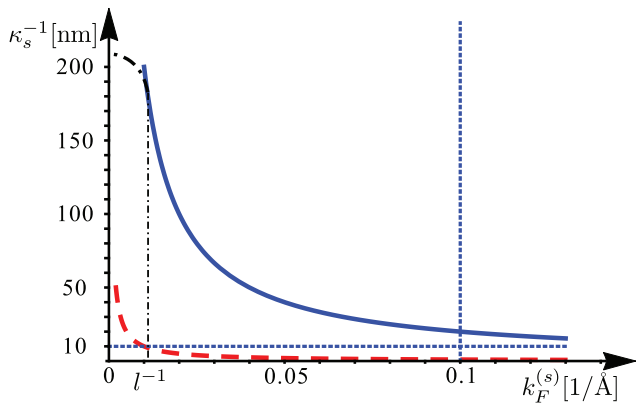

FIG. 3. (Color online) Plot of the single surface screening length κ_s^{-1}. The red curve (large dashes) is the lower bound (corresponding to $\alpha = 1$) of the screening length. The solid, blue curve is the screening length for Bi_2Se_3 film with experimental parameters given in Table I in Sec. VI B. For the latter, the required minimal thickness and maximal Fermi momentum are also depicted (dotted blue lines). The disorder-induced regularization of the divergence at small Fermi momentum is schematically represented by the black dot-dashed curve.

The density of states for the linear (Dirac) spectrum is $\nu(\mu_s) = k_F^{(s)}/2\pi\hbar v_F$, where $k_F^{(s)}$ is the Fermi wave vector of the sth surface state and v_F the Fermi velocity. Therefore

$$\kappa_s^{-1} = \frac{1}{\alpha}\frac{1}{k_F^{(s)}}. \tag{9}$$

We introduced the dimensionless parameter $\alpha = e^2/\epsilon\hbar v_F$, which is the effective coupling constant of the Coulomb interaction and is equal to $c/\epsilon v_F$ times the fine structure constant of quantum electrodynamics. Clearly, α plays the same role as the dimensionless density parameter r_s in conventional theories of electrons in parabolic bands. We will assume that the interaction is not too strong, $\alpha \lesssim 1$, otherwise the system may become unstable, see a discussion at the end of Sec. II C.

The dashed red curve in Fig. 3 represents the lower bound (corresponding to $\alpha = 1$) of κ_s^{-1} as a function of $k_F^{(s)}$. The actual value of κ_s^{-1} for an exemplary case of Bi_2Se_3 (experimental parameters can be found in Table I below) is depicted by the blue solid curve. We see that the screening length can by far exceed the thickness of the topological insulator slab. Indeed, the Bi_2Se_3 experiments[16,31–33] are performed on probes of thickness $d \simeq 1$–100 nm. For this material, our assumption of separate gapless surface states (no tunneling) is both numerically[52] and experimentally[53] shown to be valid down to $d \simeq 10$ nm (blue horizontal dashed line). Thus relevant experimental values of d in the experiments of interest range from $d \simeq 10$ nm up to $d \simeq 100$ nm. On the other hand, surface electrons have a maximal Fermi wave vector of $k_F \sim 0.1/\text{Å}$ associated with $\mu = \Delta_{\text{bulk}} = 0.3$ eV, see blue vertical dashed line. For the lowest concentration, increase of the screening length is limited by disorder. In this way, we estimate the range of κ_s^{-1} as 20–200 nm, so that the condition $\kappa_s^{-1} > d$ can be easily fulfilled. This is particularly the case for relatively thin films ($d \simeq 10$ nm) and in the vicinity of surface Dirac point.

The above analysis proves the relevance of the intersurface electron-electron interaction. In fact, in course of this analysis, we have made several simplifying assumptions that require certain refinements; we list them for the reader's benefit. First, in general, the coating material (ϵ_1), the topological insulator (ϵ_2), and the substrate (ϵ_3) are all dielectrics with different dielectric constants $\epsilon_1 \neq \epsilon_2 \neq \epsilon_3$. In order to determine the exact Coulomb interaction, one has to solve the electrostatic problem of a point charge in such a sandwich structure of dielectrics,[54–56] see Appendix B. Second, the long-range Coulomb interaction is accompanied by short-range contributions, which, in particular, induce corrections to the polarization operator which affect the screening length. More precise calculations taking Fermi liquid corrections into account can be found in Sec. III F 4 and Appendix C. Finally, we neglected the dependence of the Fermi velocity v_F on the chemical potential μ_s, see Sec. II C. However, all these refinements do not modify our conclusion of the importance of interaction between the surface states. We now proceed with a presentation of the field-theoretical formalism that will allow us to explore the problem.

C. Microscopic Hamiltonian

The model under consideration is schematically depicted in Fig. 4. It is described in path integral technique

$$\mathcal{Z} = \int \mathcal{D}[\bar{\psi},\psi]\, e^{-S[\bar{\psi},\psi]} \tag{10}$$

by the following microscopic Matsubara action:

$$S[\bar{\psi},\psi] = \int_{\tau,\boldsymbol{x}} \bar{\psi}(\partial_\tau + H_0 + H_{\rm dis})\psi + S_{\rm int}. \tag{11}$$

The notation $\int_{\tau,\boldsymbol{x}} = \int d^2x \int_0^\beta d\tau$ will be used throughout the article, where, as usual, $\beta = 1/T$ is the inverse temperature. If not specified otherwise, we set Boltzmann's constant, Planck's constant, and the speed of light $k_B = \hbar = c = 1$ in the remainder. The fermionic fields $\bar{\psi}(\boldsymbol{x},\tau) = (\bar{\psi}_1^\uparrow,\bar{\psi}_1^\downarrow,\bar{\psi}_2^\uparrow,\bar{\psi}_2^\downarrow)$ and $\psi(\boldsymbol{x},\tau) = (\psi_1^\uparrow,\psi_1^\downarrow,\psi_2^\uparrow,\psi_2^\downarrow)^T$ describe the spinful $(\uparrow,\downarrow)$ excitations living on surfaces $s = 1$ and $s = 2$. The one-particle Hamiltonian that characterizes the surface s is

$$(H_0 + H_{\rm dis})_s = [V_s(\boldsymbol{x}) - \mu_s] \otimes \mathbf{I}_\sigma + i(-)^s v_F^{(s)} \nabla \wedge \vec{\sigma}, \tag{12}$$

where $\mathbf{I}_\sigma$ is the unit matrix in spin space and we define $\boldsymbol{a} \wedge \boldsymbol{b} = a_x b_y - a_y b_x$. The disorder potentials $V_s(\boldsymbol{x})$ for two surfaces are assumed to be white-noise distributed and uncorrelated:

$$\langle V_s(\boldsymbol{x}) V_{s'}(\boldsymbol{x}')\rangle = \frac{\delta(\boldsymbol{x}-\boldsymbol{x}')\delta_{ss'}}{\pi \nu_s \tau_s}. \tag{13}$$

The disorder strengths $1/\pi\nu_s\tau_s$ may be different for two surfaces.

It is worth emphasizing the following physical implications of this Hamiltonian.

(1) The model (and its analysis below) corresponds to the general case in which the chemical potentials μ_1, μ_2 and hence the carrier densities of the two surfaces may differ.

(2) Since the disorder potentials are different for two surfaces, no intersurface diffuson and cooperon modes will arise. Note that the considered model of fully uncorrelated disorder correctly describes the low-energy physics of the majority of experimental setups, even in the presence of moderate intersurface correlations of disorder. Indeed, any mismatch in chemical potentials and/or disorder configurations leads to an energy gap in the intersurface soft modes. Two physical regimes are conceivable: (i) almost identical surfaces in almost fully correlated random potentials, $|\mu_1 - \mu_2| \ll 1/\tau_s$ and $\langle [V_1(\boldsymbol{x}) - V_2(\boldsymbol{x}')]^2\rangle \ll \sum_{s=1,2}\langle V_s(\boldsymbol{x})V_s(\boldsymbol{x}')\rangle$, and (ii) all other parameter regimes, when at least one of the conditions in (i) is not fulfilled.

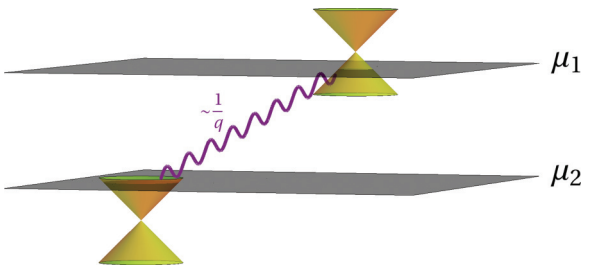

FIG. 4. (Color online) Pictographic representation of the microscopic model: diffusively propagating surface states at different chemical potentials that interact with each other by means of long-range Coulomb interaction.

Our model is designed for the case (ii), where the gap is comparable to the elastic scattering rate and intersurface soft modes do not enter the diffusive theory at all. It also applies to the case (i) in the ultimate large-scale limit (i.e., at energy scales below the gap). In this case, there will be, however, an additional, intermediate regime in the temperature dependence (or ac frequency dependence), which is not considered in our work.

(3) $\vec{\sigma}$ in Eq. (12), in general, does not describe the physical spin. For example, in Bi_2Se_3 structures, the effective spin σ is determined by a linear combination of real spin and the parity (band) degrees of freedom. The mixing angle depends on how the crystal is cut.[57] In this case, also the Fermi velocity becomes anisotropic.

(4) Because of interaction effects, the true dispersion relation is not linear but contains logarithmic corrections (or more generally is subjected to "ballistic" RG[58–60]), which leads to dependence of the Fermi velocity on the chemical potential. This is reflected in the notation $v_F^{(s)} \equiv v_F(\mu_s)$.

(5) Similarly, also the strength of the disorder may be substantially different for both surfaces, so that the (quantum) mean-free times τ_s are considered as two independent input parameters. This is primarily because the vicinity to the substrate or, respectively, to the coating material makes the impurity concentration on both surfaces *a priori* different. In addition, τ_s acquire renormalization corrections, leading to a logarithmic dependence on μ_s.[59,61–63]

(6) The (pseudo-)spin texture on the top and bottom surfaces is opposite [denoted by the factor $(-)^s$].

(7) Finally, in some materials (in particular, in Bi_2Te_3), the Dirac cone is strongly warped. We neglect the warping as it does not affect the main result of this paper, namely, the (universal) RG equations. Recently,[64] it has been shown that warping only influences the dephasing length (i.e., the length scale at which the RG flow is stopped).

The interaction is mediated by the Coulomb potential, see Eq. (4) and Appendix B. With the definition $\rho_s(\tau,\boldsymbol{x}) = \bar{\psi}_s(\tau,\boldsymbol{x})\psi_s(\tau,\boldsymbol{x})$, the corresponding contribution to the action is given by

$$S_{\rm int} = \frac{1}{2}\sum_{ss'}\int_{\tau,\boldsymbol{x},\boldsymbol{x}'} \rho_s(\tau,\boldsymbol{x}) U_{0,ss'}(|\boldsymbol{x}-\boldsymbol{x}'|)\rho_{s'}(\tau,\boldsymbol{x}'). \tag{14}$$

For equal surfaces $[v_F^{(1)} = v_F^{(2)}]$, a simple rescaling of Eqs. (11) and (14) shows that the effective coupling to the Coulomb interaction is α. It can, in general, become of the order of unity. Since the perturbation theory is insufficient in such a case, we adopt the more general, yet phenomenological, Fermi liquid theory to access the behavior for energies down to the elastic scattering rates $\tau_{1,2}^{-1}$ (see Secs. III F3, III F4, and Appendix C). This (clean) Fermi liquid theory will then be a starting point for the interacting diffusive problem at energies below the elastic scattering rate.

If the interaction becomes too strong, it might in principle drive the system into a phase with spontaneously broken symmetry.[65] Examples are the Stoner instability[66] as well as more exotic phenomena such as topological exciton condensation,[67] which is specific to 3D TI thin films.

Throughout our analysis, we assume that the system is not in a vicinity of such an instability. To our knowledge, this assumption is consistent with all transport experiments on 3D TI slabs addressed in this work.

III. σ-MODEL DESCRIPTION

We are interested in the low-energy (low-temperature, long-length-scale) physics of the 3D TI problem defined by Eqs. (1) and (2). This physics is controlled by coupled diffuson and cooperon modes. In this section, we derive the effective field theory—diffusive nonlinear σ model—that describes the system in this regime.

A. Symmetries of the action

The structure of the effective low-energy theory, the diffusive NLσM, is controlled by symmetries of the microscopic action. The information about other microscopic details enters the theory only via the values of the coupling constants. We thus begin by analyzing symmetries of the problem.

First, our system obeys the time reversal symmetry $H = \sigma_y H^T \sigma_y$. Second, we assume no intersurface tunneling, i.e., the particle number is conserved in each surface separately. This implies invariance of the action with respect to $\mathbf{U}(1) \times \mathbf{U}(1)$ transformations (global in space and time).

The presence of Coulomb interaction promotes the $\mathbf{U}(1)$ symmetry in the total-density channel, $\rho_1 + \rho_2$, to transformations that are local in time but global in space. In other words, rotations of fermionic fields, $\bar{\psi}_s(\tau,\boldsymbol{x}) \to \bar{\psi}_s(\tau,\boldsymbol{x}) \exp[-i\chi_s(\tau)]$, $\psi_s(\tau,\boldsymbol{x}) \to \exp[i\chi_s(\tau)]\psi_s(\tau,\boldsymbol{x})$, with equal phases $\chi_1(\tau) = \chi_2(\tau)$ leave the action (11) invariant. This is a special case of "$\mathcal{F}$ invariance"[68] and has important consequences for the present problem. The $\mathcal{F}$ invariance (it is intimately linked to gauge invariance) generally states that in each channel with long-range interaction, time-dependent but spatially constant $\mathbf{U}(1)$ rotations are symmetries of the action. In our problem, as it follows from the $q \to 0$ limit of the Coulomb interaction:

$$\underline{U}(\boldsymbol{q}) \overset{q\to 0}{\propto} \frac{1}{q}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \tag{15}$$

only the interaction between the total densities is long ranged. The structure of Eq. (15) remains true also in the case of asymmetric dielectric environment, see Appendix C 4.

To make the time-reversal symmetry explicit, we define particle-hole bispinors by combining ψ and $\bar{\psi}$ fields.[41,69] In the momentum space, the bispinors read

$$\Phi_n(\boldsymbol{k}) = \frac{1}{\sqrt{2}}\begin{pmatrix} \bar{\psi}_n(-\boldsymbol{k})^T \\ i\sigma_y \psi_n(\boldsymbol{k}) \end{pmatrix} \tag{16}$$

and

$$\bar{\Phi}_n(\boldsymbol{k}) = [C\Phi_n(-\boldsymbol{k})]^T \text{ with } C = i\sigma_y\tau_x, \tag{17}$$

where n is the index associated to the fermionic Matsubara frequency $i\epsilon_n$, and τ matrices act in the particle-hole space. This allows us to rewrite the one-particle Hamiltonian as

$$S^{\text{free}} = -\sum_n \int_{\boldsymbol{k}} \bar{\Phi}_n(\boldsymbol{k})[i\epsilon_n - H^T(-\boldsymbol{k})]\Phi_n(\boldsymbol{k}). \tag{18}$$

It is convenient to perform a rotation of bispinors

$$\eta = \sqrt{\tau_x}\Phi, \tag{19}$$

where $\sqrt{\tau_x} = e^{-i\pi/4}(\mathbf{I}_\tau + i\tau_x)/\sqrt{2}$. The free action then takes the form

$$S^{\text{free}} = -\sum_s \int_{\boldsymbol{x}} \eta_s^T \{[i\hat{\epsilon} - V_s + \mu_s](-i\sigma_y) \tag{20}$$

$$+ (-)^{s+1} v_F^{(s)}(\partial_x - i\partial_y\sigma_z)\}\eta_s. \tag{21}$$

The Matsubara frequency summation is incorporated into the scalar product $\eta^T(\dots)\eta$. In these notations, $\hat{\epsilon}$ is a diagonal matrix in the Matsubara space consisting of entries ϵ_n.

In order to perform the average over disorder, we replicate the theory N_R times. Furthermore, in order to implement the $\mathbf{U}(1)$-gauge invariance in the framework of the NLσM, we apply a double cutoff truncation procedure with $N_M \ll N'_M$ for the Matsubara frequencies.[68] Here, N'_M and N_M are the numbers of retained Matsubara harmonics for fast (electrons of the original theory) and slow (diffusons and cooperons of the NLσM) degrees of freedom, respectively. As a consequence, η becomes a $(2_s \times 2_\sigma \times 2_\tau \times 2N'_M \times N_R)$-dimensional Grassmannian vector field. Except for the frequency term, the free action (21) is manifestly invariant under global orthogonal rotations of the kind

$$\eta_s \to (O_s \otimes \mathbf{I}_\sigma)\eta_s \text{ with } O_s \in \mathbf{O}(2_\tau \times 2N'_M \times N_R). \tag{22}$$

Since the surfaces are fully decoupled in the absence of interactions, the rotations O_1 and O_2 of the fields corresponding to the top and bottom surfaces are completely independent.

B. Quasiclassical conductivity

To obtain the quasiclassical conductivity, we first find the fermionic self-energy within the self-consistent Born approximation (SCBA):

$$\Sigma_n^s = \frac{-2i\sigma_y}{\pi\nu_s\tau_s}\langle \eta_{\boldsymbol{x},s}\eta^T_{\boldsymbol{x},s}\rangle_{\text{SCBA}}. \tag{23}$$

Here, $\langle\dots\rangle_{\text{SCBA}}$ denotes the self-consistent treatment, i.e., a shift $\mu_s \to \mu_s + \Sigma_n^s$ in the fermionic propagator. Equation (23) yields for the imaginary part of the self-energy $\text{Im}(\Sigma_n^s) = (1/2\tau_s)\text{sgn}(\epsilon_n)$. The quasiclassical Drude dc conductance of the noninteracting problem in the absence of a magnetic field is

$$\sigma_s^D = 2\pi\nu_s D_s \frac{e^2}{h}, \tag{24}$$

with $D_s = (v_F^{(s)})^2\tau_s$. Note that the transport time is twice the quantum mean-free time τ_s. In the diagrammatic language, this is a consequence of vertex corrections.

C. Fermionic currents and bosonization rules

To derive the NLσM, we use the method of non-Abelian bosonization.[70–74] An advantage of this approach is that nontrivial topological properties of the Dirac fermions are translated into the field theory in a particularly transparent way.

In the first step, the kinetic term (see Sec. III D) is bosonized. Subsequently, we bosonize also the terms induced by the

chemical potential, disorder, and frequency (see Sec. III E). Since only interaction couples the two surfaces, we omit the surface index s in Sec. III D and Sec. III E. This index is restored later in Sec. III F where the interaction is included.

Local left ($\eta_\uparrow \to O_L\eta_\uparrow$) and right ($\eta_\downarrow \to O_R\eta_\downarrow$) rotations define the left and right currents. The bosonization rules for these currents as well as for the mass term are

$$j_+ = v_F\eta_\uparrow\eta_\uparrow^T \leftrightarrow \frac{1}{8\pi}(O\partial_+O^T), \tag{25a}$$

$$j_- = v_F\eta_\downarrow\eta_\downarrow^T \leftrightarrow \frac{1}{8\pi}(O^T\partial_-O), \tag{25b}$$

$$\eta_\uparrow\eta_\downarrow^T \leftrightarrow i\lambda O, \tag{25c}$$

where $\partial_\pm = \partial_x \pm i\partial_y$. The energy scale λ is of the order of the ultraviolet (UV) cutoff and is introduced here for dimensional reasons (see Secs. III E 1 and IV B for a discussion of its physical meaning). Note that in general, the UV cutoff is different for the top and bottom surfaces, $\lambda_1 \neq \lambda_2$. Further, O is an orthogonal $(2_\tau \times 2N'_M \times N_R) \times (2_\tau \times 2N'_M \times N_R)$ matrix field. Below, we will need the following constant matrices in this space:

$$\begin{aligned}\Lambda^{\tau_1\tau_2;\alpha\beta}_{nm} &= \mathrm{sgn}\,(n)\,\delta^{\tau_1\tau_2}\delta^{\alpha\beta}\delta_{nm},\\ \hat\eta^{\tau_1\tau_2;\alpha\beta}_{nm} &= n\delta^{\tau_1\tau_2}\delta^{\alpha\beta}\delta_{nm},\\ \left(I^{\alpha_0}_{n_0}\right)^{\tau_1\tau_2;\alpha\beta}_{nm} &= \delta^{\tau_1\tau_2}\delta^{\alpha_0\alpha}\delta^{\alpha_0\beta}\delta_{n-m,n_0}.\end{aligned} \tag{26}$$

Here and throughout the paper, we use a convention that $\alpha,\beta \in \{1,\ldots,N_R\}$ denote replicas and $m,n \in \{-N'_M,\ldots,N'_M-1\}$ Matsubara indices. The double cutoff regularization scheme[68] prescribes that matrices O have nontrivial matrix elements O_{nm} only for low-energy excitations $n,m \in \{-N_M,\ldots,N_M-1\}$ and stay equal to the origin O_0 of the σ model manifold outside this low-energy region. As explained below, $O_0 = \Lambda$.

D. Bosonization of the kinetic part

The kinetic part of Eq. (21) is nothing but the Euclidean counterpart of the model considered in Ref. 70. Upon non-Abelian bosonization, it yields the Wess-Zumino-Novikov-Witten (WZNW) action

$$S_{\mathrm{WZNW}} = \int_x \frac{1}{16\pi}\mathrm{tr}\nabla O\nabla O^{-1} + \frac{i}{24\pi}\Gamma_{\mathrm{WZ}}, \tag{27}$$

where Γ_{WZ} is the Wess-Zumino (WZ) term

$$\Gamma_{\mathrm{WZ}} = \int_{x,w}\epsilon_{\mu\nu\rho}\mathrm{tr}[(\tilde O^{-1}\partial_\mu\tilde O)(\tilde O^{-1}\partial_\nu\tilde O)(\tilde O^{-1}\partial_\rho\tilde O)], \tag{28}$$

where $\epsilon_{\mu\nu\rho}$ denotes the Levi-Civita symbol. The definition of the WZ term involves an auxiliary coordinate $w \in [0,1]$ and smooth fields $\tilde O(\boldsymbol{x},w)$ satisfying $\tilde O(\boldsymbol{x},w=0) = \mathrm{const}$ and $\tilde O(\boldsymbol{x},w=1) = O(\boldsymbol{x})$. As a result, the compactified two-dimensional coordinate space $\mathbb{R}^2 \cup \{\infty\} \simeq \mathbb{S}^2$ is promoted to the solid 3-ball $\mathbb{B}^3$ (i.e., the "filled" sphere).

E. Free NLσM of class AII

1. Disorder, frequency, and the chemical potential

The action (27) is the bosonized counterpart of the second (proportional to velocity) term of the microscopic action (21). Let us now consider the first term in Eq. (21), which carries information about the chemical potentials, frequency, and random potential.

Bosonization of the terms with frequency and the chemical potential in the microscopic action (21) yields

$$\delta S = 2\int_x \mathrm{tr}[(i\hat\epsilon+\mu)\eta_\uparrow\eta_\downarrow^T] \leftrightarrow -2\lambda\int_x \mathrm{tr}(\hat\epsilon - i\mu)O. \tag{29}$$

Upon disorder averaging and bosonization, the term with random potential provides the following contribution to the field theory:

$$\begin{aligned}\delta S_{\mathrm{dis}} &= -\frac{1}{\pi\nu\tau}\int_x(\mathrm{tr}\eta_\uparrow\eta_\downarrow^T)^2 + \frac{1}{\pi\nu\tau}\int_x \mathrm{tr}(\eta_\uparrow\eta_\downarrow^T)^2\\ &\leftrightarrow \frac{\lambda^2}{\pi\nu\tau}\int_x(\mathrm{tr}O)^2\\ &\quad+\frac{\lambda^2}{2\pi\nu\tau}\int_x \mathrm{tr}(O^T-O)^T(O^T-O).\end{aligned} \tag{30}$$

As we see, disorder induces mass terms for O matrices. Both mass terms in Eq. (30) are strictly non-negative. Therefore they are minimized by arbitrary traceless symmetric orthogonal matrix. It is convenient to choose the specific saddle-point solution as

$$O = \Lambda. \tag{31}$$

This saddle-point solution coincides with the SCBA. Indeed, Eq. (23) can be written as

$$\begin{aligned}\frac{i}{2\tau}\Lambda\otimes\mathbf{1}_\sigma &= \frac{2}{\pi\nu\tau}\left\langle\begin{pmatrix}-\eta_\downarrow\eta_\uparrow^T & -\eta_\downarrow\eta_\downarrow^T\\ \eta_\uparrow\eta_\uparrow^T & \eta_\uparrow\eta_\downarrow^T\end{pmatrix}\right\rangle_{\mathrm{SCBA}}\\ &\leftrightarrow \frac{2}{\pi\nu\tau}\left\langle\begin{pmatrix}i\lambda O^T & \frac{-1}{8\pi v_F}O^T\partial_-O\\ \frac{1}{8\pi v_F}O\partial_+O^T & i\lambda O\end{pmatrix}\right\rangle.\end{aligned} \tag{32}$$

It is solved by the saddle-point solution (31) provided the auxiliary UV energy scale λ introduced in Eq. (25) is related to the density of states (i.e., to the chemical potential),

$$\lambda = \frac{\pi\nu}{4} = \frac{|\mu|}{8v_F^2}. \tag{33}$$

We will rederive this relation from a different viewpoint below, see Sec. IV B.

Equation (31) is not the only solution of the saddle point equation. It is easy to see that rotations

$$O \to O_{\mathrm{soft}}^T O\, O_{\mathrm{soft}},\ O_{\mathrm{soft}} \in \mathbf{G} = \mathbf{O}(2_\tau \times 2N_M \times N_R) \tag{34}$$

leave the mass term unaffected. On the other hand, the saddle-point $O = \Lambda$ is invariant under rotations from a smaller group, $O_{\mathrm{soft}} \in \mathbf{K} = \mathbf{O}(2_\tau \times N_M \times N_R) \times \mathbf{O}(2_\tau \times N_M \times N_R)$. This can be understood as a breakdown of symmetry $\mathbf{G} \to \mathbf{K}$. We thus obtain a nontrivial manifold of saddle-points annihilating the mass term. Allowing for a slow variation of O_{soft} and restricting other terms in the action to this manifold, we will obtain the NLσM action.

2. Free NLσM with $\mathbb{Z}_2$ topological term

As we have just discussed, we keep only the soft modes

$$Q = O_{\mathrm{soft}}^T\Lambda O_{\mathrm{soft}} \text{ with } O_{\mathrm{soft}} \in \mathbf{G}. \tag{35}$$

The subscript "soft" will be omitted in the remainder. The NLσM manifold $\mathcal{M} = \mathbf{G}/\mathbf{K}$. We also rename the coupling constants according to the conventional notation of diffusive NLσMs and restore the surface index s,

$$S^{\text{free}} = \sum_s \int_x \frac{\sigma_s}{16} \text{tr}\,(\nabla Q_s)^2 - 2\pi T z_s \text{tr}\,(\hat{\eta} Q_s) + i S_s^{(\theta)}. \quad (36)$$

As will become clear from linear response theory (see Sec. III G 3), σ_s measures the dc conductivity of surface s (in units e^2/h). Its bare value is the Drude conductance depending on the chemical potential μ_s, as can be directly verified (see Appendix A 1). The coupling constants z_s determine the renormalization of the specific heat.

The nontrivial second homotopy group of the NLσM manifold $\pi_2(\mathcal{M}) = \mathbb{Z}_2$ allows for topological excitations (instantons), similarly to the QHE theory. A crucial difference is that in the QHE case the second homotopy group is $\mathbb{Z}$, so that any integer topological charge (number of instantons) is allowed. Contrary to this, in the present case, any configuration of an even number of instantons can be continuously deformed to the trivial, constant vacuum configuration. Therefore the theta term $S_s^{(\theta)}$ appearing in Eq. (36) only distinguishes between an even ($S_s^{(\theta)} = 0 \mod 2\pi$) and odd ($S_s^{(\theta)} = \pi \mod 2\pi$) numbers of instantons.

Such a $\mathbb{Z}_2$ theta term $S^{(\theta)}$ does not appear in the case of usual metals with strong spin-orbit coupling; it results from the Dirac-fermion nature of carriers and is a hallmark of topologically protected metals (in our case, the surface of a topological insulator). The topological term flips the sign of the instanton effects (as compared to the case of a usual metal with spin-orbit interaction) from localizing to delocalizing. Thus the theta term translates the protection against Anderson localization into the NLσM approach.

We are now going to show that $S_s^{(\theta)}$ is nothing but the WZ term (obtained from non-Abelian bosonization) restricted to the smaller symmetry group:

$$S_s^{(\theta)} = \frac{1}{24\pi} \Gamma_{\text{WZ},s}|_{\tilde{O}_s(\boldsymbol{x},w=1)=Q_s(\boldsymbol{x})=Q_s^T(\boldsymbol{x})}. \quad (37)$$

Note that, since the second homotopy group of the NLσM manifold is nontrivial, the definition of the WZ term requires that away from $w = 1$ the extended fields can take values in the big orthogonal group $\mathbf{G}$.

To show that Eq. (37) is indeed the $\mathbb{Z}_2$ theta term, we proceed in the same way as was recently done for the symmetry class CII.[75] First of all, it is straightforward to check that $S_s^{(\theta)}$ is invariant under small variations of the σ-model field, $Q_s \to Q_s' = Q_s + \delta Q_s$ ($Q_s'^2 = \mathbf{1} = Q_s^2$). Thus $S_s^{(\theta)}$ only depends on the topology of the field configuration. This immediately implies that it is zero in the topologically trivial sector. In order to proof that $S_s^{(\theta)}$ also returns the correct value $S_s^{(\theta)} = \pi \mod 2\pi$ in the topologically nontrivial sector, it is sufficient to insert a single instanton into $S_s^{(\theta)}$. Instantons are field configurations that per definition can not be continuously deformed into the vacuum configuration. Introducing the third dimension and allowing the field to take values in the entire orthogonal group we can continuously shrink the instanton in the $w = 1$ sphere to the constant at $w = 0$. A necessary condition for this untwisting to happen is that for some subinterval of (0,1) the field leaves the NLσM manifold for the larger orthogonal group. A direct calculation shows that the group volume covered while untwisting indeed yields the value $iS_s^{(\theta)} = i\pi$ (see Appendix A 2).

There have been alternative derivations of the $\mathbb{Z}_2$ term before.[76,77] Viewing this theta term as a symmetry-broken WZ term, Eq. (37) yields a local expression for it and implies the following advantages. First, this form is very useful for understanding the crossover between 3D topological insulators of class DIII and AII. Second and more importantly, an analysis of the response of the system to an external electric field requires coupling of the diffusive matter fields to $\mathbf{U}(1)$ gauge potentials. In particular, one should gauge the topological term, which can be done in a standard way by using a local expression for it. We will show in Sec. III G4 that such a procedure yields the correct linear response theory for the anomalous quantum Hall effect of Dirac fermions.

In addition to a nontrivial second homotopy group π_2, the sigma model manifold of the class AII possesses also a nontrivial first homotopy group, $\pi_1(\mathcal{M}) = \mathbb{Z}_2$. For this reason, the RG flow in 2D systems of class AII (as well as in other classes with a nontrivial π_1 group, namely, AIII, BDI, CII, and DIII) is affected by vortices, as was shown in Ref. 75. In the case of AII (and DIII) class these are $\mathbb{Z}_2$ vortices,[75] i.e., a vortex is identical to an antivortex. In a recent work,[78] it was argued that such vortices are crucial for establishing localization in the class AII. Conversely, the robustness of a nonlocalized state on the surface of a weak topological insulator and of the critical state separating 2D trivial and topological insulators were explained by vanishing of the corresponding fugacity.

On the surface of a strong 3D TI, the effect of vortices is erased by the $\mathbb{Z}_2$ topological term, in the same way as argued previously[75] for the case of the symmetry class CII. Specifically, due to the $\mathbb{Z}_2$ theta term, the vortices acquire an internal degree of freedom which, upon averaging, annihilates the contribution of vortices to renormalization. For this reason, the vortices need not be taken into account in the present context.

F. Interacting NLσM

In the previous section, we have derived the diffusive nonlinear σ model for noninteracting particles. The next step is to include the electron-electron interactions.

1. *Interacting Fermi gas*

We concentrate first on the case of a weak Coulomb interaction ($\alpha \ll 1$). At length scales larger than the screening length, the interaction is effectively pointlike:

$$S_{\text{int}} = \frac{T}{2} \sum_{m,\alpha;ss'} \int_x \text{tr}\big(I_m^\alpha \psi_s \bar{\psi}_s\big) U_{ss'}^q \big(I_{-m}^\alpha \psi_{s'} \bar{\psi}_{s'}\big), \quad (38)$$

where $U_{ss'}^q$ is the "overscreened" Coulomb interaction matrix, i.e., the $q \to 0$ limit of Eq. (7) (for its generalization in case of an asymmetric dielectric environment, see Appendix B). We use the bosonization rule

$$\begin{aligned} \text{tr} I_m^\alpha \psi_s \bar{\psi}_s &= \text{tr} I_m^\alpha (1-\tau_y)\eta_{s,\uparrow}\eta_{s,\downarrow}^T - \text{tr} I_m^\alpha (1-\tau_y)\eta_{s,\downarrow}\eta_{s,\uparrow}^T \\ &\to i\lambda \big[\text{tr} I_m^\alpha (1-\tau_y)\big(O_s + O_s^T\big)\big]. \end{aligned} \quad (39)$$

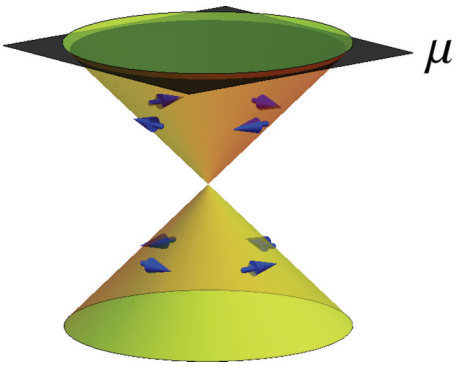

FIG. 5. (Color online) Schematic representation of the Dirac cone and the strong Rashba spin-orbit coupling. If the chemical potential (black plane) is large compared to the typical energy scale E (e.g., temperature), only one kind of helical states can take part in the dynamics.

When disorder is introduced, the matrices O become restricted to the σ-model manifold $\mathcal{M}$, and we obtain

$$S_{\rm int} = -\lambda^2 8T \sum_{m,\alpha;ss'} \int_{\boldsymbol{x}} {\rm tr}\big[J^{\alpha}_{-m} Q_s\big] U^{q}_{ss'} {\rm tr}\big[J^{\alpha}_{m} Q_{s'}\big]. \quad (40)$$

Here, we have defined $J^{\alpha}_{n} = I^{\alpha}_{n} \frac{1+\tau_y}{2}$. As has been already emphasized, we want to treat the general case of strong interactions up to $\alpha \sim 1$. Therefore, in the following (and in more detail in Appendix C), we present the Fermi liquid (FL) treatment of strongly interacting surface states of a thin 3D TI film.

2. Effective spinless theory

One of the most striking peculiarities of the surface states of 3D topological insulators is their Rashba-like kinetic term. As a consequence, spin and momentum are locked in a manner visualized in Fig. 5. Such states are called helical; one associates helicity eigenvalues $+1$ (-1) with states with positive (respectively, negative) kinetic energy. As has been stated above, we will be interested in the low-energy regime $E \ll |\mu_{1,2}|$. Hence, at each of the surfaces, only one type of helical states represents dynamical low-energy degrees of freedom, while the other one is suppressed by a mass $\approx 2|\mu_{1,2}|$. Therefore we project onto the appropriate helicity eigenstate of each surface using the following projection operator:

$$\mathcal{P}_s = |\mu_s, \boldsymbol{p}\rangle\langle\mu_s, \boldsymbol{p}| \text{ with } |\mu_s, \boldsymbol{p}\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i\,{\rm sgn}\mu_s\, e^{i\phi(\boldsymbol{p})} \end{pmatrix}, \quad (41)$$

where we have defined the polar angle ϕ of the momentum, $p_x \equiv |\boldsymbol{p}| \cos\phi$ and $p_y \equiv |\boldsymbol{p}| \sin\phi$. The clean single-particle action becomes effectively spinless:

$$S^{(s)}_0 = -\sum_s \int_{\boldsymbol{p}} \bar{\zeta}_s(\boldsymbol{p}) \big[i\epsilon + {\rm sgn}(\mu_s)\big(|\mu_s| - v^s_F |\boldsymbol{p}|\big)\big] \zeta_s(\boldsymbol{p}), \quad (42)$$

where ζ_s, $\bar{\zeta}_s$ are the fields associated with the helicity eigenstates, $\zeta_s = \langle \mu_s, \boldsymbol{p}|\psi_s$ and $\bar{\zeta}_s = \bar{\psi}_{s,\sigma}|\mu_s, \boldsymbol{p}\rangle$.

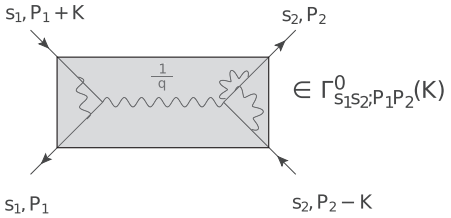

FIG. 6. An example of contribution to a one Coulomb-line reducible small angle scattering amplitude. Independently of ${\rm sgn}(\mu_s)$, ingoing arrows denote fields ζ_s, outgoing arrows $\bar{\zeta}_s$.

3. Scattering channels

In the presence of a Fermi surface, the electron-electron interaction at low energies decouples into separate scattering channels defined by small energy-momentum transfer and by the tensor structure in the surface space:

$$S_{\rm int} = -\frac{T}{2} \int_{P_1,P_2,K} \sum_{\alpha} \big[\mathcal{O}^{IA}_{0+1} + \mathcal{O}^{IA}_{2} + \mathcal{O}^{IA}_{c}\big] \quad (43)$$

with

$$\mathcal{O}^{IA}_{0+1} = \sum_{s_1 s_2} \big[\bar{\xi}^{\alpha}_{s_1}(P_1)\zeta^{\alpha}_{s_1}(P_1+K)\big] \times \Gamma^{0+1,q}_{s_1,s_2;\hat{p}_1,\hat{p}_2} \big[\bar{\xi}^{\alpha}_{s_2}(P_2)\zeta^{\alpha}_{s_2}(P_2-K)\big], \quad (44)$$

$$\mathcal{O}^{IA}_{2} = \sum_{s_1 s_2} \big[\bar{\xi}^{\alpha}_{s_1}(P_2)\zeta^{\alpha}_{s_1}(P_1+K)\big] \times \Gamma^{2,q}_{s_1,s_2;\hat{p}_1,\hat{p}_2} \big[\bar{\xi}^{\alpha}_{s_2}(P_1)\zeta^{\alpha}_{s_2}(P_2-K)\big], \quad (45)$$

and

$$\mathcal{O}^{IA}_{c} = \sum_{s_1 s_2} \big[\bar{\xi}^{\alpha}_{s_1}(P_2)\zeta^{\alpha}_{s_1}(-P_1+K)\big] \times \Gamma^{c,q}_{s_1,s_2;\hat{p}_1,\hat{p}_2} \big[\bar{\xi}^{\alpha}_{s_2}(-P_2+K)\zeta^{\alpha}_{s_2}(P_1)\big]. \quad (46)$$

Here, the capital letters denote $2+1$ momenta. The smallness of $K = (\omega_m, \boldsymbol{q})$ means that the following conditions hold $(\omega_m, |\boldsymbol{q}|) \ll (|\mu_s|, p^{(s)}_F)$ for both $s = 1,2$. We emphasize that all "Dirac factors" of 3D surface electrons are included in the angular dependence of the scattering amplitudes (subscripts $\Gamma_{\hat{p}_1,\hat{p}_2}$).

We refer to the three scattering channels as small angle scattering channel (Γ^{0+1}), large angle scattering channel (Γ^{2}), and the Cooper channel (Γ^{c}). The quantities entering Eq. (43) are the static limit of the corresponding scattering amplitude, $\Gamma(\omega_m = 0, \boldsymbol{q})$. They already include static screening and do not acquire any tree-level corrections due to disorder.[40,42] Exemplary diagrams are given in Figs. 6–9. There, the small

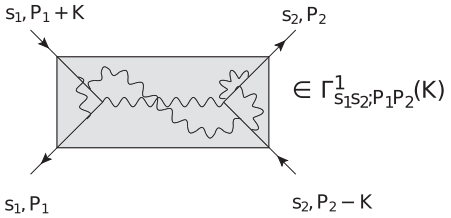

FIG. 7. An example of contribution to a one-Coulomb-line irreducible small-angle scattering amplitude.

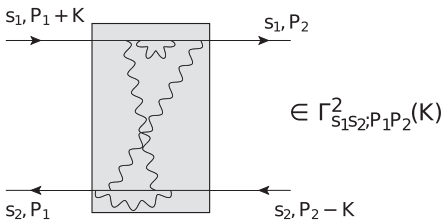

FIG. 8. An example of contribution to a large-angle scattering amplitude.

angle scattering amplitude is subdivided into its one Coulomb line reducible (Γ^0) and irreducible (Γ^1) parts such that

$$\Gamma^{0+1} = \Gamma^0 + \Gamma^1. \tag{47}$$

The irreducible part Γ^1 also includes the short-range interaction induced by the finite thickness of the 3D TI film (see Appendixes B and C 6).

For the short-range interaction amplitudes (Γ^1, Γ^2, Γ^c), the static limit coincides with the "q-limit" $\Gamma^q = \lim_{q\to 0}\Gamma(\omega_m = 0, \boldsymbol{q})$, see also Appendix C. It should be kept in mind that for the one-Coulomb-line-reducible part Γ^0 (it is long-ranged), the "q-limit" $\Gamma^{0,q}$ is only a valid approximation if the mean-free path l exceeds the screening length. This applies to most realistic situations. (In the opposite case, Γ^0 is parametrically small. On top of this, the q dependence of the Coulomb potential implies a strong scale dependence of both conductivity corrections and the interaction amplitude until the running scale reaches the screening length at which $\Gamma^0 \approx \Gamma^{0,q}$ is again justified.)

We conclude this section with a side remark concerning the topological exciton condensation.[67] In order to find the conventional pole structure of the FL Green's functions for the case $\mathrm{sgn}(\mu_s) = -1$, one needs to transpose the bilinear form in action (42) and swap the notation $\zeta_s(\epsilon_n) \leftrightarrow \bar{\zeta}_s(-\epsilon_n)$. If $\mathrm{sgn}(\mu_1\mu_2) = -1$, this interchange of notations obviously happens in only one surface. In this case, the large-angle scattering amplitude Γ^2_{12} and the Cooper-channel amplitude Γ^c_{12} are interchanged. Even though this procedure illustrates the analogy between exciton condensation (divergence in Γ^2_{12}) and Cooper instability (divergence in Γ^c_{12}), in the following, we choose to keep our original notation of ζ_s and $\bar{\zeta}_s$ also in the case of $\mu_s < 0$.

4. *Clean Fermi liquid theory*

A systematic treatment of the scattering amplitudes involves the field theory of the FL[79–81] (see Appendix C). It is valid down to energy scales $\sim\tau_{1,2}^{-1}$ and therefore constitutes

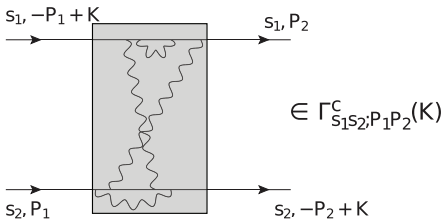

FIG. 9. An example of contribution to a scattering amplitude in the Cooper channel.

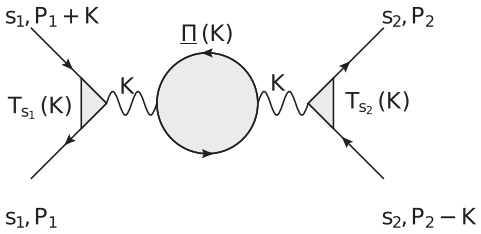

FIG. 10. A diagram contributing to Γ^0.

the starting point for the effective diffusive theory at lower energies, $T \ll \tau_{1,2}^{-1}$.

In contrast to the Green's function of the free theory, in the FL, the exact electronic propagator contains both a singular and a regular part. The singular part ("quasiparticle pole") includes a renormalized dispersion relation and its residue is no more equal to unity but rather is $a_s \in (0,1)$. As usual in the context of disordered FLs,[40] we absorb the quasiparticle residue by rescaling the fermionic fields and redefining the scattering amplitude.

The conservation of the particle number separately in each of the two surfaces leads to the following Ward identities:

$$\Pi^{\omega}_{s_1,s_2} \equiv \lim_{\omega_m\to 0}\Pi_{s_1,s_2}(\omega_m, \boldsymbol{q} = 0) = 0 \tag{48}$$

and

$$\Pi^{q}_{s_1,s_2} \equiv \lim_{|\boldsymbol{q}|\to 0}\Pi_{s_1,s_2}(\omega_m = 0, \boldsymbol{q}) = -\frac{\partial N_{s_1}}{\partial \mu_{s_2}}. \tag{49}$$

Since these identities reflect the gauge invariance, they can not be altered during the RG procedure. Thus the static polarization operator is always given by the compressibility $\partial N_{s_1}/\partial\mu_{s_2}$.

The FL theory in a restricted sense contains only short-range interactions Γ^1, Γ^2, and Γ^c. For electrons in metals, one has also to include the long-range Coulomb interaction. Following Ref. 79, the associated scattering amplitude Γ^0 is obtained by means of static RPA-screening of Coulomb interaction with the help of the FL renormalized polarization operator and triangular vertices (see Fig. 10). In Appendix C, we explicitly perform the formal FL treatment. This determines the interaction amplitudes at ballistic scales. They will serve as bare coupling constants of the diffusive NLσM (see Sec. III F 7). We now turn our attention to the disordered FL. This will allow us to find out which of the interaction channels give rise to soft modes within our problem.

5. *Diffusive Fermi liquid theory*

The full amplitudes $\Gamma^{0+1}(K)$, $\Gamma^2(K)$, and $\Gamma^c(K)$ contain, among others, diagrams describing multiple particle-hole (in the Cooper channel, particle-particle) scattering (see Appendix C). The very idea of dirty FL lies in replacing the dynamic part of these particle-hole (particle-particle) sections by their diffusive counterpart.[40,42] In particular, only the zeroth angular harmonic of the scattering amplitudes survives in the diffusive limit.

The scattering amplitude Γ^2_{12} (as well as Γ^c_{12}) contains only particle-hole (respectively, particle-particle) sections consisting of modes from opposite surfaces of the topological insulator. Since we assume the disorder to be uncorrelated between the surfaces, these modes will not become diffusive

and are hence not of interest for the present investigation. We therefore do not consider Γ^2_{12} and Γ^c_{12} any longer. As one can see from Figs. 6–8, the large angle scattering amplitudes Γ^2_{11} and Γ^2_{22} cannot be distinguished from the small angle scattering amplitudes Γ^{0+1}_{11} and Γ^{0+1}_{22}, respectively. We incorporate the effect of Γ^2_{11} and Γ^2_{22} into the "singlet channel," which has the following matrix structure in the surface space:

$$\underline{\Gamma}^{\rho} = \begin{pmatrix} \Gamma^{0+1-2}_{11} & \Gamma^{0+1}_{12} \\ \Gamma^{0+1}_{12} & \Gamma^{0+1-2}_{22} \end{pmatrix}. \tag{50}$$

Here, we used

$$\Gamma^{0+1-2} = \Gamma^{0+1} - \Gamma^2. \tag{51}$$

The intrasurface Cooper channel interaction Γ^c_{ss} will be also neglected. Its bare value is repulsive for the Coulomb interaction, so that the Cooper renormalization on ballistic scales $1/\tau \ll E \ll |\mu|$ renders it small on the UV scale of the diffusive theory (i.e., at the mean-free path). Within the diffusive RG of a single 3D TI surface it quickly becomes of the order of $1/\sqrt{\sigma}$ and thus negligible (see Ref. 40 and supplementary material of Ref. 35). Consequently, we drop the Cooper channel amplitude and do not consider the superconductive instability in this work.[122] For the opposite case of attraction in the Cooper channel, Coulomb interaction suppresses the transition temperature T_c.[44] The difference between Coulomb and short-range repulsive interaction was addressed in Ref. 83.

6. Bosonization of Fermi Liquid

The non-Abelian bosonization relies on the Dirac nature of the 2D electrons and on the associated non-Abelian anomaly. On the other hand, for $\alpha \sim 1$, the spectrum of the system gets strongly renormalized by interactions. An appropriate description in such a situation is the FL theory, which is restricted to fermionic excitations close to the Fermi level. So, one can ask whether the result of non-Abelian bosonization remains applicable for $\alpha \sim 1$. The answer is yes, for the following reasons. All terms of the bosonized theory except for the $\mathbb{Z}_2$ theta term are determined by fermionic excitations close to the Fermi energy. Therefore they equally hold for the FL if the coupling constants are appropriately redefined in terms of the corresponding FL parameters.

On the other hand, the $\mathbb{Z}_2$ theta term is a consequence of the chiral anomaly and thus the only term determined by energies far from μ. However, it is well known that anomalies in quantum field theories are insensitive to interactions. Hence, the $\mathbb{Z}_2$ term in the diffusive NLσM persists even for $\alpha \sim 1$. This follows also from the key property of the FL state: its spectrum is adiabatically connected to the free spectrum. This implies that topological implications remain unchanged. To summarize, the only difference between the NLσM for the weakly interacting Fermi gas ($\alpha \ll 1$) and the FL ($\alpha \sim 1$) is the replacement of the interaction strength by the appropriate FL constant,

$$\underline{U}^q \to -\underline{\Gamma}^{\rho}$$

in Eq. (40).

7. Bare value of scattering amplitudes

According to the formal FL treatment (see Appendix C 4), the singlet-channel interaction amplitude is given by

$$\underline{\nu}\underline{\Gamma}^{\rho}\underline{\nu} = -\underline{\nu} - \frac{\det \underline{\Pi}^q}{\Pi^q_{11} + \Pi^q_{22} + 2\Pi^q_{12}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \tag{52}$$

where $(\underline{\nu})_{ss'} = \nu_s \delta_{ss'}$ and

$$\underline{\Pi}^q = -\underline{\nu} - \underline{\nu} \begin{pmatrix} \Gamma^{1-2}_{11} & \Gamma^1_{12} \\ \Gamma^1_{12} & \Gamma^{1-2}_{22} \end{pmatrix} \underline{\nu}. \tag{53}$$

Here, $\Gamma^{1-2} = \Gamma^1 - \Gamma^2$. The remarkably simple matrix structure of $\underline{\nu} + \underline{\nu}\underline{\Gamma}^{\rho}\underline{\nu}$ is actually due to the presence of the long-range Coulomb interaction. This fact will be explained by means of $\mathcal{F}$ invariance in Sec. III G 2. It has very important consequences for the RG flow in the diffusive regime, see Sec. IV B.

8. Action of NLσM

We are now in a position to present the full action of the diffusive interacting NLσM for the problem under consideration:

$$S = \sum_s \left[S^{(\mathrm{kin})}_s + i S^{(\theta)}_s \right] + S^{(\eta + \mathrm{int})}. \tag{54}$$

It contains the kinetic term

$$S^{(\mathrm{kin})}_s = \frac{\sigma_s}{16} \int_{\boldsymbol{x}} \mathrm{tr}\, (\nabla Q_s)^2 \tag{55}$$

and the $\mathbb{Z}_2$ theta term

$$S^{(\theta)}_s = \frac{1}{24\pi}\, \Gamma_s |_{\tilde{O}_s(\boldsymbol{x}, w=1) = Q_s(\boldsymbol{x}) = Q^T_s(\boldsymbol{x})} \tag{56}$$

for each of the surfaces, as well as the frequency and interaction terms,

$$S^{(\eta + \mathrm{int})} = -\pi T \left[\sum_s 2 z_s \mathrm{tr} \hat{\eta} Q_s - \sum_{ss';n,\alpha} \mathrm{tr}\left(J^{\alpha}_n Q_s\right) \Gamma_{ss'} \mathrm{tr}\left(J^{\alpha}_{-n} Q_{s'}\right) \right]. \tag{57}$$

Here, we have introduced the notation

$$\Gamma_{ss'} = \frac{8}{\pi} \lambda_s \Gamma^{\rho}_{ss'} \lambda_{s'}. \tag{58}$$

G. Inclusion of scalar and vector potentials into the NLσM

In this section, we investigate consequences of the gauge invariance for the interacting NLσM.

1. Electromagnetic gauge invariance

We include the scalar potential Φ_s and the vector potential $A_{\mu,s}$ for surface s in the microscopic action (11) by means of covariant derivatives. This makes the action gauge invariant, i.e., unchanged under local **U**(1) rotations of the fermionic fields ψ and $\bar{\psi}$ accompanied by the corresponding gauge transformation of the potentials. Note that locality implies independent rotations on the top and bottom surfaces of the TI film.

The rotations of ψ fields imply the following rotation of bispinors:

$$\eta_s(\boldsymbol{x}) \to W_s \eta_s(\boldsymbol{x}), \tag{59}$$

where

$$W_s = \left(e^{-i\hat{\chi}_s^T} \frac{1+\tau_y}{2} + e^{i\hat{\chi}_s} \frac{1-\tau_y}{2} \right), \tag{60}$$

and we use the following convention for hatted matrices: $\hat{a} \equiv \sum_{n,\alpha} a_n^\alpha I_n^\alpha$. Let us recall that the η_s fields are considered as vectors in the Matsubara space. Upon introducing replica indices in the theory, the **U**(1) rotation angles and correspondingly the gauge potentials get replicated as well.

2. $\mathcal{F}$ algebra and $\mathcal{F}$ invariance

As a direct consequence of Eq. (59), Q matrices transform under a gauge transformation χ_s in the following way:

$$Q_s \to W_s Q_s W_s^T. \tag{61}$$

Under such rotations, in the limit $N_M', N_M \to \infty$, $N_M/N_M' \to 0$, the frequency term acquires the correction[68]

$$\delta_\chi \mathrm{tr}\hat{\eta} Q_s = 2 \sum_{n,\alpha} \left(in \chi_{s,n}^\alpha \mathrm{tr} J_{-n}^\alpha Q_s - n^2 \chi_{s,n}^\alpha \chi_{s,-n}^\alpha \right), \tag{62}$$

while the factors entering the interaction term vary as follows:

$$\delta_\chi \mathrm{tr} J_n^\alpha Q_s = -i 2n \chi_{s,n}^\alpha. \tag{63}$$

As explained in Sec. III A, the presence of the Coulomb interaction implies invariance of the fermionic action (11) under a simultaneous rotation in both surfaces by the same spatially constant ("global") but time-dependent **U**(1) phase even *without* inclusion of gauge potentials ("$\mathcal{F}$ invariance"). This symmetry has to be preserved on NLσM level, implying that

$$(\underline{z} + \underline{\Gamma}) \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0. \tag{64}$$

Here, $(\underline{z})_{ss'} = z_s \delta_{ss'}$. Since the intersurface interaction is symmetric, $\Gamma_{12} = \Gamma_{21}$, Eq. (64) yields

$$\underline{z} + \underline{\Gamma} = \mathrm{const} \times \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \tag{65}$$

This relation is consistent with Eq. (52). However, contrary to Eq. (52), the relation (65) is manifestly imposed by the symmetry ("$\mathcal{F}$ invariance") of the action (54). It should therefore remain intact under RG flow.

3. Gauging the NLσM and linear-response theory

Generally, the requirement of gauge invariance prescribes the correct coupling to the scalar and vector potentials in the action of the NLσM, Eq. (54). In particular, in the kinetic term, one has to replace $\partial_\mu Q_s \to D_{\mu,s} Q_s$ with the long derivative D_μ of the form

$$D_{\mu,s} Q_s \equiv \partial_\mu Q_s + \sum_{n,\alpha} i A_{\mu,s,-n}^\alpha \left[J_n^\alpha - \left(J_n^\alpha \right)^T, Q_s \right]. \tag{66}$$

For simplicity, the electron charge is absorbed into the vector potential here and in the following section.

As the theory is nonlocal in the imaginary time, the inclusion of the scalar potential is nonlinear. The corresponding term that should be added to the NLσM (54) reads

$$S^\Phi = -2 \sum_{n\alpha,ss'} \Phi_{n,s}^\alpha (\underline{z} + \underline{\Gamma})_{ss'} \mathrm{tr} J_n^\alpha Q_s + \frac{1}{\pi T} \sum_{n\alpha,ss'} \Phi_{n,s}^\alpha (\underline{z} + \underline{\Gamma})_{ss'} \Phi_{-n,s}^\alpha. \tag{67}$$

The inclusion of the scalar and vector potentials allow us to express the density-density correlation function and the conductivity in terms of the matrix fields Q_s by means of the linear-response theory. In particular, a double differentiation of the partition function with respect to the scalar potential yields the density-density response:

$$\Pi_{ss'}^{\mathrm{RPA}}(\omega_n, \boldsymbol{q}) = -\frac{2}{\pi}(z+\Gamma)_{ss'} + 4T \sum_{s_1,s_2} (z+\Gamma)_{ss_1} \left\langle \mathrm{tr} J_n^\alpha Q_{s_1}(\boldsymbol{q}) \times \mathrm{tr} J_{-n}^\alpha Q_{s_2}(-\boldsymbol{q}) \right\rangle (z+\Gamma)_{s_2 s'}. \tag{68}$$

Here, $\langle \ldots \rangle$ denotes average with respect to the action (54). The superscript "RPA" emphasizes that the quantity appearing in the total density-density response includes RPA resummation. It is thus one-Coulomb-line-reducible, and only its irreducible part corresponds to the polarization operator.

In the same spirit, we obtain the expression for the conductivity (in units of e^2/h) at a finite, positive frequency ω_n:

$$\sigma_{ss'}'(\omega_n) = B_1^{(s)} \delta_{ss'} + B_2^{(ss')}. \tag{69}$$

Here, we introduced two correlators:

$$B_1^{(s)} = \frac{\sigma_s}{8n} \left\langle \mathrm{tr} \left[J_n^\alpha - \left(J_n^\alpha \right)^T, Q_s \right] \left[J_{-n}^\alpha - \left(J_{-n}^\alpha \right)^T, Q_s \right] \right\rangle \tag{70}$$

and

$$B_2^{(ss')} = \frac{\sigma_s \sigma_{s'}}{128n} \int_{\boldsymbol{x}-\boldsymbol{x}'} \sum_{\mu=x,y} \left\langle \mathrm{tr} \left\{ \left[J_n^\alpha - \left(J_n^\alpha \right)^T, Q_s \right] \partial_\mu Q_s \right\}_{\boldsymbol{x}} \times \mathrm{tr} \left\{ \left[J_{-n}^\alpha - \left(J_{-n}^\alpha \right)^T, Q_{s'} \right] \partial_\mu Q_{s'} \right\}_{\boldsymbol{x}'} \right\rangle. \tag{71}$$

Substituting the saddle-point value $Q_s = \Lambda$, we obtain the classical value $\sigma_{ss'}'(\omega_n) = \sigma_s \delta_{ss'}$. Hence the dimensionless coupling constant of the NLσM has been identified with the physical conductivity in units of e^2/h.

4. Gauging the theta term and anomalous quantum Hall effect

The local expression of the $\mathbb{Z}_2$ theta term, i.e, the WZW term, Eq. (56), also allows of inclusion of gauge potentials.[84–89] However, the situation is more subtle here. Specifically, it turns out that the contribution of nonsingular gauge potentials to the topological term $S^{(\theta)}$ vanishes. We explicitly show this in Appendix A 1.

The situation changes when the time-reversal symmetry is broken (at least, in some spatial domain at the surface) by a random or/and uniform magnetic field. Subjected to a strong magnetic field, 3D TI surface states display the characteristic quantum Hall effect of Dirac electrons[25,90] with quantized transverse conductance:

$$\sigma_{xy} = g \left(n \pm \frac{1}{2} \right) \frac{e^2}{h}, \; n \in \mathbb{Z}, \tag{72}$$

where g is the degeneracy of Dirac electrons, e.g., $g = 2$ for two 3D TI surfaces. It is intimately linked to the topological magnetoelectric effect.[91–94] Theoretically, the anomalous quantum Hall effect was explained and discussed in a previous work by three of the authors.[95] We will explain in the following how to understand it in the framework of the linear response theory within the NLσM. As it turns out, the crucial point is that gauge potentials drop from $S^{(\theta)}$.

We first briefly recall the NLσM field theory describing the ordinary integer QHE (i.e., for electrons with quadratic dispersion). It contains Pruisken's theta term,[96] which assumes the following form upon inclusion of the vector potential:[68]

$$S^{\mathrm{QHE}} = \frac{\vartheta}{16\pi}\int_x \epsilon_{\mu\nu}\mathrm{tr}\, Q_U\partial_\mu Q_U\partial_\nu Q_U \tag{73a}$$

$$+\frac{i\vartheta}{4\pi}\int_x \epsilon_{\mu\nu}\mathrm{tr}\partial_\mu \hat{A}_\nu Q_U \tag{73b}$$

$$+\frac{\vartheta}{4\pi}\int_x \epsilon_{\mu\nu}\sum_{n,\alpha} n A^{\alpha}_{\mu,n}A^{\alpha}_{\nu,-n}. \tag{73c}$$

Here, $Q_U = U^{-1}\Lambda U$ with $U \in \mathbf{U}(2N_M \times N_R)$, $\epsilon_{\mu\nu} = -\epsilon_{\nu\mu}$ is the 2D antisymmetric symbol ($\epsilon_{xy} \overset{\text{def.}}{=} 1$), and ϑ is the theta angle of the Pruisken's NLσM. We emphasize that the last two terms [see Eqs. (73b) and (73c)] determine the effective electromagnetic response and thus prescribe the relation between the physical observable σ_{xy} (in units of e^2/h) and the theta angle ϑ. In particular, $\vartheta/2\pi$ is identified as the bare value of the Hall conductance.[97]

Let us now turn to a single Dirac surface state. As has been discussed above, all gauge potentials drop from $S^{(\theta)}$. Let us first add a random magnetic field (keeping zero average magnetic field) to the gauged NLσM. This implies a breakdown of the symmetry:

$$\mathcal{M} \to \frac{\mathbf{U}(2N_MN_R)}{\mathbf{U}(N_MN_R)\times\mathbf{U}(N_MN_R)}. \tag{74}$$

The $\mathbb{Z}_2$ theta term becomes the Pruisken's theta term[98] (recall $\theta = \pi \bmod 2\pi$):

$$S^{(\theta)}_U = \frac{\theta}{16\pi}\int_x \epsilon_{\mu\nu}\mathrm{tr}\, Q_U\partial_\mu Q_U\partial_\nu Q_U. \tag{75}$$

We emphasize that together with the gauged kinetic term $S^{(\theta)}_U$ is the complete gauged theory, no extra terms of the types (73b) and (73c) appear. Being topological, the Pruisken's theta term is invariant under smooth $\mathbf{U}(1)$ rotations. Recall that exactly the terms (73b) and (73c) provided a link between ϑ and σ_{xy} in the conventional (non-Dirac) QHE setting. Their absence in Eq. (75) is thus physically very natural: without a net magnetic field the Hall conductivity is zero.

We consider now the case when the average magnetic field is nonzero. The action of the NLσM describing a Dirac fermion is then given by a sum of Eqs. (73) and (75). The renormalization of the action of the NLσM is governed by the full theta angle $\vartheta + \theta$. On the other hand, only ϑ is related with the bare value of σ_{xy}. Then standard arguments for the quantization of the Hall conductivity[47] lead to the result (72) for the anomalous QHE.

IV. ONE-LOOP RG

In the preceding section, we have derived the diffusive NLσM, Eq. (54). We will now investigate its behavior under renormalization. This will allow us, in particular, to deduce the scale dependence of the conductivity. The most important steps of the calculation are presented in the main text (further details can be found in Appendix D).

We calculate the renormalization of the NLσM parameters within the linear-response formalism (rather than the background-field method). This is favorable since it implies a more direct physical interpretation of the NLσM coupling constants. Furthermore, this way, one can in principle treat simultaneously different infrared regulators, such as temperature or frequency. However, for the sake of clarity of presentation, we restrict ourselves to a purely field-theoretical regularization scheme and add a mass term to the action

$$S_L = -\sum_{s=1,2}\frac{\sigma_s L^{-2}}{8}\int_x \mathrm{tr}\Lambda Q_s. \tag{76}$$

The connection between the running length scale L and the physical regulators temperature or frequency was analyzed in Ref. 99. Roughly speaking, in the presence of a single infrared scale E, e.g., when calculating dc conductance at finite temperature and assuming an infinite sample, one can replace L by L_E in the results.

We will calculate all UV-divergent contributions in the dimensional regularization scheme. This allows us to preserve the local $\mathrm{O}(2_\tau \times N_M \times N_R)\times \mathrm{O}(2_\tau \times N_M \times N_R)$ symmetry of the Q matrix (35) and to ensure the renormalizability of the theory.

A. Diffusive propagators

We employ the exponential parameterization of the matrix fields $Q_s = \Lambda \exp W_s$. The antisymmetric fields

$$W_s = \begin{pmatrix} 0 & q_s \\ -q_s^T & 0 \end{pmatrix}$$

anticommute with Λ. Further, we define a set of real matrices in the particle-hole space: $\tilde{\tau}_\mu \equiv 2^{-1/2}(\mathbf{1}, \tau_x, i\tau_y, \tau_z)$. This allows us to introduce the fields $q^{(\mu)} \equiv \mathrm{tr}^\tau q\tilde{\tau}^T_\mu$, where tr^τ is the trace in the particle-hole space only. With these definitions at hand, we expand the action, Eqs. (54) and (76), to quadratic order in $q^{(\mu)}$ and obtain the NLσM propagators that describe the diffusive motion in the particle-hole (diffusons) and particle-particle (cooperons) channels.

The fields $q^{(1)}$ and $q^{(3)}$ describe cooperons. Their propagator is unaffected by interaction (since we have discarded the interaction in the Cooper channel),

$$\left\langle \left[q^{(\mu)}_s(\boldsymbol{p})\right]^{\alpha_1\alpha_2}_{m_1m_2}\left[q^{(\nu)}_{s'}(-\boldsymbol{p})\right]^{\beta_1\beta_2}_{n_1n_2}\right\rangle = \frac{4}{\sigma_s}D_s(\omega_{n_{12}},\boldsymbol{p})\delta_{ss'}\delta_{\mu\nu}\delta_{n_1m_1}\delta_{n_2m_2}\delta_{\alpha_1\beta_1}\delta_{\alpha_2\beta_2}(\delta_{\mu 1}+\delta_{\mu 3}), \tag{77}$$

where

$$[D_s(\omega_{n_{12}},\boldsymbol{p})]^{-1} = \boldsymbol{p}^2 + L^{-2} + \frac{4z_s}{\sigma_s}\omega_{n_{12}}. \tag{78}$$

The Matsubara indices n_1, m_1 are non-negative, while the indices n_2, m_2 are negative; we have also defined $n_{12} \equiv n_1 - n_2 > 0$ and $m_{12} \equiv m_1 - m_2 > 0$.

Next, we consider the diffusons $q^{(0)}$ and $q^{(2)}$. Their Green's function, written as a matrix in surface space, is

$$\begin{aligned}\left\langle \left[q_s^{(\mu)}(\boldsymbol{p})\right]_{m_1 m_2}^{\alpha_1\alpha_2} \left[q_{s'}^{(\nu)}(-\boldsymbol{p})\right]_{n_1 n_2}^{\beta_1\beta_2} \right\rangle \\ = \frac{4}{\sigma_s} D_s(\omega_{n_{12}}, \boldsymbol{p}) \delta_{\mu\nu} \delta_{n_{12},m_{12}} \delta_{\alpha_1\beta_1} \delta_{\alpha_2\beta_2} (\delta_{\mu 0} + \delta_{\mu 2}) \\ \times \left\{ \delta_{n_1 m_1} \delta_{ss'} - \frac{8\pi T}{\sigma_{s'}} \delta^{\alpha_1\alpha_2} [\underline{\Gamma D^c}(\omega_{n_{12}}, \boldsymbol{p})]_{ss'} \right\}. \quad (79)\end{aligned}$$

Here, we have introduced

$$[D^c(\omega_{n_{12}}, \boldsymbol{p})]^{-1}_{ss'} = D_s^{-1}(\omega_{n_{12}}, \boldsymbol{p}) \delta_{ss'} + \frac{4\omega_{n_{12}}}{\sigma_s} \Gamma_{ss'}. \quad (80)$$

B. RG invariants

The bare action contains, aside from the mass L^{-1}, seven running coupling constants: σ_1, σ_2, z_1, z_2, Γ_{11}, Γ_{22}, and Γ_{12}. We are now going to show that three linear combinations of them are conserved under RG. To this end, we evaluate the density-density response (68) at the tree level:

$$\underline{\Pi}^{\mathrm{RPA}}(\omega, \boldsymbol{p}) = -\frac{2}{\pi} [\underline{z} + \underline{\Gamma}](1 - 4\omega \underline{\sigma^{-1} D^c}(\omega, \boldsymbol{p})[\underline{z} + \underline{\Gamma}]), \quad (81)$$

where $(\underline{\sigma})_{ss'} = \sigma_s \delta_{ss'}$. There is no need for infrared regularization here, and we therefore omit the mass term (76).

On the other hand, the density-density response function can be obtained from the fermionic formulation of the theory, see Appendix C 5:

$$\underline{\Pi}^{\mathrm{RPA}} = [\underline{\Pi}^q - \underline{\nu\Gamma^0\nu}](1 + \omega \underline{\Delta}^{\Gamma}(\omega, \boldsymbol{p})[\underline{\Pi}^q - \underline{\nu\Gamma^0\nu}]), \quad (82)$$

where

$$\Delta^{\Gamma}(\omega, \boldsymbol{p}) = [\underline{\nu D} \boldsymbol{p}^2 + \omega(\underline{\nu} + \underline{\nu\Gamma^{\rho,q}\nu})]^{-1}. \quad (83)$$

The equality of Eqs. (81) and (82) relates two functions of momentum and frequency. In the static limit, we find the following constraint connecting the NLσM coupling constants with physical FL parameters:

$$\frac{2}{\pi} (\underline{z} + \underline{\Gamma}) = -\underline{\Pi^q} + \underline{\nu\Gamma^0\nu}. \quad (84)$$

Next, from comparison of momentum dependence in Eqs. (81) and (82), we find the Einstein relation: $\sigma_s = 2\pi \nu_s D_s$. Accordingly, σ measures the conductance in units of e^2/h, consistently with what has been found in Secs. III B and III G3.

In view of gauge invariance (see Sec. III F4), the static polarization operator entering Eq. (84) is nothing but the compressibility

$$\Pi^{ss',q} = -\frac{\partial N_s}{\partial \mu_{s'}}.$$

Its value is not renormalized because it can be expressed as a derivative of a physical observable with respect to the chemical potentials. On ballistic scales, the chemical potential enters logarithmically divergent corrections only as the UV cutoff of the integrals. In the diffusive regime, the UV cutoff is provided by the scattering rates $\tau_s^{-1} \ll |\mu_s|$. Therefore diffusive contributions to the derivative with respect to the chemical potential vanish.[40] Since $\underline{\nu\Gamma^0\nu}$ only depends on $\underline{\Pi}^q$ (see Appendix C 4), it is not renormalized as well. Therefore the right-hand side of Eq. (84) is not renormalized and hence neither is its left-hand side, i.e., $\underline{z} + \underline{\Gamma}$. This matrix constraint yields three RG invariants: $z_1 + \Gamma_{11}$, $z_2 + \Gamma_{22}$, and Γ_{12}. Thus, only four out of seven NLσM parameters are independent running coupling constants. We emphasize that, in contrast to Eq. (65), this reasoning is valid also in the absence of long-range interaction.

Finally, let us evaluate Eq. (84) on the bare level. Expressing the static polarization operator as $\underline{\Pi}^q = -\underline{\nu} - \underline{\nu\Gamma^{1-2}\nu}$ and using the definition of z_s in Sec. III E1 one can find the following relations for the bare values:

$$\frac{4\lambda_s}{\pi} \equiv \frac{2}{\pi} z_s = \nu_s. \quad (85)$$

Equivalently, the same relationship between λ_s and ν_s can be obtained by comparing the bare definition of $\underline{\Gamma}$ [see Eq. (58)] with the right-hand side of Eq. (84). The relation (85) has been foreseen earlier on the basis of SCBA, see Eq. (33). In conclusion, the SCBA and the density response independently show that the UV cutoff scale for the bosonization is automatically set by the chemical potential (which is also very natural from the physical point of view).

C. Renormalization of conductivities

1. Correlator B_1

We will first analyze the correlator $B_1^{(s)}$, see Eq. (70). The one-loop correction is determined by the expansion to second order in $q^{(\mu)}$. The tensor structure in particle-hole space implies that the diffuson contribution ($\mu = 0,2$) vanishes. The classical value together with the cooperon contribution ($\mu = 1,3$) is

$$B_1^{(s)} = \sigma_s + 2 \int_{\boldsymbol{p}} D_s(\omega_n, \boldsymbol{p}). \quad (86)$$

We evaluate this term in the announced regularization scheme:

$$\begin{aligned} B_1^{(s)} &= \sigma_s + 2\mathrm{I}_1^{(2+\epsilon)} & (87) \\ &= \sigma_s + \frac{1}{2\pi} \left(-\frac{2}{\epsilon} + 2 \ln L/l + \mathrm{const} \right). & (88) \end{aligned}$$

For dimensional reasons, we have introduced the reference length scale l, which for the present diffusive problem is set by the mean-free path $l = \max_{s=1,2} l_s$. We have further evaluated the following standard dimensionless integral:

$$\begin{aligned} \mathrm{I}_1^{(D)} &\equiv l^{D-2} \int \frac{d^D p}{(2\pi)^D} \frac{1}{\boldsymbol{p}^2 + L^{-2}} \\ &= \frac{\left(\frac{l^2}{L^2}\right)^{\frac{D}{2}-1}}{(4\pi)^{\frac{D}{2}}} \Gamma\left(1 - \frac{D}{2}\right) \\ &\overset{D=2+\epsilon}{=} \frac{1}{4\pi} \left[-\frac{2}{\epsilon} + 2 \ln L/l + \ln 4\pi - \boldsymbol{\gamma} + \mathcal{O}(\epsilon) \right], \end{aligned}$$

where $\boldsymbol{\gamma} \approx 0{,}577$ is the Euler-Mascheroni constant. The logarithmic term in Eq. (88) is nothing but the well-known weak-antilocalization effect.[100]

2. *Correlator* B_2

Next, we turn our attention to $B_2^{(ss')}$, Eq. (71). Because of the presence of gradients, it does not contribute neither at classical nor at tree level. Furthermore, due to the absence of the Cooper channel and the uncorrelated disorder on the top and bottom surfaces, there are no quantum corrections to the transconductance σ_{12}. The correlator $B_2^{(ss')}$ can be recast into the form (see Appendix D)

$$B_2^{(ss')} = \frac{16\delta_{ss'}}{n\sigma_s}\int_{\boldsymbol{p}} \boldsymbol{p}^2 \sum_{\omega_m>0}\omega_m \times[(\underline{D\Gamma D^c})_{ss}(\omega_m,\boldsymbol{p})D_s(\omega_{m+n},\boldsymbol{p}) -(\underline{D\Gamma D^c})_{ss}(\omega_{m+n},\boldsymbol{p})D_s(\omega_{m+2n},\boldsymbol{p})]. \quad (89)$$

For its evaluation, it is instructive to separate contributions stemming from intrasurface interaction Γ_{ss} and intersurface interaction Γ_{12}. This leads to

$$B_2^{ss'} = -4\delta_{ss'}\left\{\underbrace{1-\frac{1+\gamma_{ss}}{\gamma_{ss}}\ln(1+\gamma_{ss})}_{\text{single surface}} + \underbrace{(1+\gamma_{ss})\left[\frac{\ln(1+\gamma_{ss})}{\gamma_{ss}}-\frac{\ln(1+\tilde{\gamma}_{ss})}{\tilde{\gamma}_{ss}}\right]}_{\text{intersurface interaction}}\right\} I_2^{(2+\epsilon)}$$
$$= -\frac{\delta_{ss'}}{\pi}\left[1-\frac{1+\gamma_{ss}}{\tilde{\gamma}_{ss}}\ln(1+\tilde{\gamma}_{ss})\right] \times\left(-\frac{2}{\epsilon}+2\ln L/l+\text{const}\right). \quad (90)$$

We have introduced $\gamma_{ss}=\Gamma_{ss}/z_s$, $\tilde{\gamma}_{11}=\gamma_{11}+(\sigma_1/\sigma_2)(1+\gamma_{11})$ and $\tilde{\gamma}_{22}=\gamma_{22}+(\sigma_2/\sigma_1)(1+\gamma_{22})$. Note that in the limit of $z_2+\Gamma_{22}=0$ [which corresponds to $\Gamma_{12}=0$ in view of Eq. (65)], we recover the well-known conductivity corrections to σ_{11} for a single surface (see also Sec. V B1). Further, in Eq. (90), we have evaluated the second standard diverging integral

$$\mathrm{I}_2^{(D)} \equiv l^{D-2}\int\frac{d^Dp}{(2\pi)^D}\frac{\boldsymbol{p}^2}{(\boldsymbol{p}^2+L^{-2})^2} = \frac{\left(\frac{l^2}{L^2}\right)^{\frac{D}{2}-1}}{(4\pi)^{\frac{D}{2}}}\frac{D}{2}\Gamma\left(1-\frac{D}{2}\right) \overset{D=2+\epsilon}{=} \frac{1}{4\pi}\left[-\frac{2}{\epsilon}+2\ln L/l+\ln 4\pi-1-\boldsymbol{\gamma}+\mathcal{O}(\epsilon)\right].$$

D. Renormalization of the interaction amplitudes

The renormalization of the interaction amplitudes, or equivalently, of Finkelstein parameters z_s, is intimately linked to the renormalization of the specific heat.[101] This is because the scale (e.g., temperature) dependence of the total thermodynamic potential Ω is governed by the scale dependence of z_s. In the present case of coupled surfaces, we can only extract the correction to the sum z_1+z_2 from the (one-loop) correction to the total thermodynamic potential:[99]

$$z_1'+z_2' = \frac{1}{2\pi\,\mathrm{tr}\eta\Lambda}\frac{\partial}{\partial T}\frac{\Omega}{T}. \quad (91)$$

At the classical level, Eq. (91) yields the relation $z_1'+z_2'=z_1+z_2$. Evaluating the quantum corrections in Eq. (91), we find

$$(z_1'+z_2') = (z_1+z_2)+2\sum_{s=1,2}\Gamma_{ss}\int_{\boldsymbol{p}} D_s(0,\boldsymbol{p}). \quad (92)$$

As the correction is a sum of contributions from the two opposite surfaces, it is natural to assume that the parameters z_s are renormalized *separately* (and without intersurface interaction effects):

$$z_s' = z_s+2\Gamma_{ss}\int_{\boldsymbol{p}} D_s(0,\boldsymbol{p}) = z_s+2\frac{\Gamma_{ss}}{\sigma_s}\mathrm{I}_1^{(2+\epsilon)} = z_s+\frac{1}{2\pi}\frac{\Gamma_{ss}}{\sigma_s}\left(-\frac{2}{\epsilon}+2\ln L/l+\text{const}\right). \quad (93)$$

We have directly proven this assumption of separate z_s renormalization by the background field method.[123]

E. The one-loop RG equations

Applying the minimal subtraction scheme to Eqs. (88), (90), and (93), we derive the one-loop perturbative RG equations:

$$\frac{d\sigma_1}{dy} = -\frac{2}{\pi}F\left(\gamma_{11},\frac{\sigma_1}{\sigma_2}\right), \quad (94a)$$
$$\frac{d\sigma_2}{dy} = -\frac{2}{\pi}F\left(\gamma_{22},\frac{\sigma_2}{\sigma_1}\right), \quad (94b)$$
$$\frac{d\gamma_{11}}{dy} = -\frac{\gamma_{11}(1+\gamma_{11})}{\pi\sigma_1}, \quad (94c)$$
$$\frac{d\gamma_{22}}{dy} = -\frac{\gamma_{22}(1+\gamma_{22})}{\pi\sigma_2}, \quad (94d)$$

where $y=\ln L/l$, $\gamma_{ss}=\Gamma_{ss}/z_s$, $l=\max_{s=1,2}l_s$, and

$$F(\gamma,x) = \frac{1}{2}-\frac{1+\gamma}{x\left[1+\gamma\left(1+\frac{1}{x}\right)\right]}\ln[(1+x)(1+\gamma)]. \quad (95)$$

We recall that Γ_{12}, $z_1+\Gamma_{11}$ and $z_2+\Gamma_{22}$ are not renormalized. We mention that the mass L^{-1} acquires a quantum correction[99] but it does not affect the one-loop renormalization of the other parameters σ_s, z_s, and $\Gamma_{ss'}$.

For an alternative presentation of the RG Equations (94), we introduce the total conductivity $\sigma=\sigma_1+\sigma_2$ and the ratio of the conductivities of the two surfaces $t=\sigma_1/\sigma_2$. In terms of these parameters, the RG equations take the following form:

$$\frac{d\sigma}{dy} = -\frac{2}{\pi}\left\{1-\frac{1}{t}\frac{1+\gamma_{11}}{1+\gamma_{11}\left(1+\frac{1}{t}\right)}\ln[(1+t)(1+\gamma_{11})] -t\frac{1+\gamma_{22}}{1+\gamma_{22}(1+t)}\ln\left[\left(1+\frac{1}{t}\right)\left(1+\gamma_{22}\right)\right]\right\}, \quad (96a)$$
$$\frac{dt}{dy} = -\frac{2}{\pi}\frac{1+t}{\sigma}\left\{\frac{1-t}{2}-\frac{1}{t}\frac{1+\gamma_{11}}{1+\gamma_{11}\left(1+\frac{1}{t}\right)} \times\ln[(1+t)(1+\gamma_{11})]+t^2\frac{1+\gamma_{22}}{1+\gamma_{22}(1+t)} \times\ln\left[\left(1+\frac{1}{t}\right)(1+\gamma_{22})\right]\right\}, \quad (96b)$$
$$\frac{d\gamma_{11}}{dy} = -\left(1+\frac{1}{t}\right)\frac{\gamma_{11}(1+\gamma_{11})}{\pi\sigma}, \quad (96c)$$

$$\frac{d\gamma_{22}}{dy} = -(1+t)\frac{\gamma_{22}(1+\gamma_{22})}{\pi\sigma}. \tag{96d}$$

V. ANALYSIS OF THE RG EQUATIONS

It is worthwhile to remind the reader that the RG Equations (94) describe the quantum corrections to conductivity due to the interplay of two distinct effects. First, they contain weak-antilocalization corrections (WAL) $\delta\sigma_s^{\rm WAL} = (1/\pi)\ln L/l$ due to quantum interference in a disordered system with the strong spin-orbit coupling. Second, these are interaction-induced contributions of Altshuler-Aronov (AA) type, including effects of both, long-range and short-range interactions. The result (94) was obtained perturbatively to leading order in $1/\sigma_s \ll 1$ but it is exact in the singlet interaction amplitudes. While these equations describe the experimentally most relevant case of Coulomb interaction, in Appendix F, we also present the RG equations for the case of short-range interaction.

Equation (94), which determine the flow of the coupling constants $\sigma_1,\sigma_2,\gamma_{11}$, and γ_{22} implies a rich phase diagram in the four-dimensional parameter space. Before discussing the general four-dimensional RG flow, we highlight the simpler case of two equal surfaces.

A. Two equal surfaces

Equal surfaces are defined by $\sigma_1 = \sigma_2 = \sigma/2$, $\gamma_{11} = \gamma_{22} = \gamma$, and, because of Eq. (65), $\gamma_{12} = -1-\gamma$. It can be checked that the plane of identical surfaces is an attractive fixed plane of the four-dimensional RG flow (see Appendix E). The RG equations for the two coupling constants σ and γ are

$$\frac{d\sigma}{dy} = -\frac{2}{\pi}\left[1 - \frac{2+2\gamma}{1+2\gamma}\ln(2+2\gamma)\right], \tag{97a}$$

$$\frac{d\gamma}{dy} = -\frac{2\gamma(1+\gamma)}{\pi\sigma}. \tag{97b}$$

Experimentally, the case of equal surfaces is realized if both surfaces are characterized by the same mean-free path and the same carrier density and, furthermore, if the dielectric environment of the probe is symmetric ($\epsilon_1 = \epsilon_3$).

1. Flow Diagram within the fixed plane

The RG flow within the σ-γ plane is depicted in Fig. 11. The green vertical fixed line at $\gamma = -1$ corresponds to the case of two decoupled surfaces (recall $\gamma_{12} = -1-\gamma$), and reproduces the result of Ref. 35 for a single surface of 3D TI. In this limit, the total correction to the conductivity is negative and obeys the universal law

$$\delta\sigma_{\gamma=-1} = 2\times\frac{2}{\pi}\left(\underbrace{1/2}_{\rm WAL} - \underbrace{1}_{\rm AA}\right)\ln L/l = -\frac{2}{\pi}\ln L/l. \tag{98}$$

The line of decoupled surfaces is repulsive, as can be seen from Eq. (97b). Flowing towards the infrared, the conductivity first decreases before turning up again while the system approaches the second fixed line at $\gamma = 0$. Note that on this line $\gamma_{12} = -1$: the intrasurface interaction has died out, but the intersurface interaction is maximal. Here, the conductivity correction is positive indicating the flow into a metallic state:

$$\delta\sigma_{\gamma=0} = 2\times\frac{2}{\pi}\left[\underbrace{1/2}_{\rm WAL} - \underbrace{(1-\ln 2)}_{\rm interaction}\right]\ln L/l. \tag{99}$$

The flow on this fixed line is towards the perfect-metal point

$$(1/\sigma^*, t^*, \gamma_{11}^*, \gamma_{22}^*) = (0,1,0,0)\,.$$

As discussed below, see Sec. V B 2, this is the only attractive fixed point even in the case of the general four-dimensional RG flow. On the $\gamma = 0$ fixed line, the intersurface interaction reduces the strength of the WAL effect but it is not strong enough to reverse the behavior. The region $\gamma > 0$ corresponds to attractive interaction in the singlet channel and is shown on the flow diagram for the sake of completeness.

2. Typical bare values and crossover scale

Typically, before renormalization, the intersurface interaction γ_{12} is weaker than or equal to the intrasurface interaction γ. This implies that its bare value γ_0 takes values in the range between $\gamma_0 = -1$ (decoupled surfaces, i.e., $\gamma_{12,0} = 0$) and $\gamma_0 = -1/2 = \gamma_{12,0}$. For small α, we can approximate γ_0 by its RPA value:

$$\gamma_0 = -\frac{1}{2} - \frac{1}{2}\frac{\kappa d}{1+\kappa d}. \tag{100}$$

Here, d is the system thickness and $\kappa = 2\pi\frac{e^2}{\epsilon_2}\nu$ the inverse single surface screening length obtained for the general symmetric situation: $\epsilon_1 = \epsilon_3 \neq \epsilon_2$, see Appendix B. Note that at $\kappa d = 0$ the conductivity corrections due to WAL and AA exactly compensate each other:

$$\delta\sigma_{\gamma=-1/2} = \frac{2}{\pi}\left(2\times\underbrace{1/2}_{\rm WAL} - \underbrace{1}_{\rm AA}\right)\ln L/l = 0,$$

as can also be seen in Fig. 11.

Typically $\kappa d > 0$ or, as already explained on general grounds, $-1 < \gamma_0 < -1/2$. Then the most drastic conse-

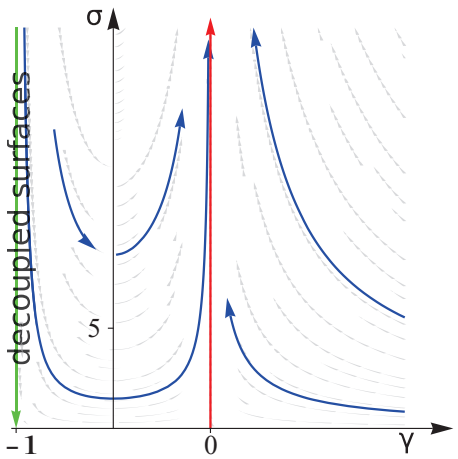

FIG. 11. (Color online) RG flow for equal surfaces in the parameter space σ (total conductivity) and γ (intrasurface interaction strength). Here and in all following RG diagrams, arrows indicate the flow towards the infrared.

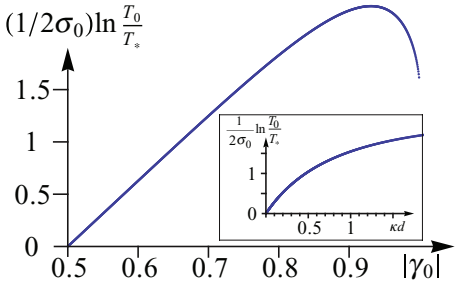

FIG. 12. (Color online) Temperature scale associated with the minimum of σ as a function of the bare values σ_0 and γ_0. (Inset) Same quantity as a function of σ_0 and κd.

quence of intersurface interaction is the nonmonotonic temperature (or length) dependence: the conductivity first decreases with lowering T but eventually the sign of $d\sigma/dT$ changes and the system is ultimately driven into the metallic phase. It is natural to ask for the temperature scale, which is associated with this sign change. The scale y_* at which the conductivity reaches its minimum can be extracted from Eq. (97) and is expressed by the integral

$$y_* = -\frac{\pi\sigma_0}{2}\int_{\gamma_0}^{\gamma_*}\frac{d\gamma'}{\gamma'}\frac{1+\gamma_0}{(1+\gamma')^2}\left(\frac{\gamma'}{\gamma_0}\right)^{1-2\ln 2} e^{2[f(\gamma')-f(\gamma_0)]}, \tag{101}$$

where $f(x) = \mathrm{Li}_2(-x) - \mathrm{Li}_2[-(1+2x)]$, Li_2 is the dilogarithm, and $\gamma_* = -1/2$.

Numerical integration of Eq. (101) yields the crossover length scale or temperature $y_* = \ln L_*/l = 1/2\ln T_0/T_*$. Its dependence on the bare values σ_0 and γ_0 is plotted in Fig. 12. Using Eq. (100), one can also investigate the dependence of y_* on κd instead of γ_0 (see inset in Fig. 12).

3. *Role of topology: Dirac electrons versus electrons with quadratic dispersion in the presence of spin-orbit interaction*

The perturbative RG Equations (94) and (97) are valid for $\sigma \gg 1$. Instanton effects are suppressed by $\exp(-2\pi\sigma)$ in this region and we therefore neglected them. As has been discussed in Sec. III E2, in the diffusive NLσM of Dirac electrons, the $\mathbb{Z}_2$ theta term reflects the topological protection from Anderson localization. This term is absent in the case of nontopological symplectic metals (NTSM) such as electrons with quadratic dispersion subjected to strong spin-orbit coupling.[124] The presence (respectively, absence) of the topological term results in the opposite signs of the instanton contribution in the two cases. However, as instantons are suppressed, our perturbative result is equally applicable to the surfaces of a 3D TI and, for example, to a double-quantum-well structure in a material with strong spin-orbit coupling. Here, we discuss nonperturbative differences between the two problems.

Let us start from the case of decoupled surfaces (green line, i.e., $\gamma = -1$, in Fig. 13). This limiting case has been analyzed before.[35] For NTSM, localizing AA corrections overcome the WAL effect and the system always flows towards localization (see Fig. 13, left). In contrast, for TI the topological protection implies $d\sigma/dy > 0$ for small σ and hence an attractive fixed point at $\sigma \sim 1$ (see Fig. 13, right).

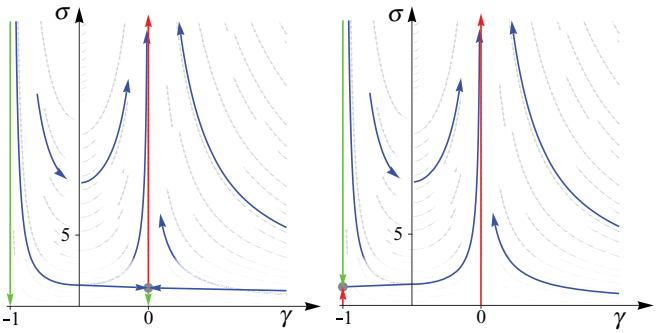

FIG. 13. (Color online) Comparison between expected RG flow for a double layer system of NTSM (left) and the coupled surfaces of a thin 3D TI film (right).

As has been explained, the $\gamma = -1$ line is unstable with respect to the intersurface interaction and the system eventually flows towards the antilocalizing red line at $\gamma = 0$. Let us now analyze this fixed line. The fact that conductivity corrections (99) are positive stems back to the (noninteracting) WAL effect. Its contribution $2\times(1/\pi)\ln L/l$ is independent of σ only for $\sigma \gg 1$. For NTSM it decreases with decreasing σ and eventually becomes negative at the metal-insulator transition (MIT) point $\sigma_{\mathrm{MIT}} \approx 2\times 1.42\, e^2/h$.[102–104] (As explained above, Sec. III E2, in a recent investigation[78] the crucial role of $\mathbb{Z}_2$ vortices for this MIT was pointed out.) Qualitatively, the picture of the MIT survives the presence of interactions, which even enhance the tendency to localization. Therefore, for the double layer system of NTSM, we expect the antilocalizing RG flow on the $\gamma = 0$ line to turn localizing below some $\sigma_{\mathrm{MIT}} \sim 1$. This MIT point is indicated by a dot in the left panel of Fig. 13.

In contrast, for the surfaces of a topological insulator the system is topologically protected from Anderson localization,[77] i.e., the β function $d\sigma/dy$ bends up when $\sigma \to 0$. There is a numerical evidence[105,106] that in a noninteracting case, this happens without any intermediate fixed points. Again, the arguments are qualitatively unchanged by the presence of (pure intersurface) interaction and this scenario is expected to hold also on the red $\gamma = 0$ line of the thin 3D TI film, see Fig. 13, right. (Strictly speaking, one cannot rule out a possibility that in the presence of interaction there emerge intermediate fixed points but we assume the simplest possible flow diagram consistent with large- and small-conductivity behavior.)

The interpolation between the limiting cases of decoupled surfaces and maximally interacting surfaces produces the two phase diagrams shown in Fig. 13. For a double layer of NTSM, there is a separatrix connecting the weak coupling, decoupled layers fixed point $(\gamma, 1/\sigma) = (-1,0)$ with the critical MIT point $(\gamma, 1/\sigma) \sim (0,1)$ that we introduced above. (Strictly speaking, we cannot exclude the possibility that this fixed point might lie slightly off the $\gamma = 0$ line.) Below the separatrix, the conductivity renormalizes down to $\sigma = 0$, i.e., the system is in the Anderson-localized phase. In contrast, above the separatrix, the characteristic nonmonotonic conductivity behavior leads to the metallic state. As the horizontal position in the phase diagram is controlled by the parameter κd, we predict a quantum phase transition between metal and insulator as a

function of the interlayer distance. On the other hand, in the case of the coupled top and bottom surfaces of a thin 3D TI film, the flow is always towards the metallic phase. The critical point of decoupled surfaces at $\gamma = -1$ with $\sigma \sim 1$ is unstable in the direction of γ.

It is worth recalling that in this paper, we neglected the tunneling between the opposite surfaces of the 3D TI. If such a tunneling is included, it introduces a corresponding exponentially small energy scale below which the two surfaces behave as a single-layer NTSM. This would in particular imply a crossover to localizing behavior at such exponentially low temperatures.

B. General RG flow

We now turn our attention to the complete analysis of RG Equations (94), which, in general, describe the case of different carrier density, disorder, and interaction strength on the top and bottom surfaces of a 3D TI film. The renormalization of interaction parameters γ_{11} and γ_{22}, Eqs. (96c) and (96d), determines four fixed planes of the RG flow: (1) $\gamma_{11} = -1 = \gamma_{22}$. Repulsive fixed plane of two decoupled surfaces with only intrasurface Coulomb interaction. This problem has been studied in Ref. 35. (2) $\gamma_{11} = 0$, $\gamma_{22} = -1$ or vice versa, a fixed plane describing a 3D TI film with strongly different surface population. This case in analyzed in Sec. V B1 below. (3) $\gamma_{11} = 0 = \gamma_{22}$, an attractive fixed plane. Intrasurface interaction has died out and only intersurface interaction survived. This case is analyzed in Sec. V B2 below.

Concerning the repulsive fixed planes, one should keep in mind that the renormalization of interaction amplitudes is suppressed by the small factor $1/\sigma$. Therefore, even if the conditions on γ_{11} and γ_{22} are only approximately fulfilled, the behavior in the fixed plane dictates the RG flow in a large temperature-frequency window. RPA estimates of the bare values of interaction amplitudes can be found in Appendix C 6.

We also remind the reader that the RG equations describing the model with the finite-range interaction (and thus the whole crossover between the problem with Coulomb interaction and the noninteracting system) is discussed in Appendix F.

1. *Strongly different surface population*

We investigate here the fixed plane of Eq. (94) in which $\gamma_{11} = 0$ and $\gamma_{22} = -1$. (Clearly, the reversed situation $\gamma_{11} = -1$ and $\gamma_{22} = 0$ is completely analogous.) Both fixed planes are "saddle-planes" of the RG flow, i.e., they are attractive in one of the γ directions and repulsive in the other.

Before analyzing this fixed plane, it is worth explaining why this limit is of significant interest for gate-controlled transport experiments, in particular, those on Bi_2Se_3. As for this material, the Fermi energy is normally located in the bulk conduction band, an electrostatic gate is conventionally used to tune the chemical potential into the bulk gap and hence to bring the system into a topologically nontrivial regime. A situation as depicted in Fig. 14 is then believed to arise in a certain range of gate voltages:[29] one of the two surfaces (here surface 1) is separated by a depletion region from a relatively thick bulk-surface layer.

Recently,[107–109] disorder-induced interference corrections for 3D TI bulk electrons have been investigated

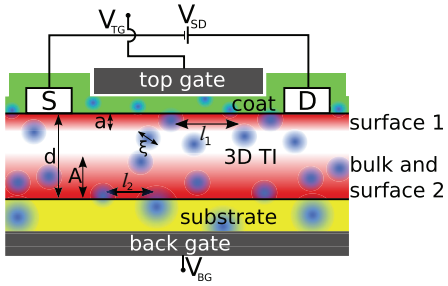

FIG. 14. (Color online) Typical scenario for gate-controlled transport experiments: a topologically protected surface separated from a thick bulk-surface layer.

theoretically;[125] while at small length scales, additional symmetries of the Hamiltonian provide nontrivial localization behavior, at sufficiently large scales, the usual WAL effect sets in. The strong coupling between electron states in the conducting part of the bulk and at surface 2 does not alter this universal low-energy property. In conclusion, at sufficiently large length scales, the symplectic class NLσM, Eq. (54), is the adequate description of such a system (under the assumption of negligible tunneling between surface 1 and the conducting part of the bulk).

Since the bulk-surface layer has a much higher carrier density than the carrier density on the spatially separated surface 1 we can expect that $\kappa_2 \gg \kappa_1$. Provided $\kappa_1 d \ll 1$, the electron-electron interaction on the spatially separated surface 1 is effectively screened out such that $|\gamma_{11}| \approx (\kappa_1/\kappa_2)(1 + 2\kappa_2 d) \ll 1$ [see Eq. (7)]. Conversely, the effect of screening by electrons on the surface 1 is negligible for Coulomb interaction of the bulk states: $1 + \gamma_{22} \approx \kappa_1/\kappa_2 \ll 1$.

Substituting $\gamma_{11} = 0$ and $\gamma_{22} = -1$ into Eq. (96), we find that the RG equations in this fixed plane are as follows:

$$\frac{d\sigma}{dy} = -\frac{2}{\pi}\left[1 - \frac{1}{t}\ln(1+t)\right], \tag{102a}$$

$$\frac{dt}{dy} = -\frac{2}{\pi}\frac{1+t}{\sigma}\left[\frac{1-t}{2} - \frac{1}{t}\ln(1+t)\right]. \tag{102b}$$

They can equivalently be written in terms of conductivities σ_1 and σ_2:

$$\frac{d\sigma_1}{dy} = -\frac{2}{\pi}\left[\frac{1}{2} - \frac{\sigma_2}{\sigma_1}\ln\left(1 + \frac{\sigma_1}{\sigma_2}\right)\right], \tag{103a}$$

$$\frac{d\sigma_2}{dy} = -\frac{1}{\pi}. \tag{103b}$$

We emphasize that the limit $\gamma_{11} = 0$ and $\gamma_{22} = -1$ is very peculiar. Indeed, due to the relation (65), this limit implies that the condition $z_1/z_2 = 0$ holds. Equations (102) and (103) are written under assumption that the ratio $t = \sigma_1/\sigma_2$ is finite in spite of the fact that $z_1/z_2 = 0$. In the experiment it corresponds to the case in which $\kappa_1/\kappa_2 \ll 1$ but the ratio $D_1/D_2 \gg 1$, where $D_s = \sigma_s/4z_s$ is the diffusion coefficient.

Equations (103) become decoupled for $\sigma_1/\sigma_2 = 0$. Then, as expected, $\delta\sigma_1 = \frac{1}{\pi}\ln L/l$ (WAL, no interaction on the surface 1) and $\delta\sigma_2 = -\frac{1}{\pi}\ln L/l$ (WAL and AA due to Coulomb interaction on the surface 2). However, the line $t = 0$ is unstable. As one can see from Eq. (102b), due to the very same quantum corrections the initially small parameter $t = \sigma_1/\sigma_2$

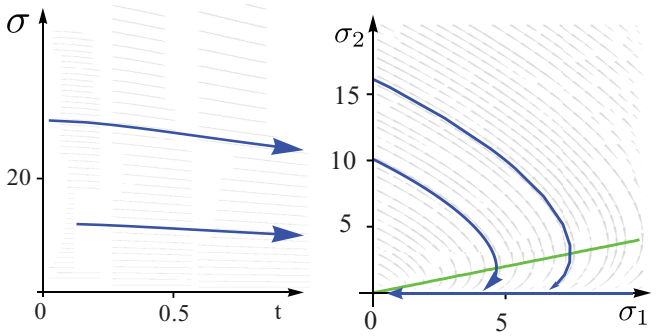

FIG. 15. (Color online) Perturbative RG flow in the fixed plane $\gamma_{11} = 0$, $\gamma_{22} = -1$. In the experimentally motivated scenario (see Fig. 14), the flow starts at $t = \sigma_1/\sigma_2 \ll 1$. The green line in the right panel is a line of zeros of the right-hand side of Eq. (103a); it determines the maximum in the RG flow of σ_1.

increases under RG. The ultimate limit of the perturbative RG flow is $\sigma \to 0$ and $t \to \infty$, see Fig. 15. The scale dependence of σ_1 is nonmonotonous; the position of the corresponding maximum is determined by zeros of the right-hand side of Eq. (103a) shown by a green line in the right panel of Fig. 15.

As has been already emphasized, the perturbative RG equations are sufficient only in the regime of large σ_s. We now discuss the topological effects at small values of conductivities. In the limit $\gamma_{11} = 0$, $\gamma_{22} = -1$ the renormalization of σ_2 is *exactly* independent of the surface 1. Indeed, in the conductivity corrections, the two surfaces influence each other only via mutual RPA screening. In the NLσM description, the interaction amplitudes in the full action (54) and hence in the propagators (80) (diffusons and cooperons) fully account for this effect. Since the layer 2 includes a single TI surface, we know that σ_2 is topologically protected and flows towards σ_2^* of the order of the quantum of conductance ("interaction-induced criticality"[35]). Before this happens, the flow of σ_1 becomes reversed from antilocalizing to localizing, see Eq. (103a). However, since the surface 1 is also topologically protected, its states can not be strongly localized and $\sigma \to \sigma_1^* > 0$.[126] Thus both surfaces are at the quantum critical points with conductivities of order e^2/h. The conclusion concerning the surface 1 is particularly remarkable: even though $\gamma_{11} = 0$, there is "intersurface-interaction-induced criticality" on the surface 1.

2. Attractive fixed plane

According to Eqs. (96c) and (96d), any $\gamma_{ss} \notin \{0, -1\}$ is renormalized to zero. The $\gamma_{11} = \gamma_{22} = 0$ is thus an attractive fixed plane of the general RG flow. The flow within this plane has the form determined by the following RG equations:

$$\frac{d\sigma}{dy} = -\frac{2}{\pi}\left[1 - \frac{1}{t}\ln(1+t) - t\ln\left(1+\frac{1}{t}\right)\right], \tag{104a}$$

$$\frac{dt}{dy} = -\frac{2}{\pi}\frac{1+t}{\sigma}\left[\frac{1-t}{2} - \frac{1}{t}\ln(1+t) + t^2\ln\left(1+\frac{1}{t}\right)\right], \tag{104b}$$

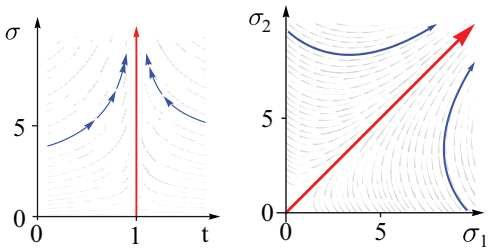

FIG. 16. (Color online) The RG flow in the attractive fixed plane $\gamma_{11} = 0 = \gamma_{22}$. The zero of Eq. (104b) is displayed by the red line.

or, equivalently,

$$\frac{d\sigma_1}{dy} = -\frac{2}{\pi}\left[\frac{1}{2} - \frac{\sigma_2}{\sigma_1}\ln\left(1+\frac{\sigma_1}{\sigma_2}\right)\right], \tag{105a}$$

$$\frac{d\sigma_2}{dy} = -\frac{2}{\pi}\left[\frac{1}{2} - \frac{\sigma_1}{\sigma_2}\ln\left(1+\frac{\sigma_2}{\sigma_1}\right)\right]. \tag{105b}$$

Even though the single-surface conductivities σ_s display nonmonotonic behavior within this plane, eventually all quantum corrections are antilocalizing, see Fig. 16. The ratio of conductivities flows to the symmetric situation $t = \frac{\sigma_1}{\sigma_2} = 1$, as has been discussed in Sec. V A. We reiterate that at the corresponding fixed line the WAL effect is competing with a contribution of the opposite sign due to intersurface interaction. While the WAL wins, the antilocalizing flow is slower than for free electrons, see Eq. (99).

3. General RG flow

After having analyzed the RG flow in various fixed planes, we briefly discuss the general RG flow. According to Eqs. (94c) and (94d), there is a single attractive fixed point of the overall RG flow—the metallic fixed point with zero intrasurface interaction, $\sigma_1 = \sigma_2 \to \infty$ and $\gamma_{11} = \gamma_{22} = 0$. On the other hand, for the values of γ_{ss} close to -1, the corresponding conductivity σ_s is first subjected to localizing quantum corrections and will thus show a nonmonotonic behavior towards antilocalization. There also exists a range of initial parameters for the RG flow for which the conductivity at one surface demonstrates monotonous antilocalizing behavior, while the conductivity in the other surface flows in the described nonmonotonous manner.

VI. DISCUSSION AND EXPERIMENTAL PREDICTIONS

In the preceding section, we have performed a general analysis of the renormalization group flow determined by the RG Equations (94). The purpose of the present section is to apply these results to specific experimentally relevant materials.

A. Parameters

As explained in Sec. II A, the RG Equations (94) apply in the case of the following hierarchy of length scales:

$$l \ll L_E, \tag{106a}$$

$$d \ll l. \tag{106b}$$

In order to deal with q-independent interaction amplitudes, an additional requirement occurs in the case $\kappa_s d \ll 1$ for both $s = 1$ and $s = 2$:

$$l_{\rm scr} \ll L_E. \tag{106c}$$

In view of condition (106a), the constraint (106c) is fulfilled in the entire diffusive regime if $l_{\rm scr} \ll l$.

Further, we have assumed that the intersurface tunneling is negligible; the corresponding condition reads

$$a \ll d. \tag{106d}$$

In this section, we will concentrate on the case when the RG scale is set by temperature, $L_E = l_T$. We recall the definition of the length scales entering the above conditions: $l = \max_{s=1,2} l_s$ is the larger mean-free path, $l_T = \min_{s=1,2} \sqrt{D_s/kT}$ the smaller thermal length, d the sample thickness, a the penetration depth, κ_s the inverse Thomas-Fermi screening length for the surface s and $l_{\rm scr}$ the total screening length for the 3D TI film. The situation in which only one of the two surfaces is in the diffusive regime while the other one is in the ballistic regime (i.e., $T\tau_1 \ll 1$ and $T\tau_2 \gg 2$ or vice versa) is also a conceivable and interesting scenario. However, we do not address it in the present paper.

The effect of intersurface interaction becomes prominent if the sample thickness does not exceed too much at least one of the single surface screening lengths κ_s^{-1}. As discussed above (see Sec. V), this condition implies that the bare values of interaction γ_{11} and γ_{22} are not too close to -1.

It is useful to present expressions for the length scales appearing in the conditions (106a)–(106d) in terms of standard parameters characterizing samples in an experiment. For simplicity, we assume $v_F^{(1)} = v_F^{(2)}$ and $\tau_1 = \tau_2$ in these formulas.

The densities of states (DOS) and inverse screening lengths for the top and bottom surfaces are

$$\nu_s = \sqrt{\frac{n_s}{\pi v_F^2}}, \qquad \kappa_s \equiv \frac{2\pi e^2}{\epsilon_2}\nu_s = 2\pi\alpha\sqrt{\frac{n_s}{\pi}}, \tag{107}$$

where n_s are the corresponding electron densities. If the electron densities for each surface separately are not known, the total density $n_{\rm tot} = n_1 + n_2$ can be used to estimate the DOS and the screening lengths:

$$\nu_1^2 + \nu_2^2 = \frac{n_{\rm tot}}{\pi v_F^2}, \qquad \kappa_1^2 + \kappa_2^2 = (2\pi\alpha)^2\frac{n_{\rm tot}}{\pi}. \tag{108}$$

The mean-free path can be expressed as

$$l = v_F\tau_{\rm tr} = \frac{\sigma}{\pi v_F(\nu_1 + \nu_2)}. \tag{109}$$

The thermal length in the diffusive regime is given by

$$l_T = \sqrt{\frac{D}{kT}} = \sqrt{\frac{v_F l}{2kT}} = \sqrt{\frac{\sigma}{kT 2\pi(\nu_1 + \nu_2)}}. \tag{110}$$

Hence the condition (106a) is fulfilled for temperatures

$$kT \ll kT_{\rm Diff}, \tag{111}$$

where

$$kT_{\rm Diff} = \frac{v_F}{2l} = \frac{1}{\sigma[e^2/h]}\left(\frac{v_F^2}{2}\right)2\pi(\nu_1 + \nu_2) \tag{112}$$

is the temperature scale at which the diffusion sets in.

In order to obtain $l_{\rm scr}$ entering Eq. (106c), we have to consider the full (inter and intrasurface) Coulomb interaction, see Appendix B. As explained in Sec. II B, it is only a meaningful quantity provided $\kappa_s d \ll 1$. Taking into account the influence of the surrounding dielectrics, we find

$$l_{\rm scr} = \frac{\epsilon_1 + \epsilon_3}{2\epsilon_2}\frac{1}{\kappa_1 + \kappa_2}. \tag{113}$$

When deriving Eq. (113), we assumed for simplicity that $\epsilon_2 \lesssim \epsilon_1 + \epsilon_3$. Regarding the experimental setups discussed in Sec. VI B, this condition is well fulfilled for Bi_2Se_3 but only marginally for HgTe. Thus in the latter case, Eq. (113) should be considered as a rough estimate.

Finally, to check the validity of the condition (106d), one needs to know the value of the penetration depth a. The latter can be estimated from the condition

$$\frac{v_{F,\perp}p_\perp}{\Delta_{\rm bulk}} \sim 1, \tag{114}$$

where $p_\perp \sim 1/a$ denotes typical momenta perpendicular to the surface. Provided $v_{F,\perp} \sim v_F$, it yields

$$a \sim \frac{v_F}{\Delta_{\rm bulk}}. \tag{115}$$

We are now going to consider two exemplary materials for 3D TIs: Bi_2Se_3 and strained HgTe. We shall estimate numerically all the relevant parameters and present characteristic plots for temperature dependence of conductivities.

B. Exemplary 3D TI materials

1. Bi_2Se_3

Bi_2Se_3 is currently the most conventional material for experimental realization of the 3D TI phase. Typical experimental data (extracted from the point of the minimal conductance in Refs. 31 and 110) are summarized in the upper part of Table I. Using Eqs. (107)–(115), we can estimate the

TABLE I. Experimental values of sample parameters at the point of the minimal carrier density and associated length scales for transport experiments on Bi_2Se_3 films of Refs. 31 and 110. The dots "..." separate values for the symmetric ($n_1 = n_2$) and totally asymmetric ($n_1 = n_{\rm tot}, n_2 = 0$) cases. The bare interaction amplitudes are estimated in the random phase approximation (RPA).

Fermi velocity	$v_F \sim 5 \times 10^5$ m/s
Bulk gap	$\Delta_{\rm bulk} \sim 0.3$ eV
Sample thickness	$d \sim 10$ nm
	Coat: $\epsilon_1 \sim 1$
Dielectric properties	3D TI (Bi_2Se_3): $\epsilon_2 \sim 100$
	Substrate ($SrTiO_3$): $\epsilon_3 \sim 10^3$–10^4
Carrier density	$n_{\rm tot} \sim 3 \times 10^{12}$ cm^{-2}
Mobility	$\mu_{el} \sim 100 \ldots 1000$ cm^2/V s
Sheet resistance	$1/\sigma \sim 0.097\, h/e^2$ at $T \sim 50$ mK
Effective coupling	$\alpha \sim 4 \times 10^{-2}$
Chemical potential	$\mu_1^2 + \mu_2^2 = (0.2\,\text{eV})^2$
Penetration depth	$a \sim 1$ nm
Mean-free path	$l \sim 24 \ldots 34$ nm
Diff. temperature	$T_{\rm Diff} \sim 80 \ldots 57$ K
Screening length	$\kappa_1^2 + \kappa_2^2 \sim (37\text{ nm})^{-2}$
Scr. length (total)	$l_{\rm scr} \sim 132 \ldots 186$ nm, for $\epsilon_3 = 10^3$
Bare interaction (RPA)	top surface: $\gamma_{11} \sim -0.6 \cdots -1$
	bottom surface: $\gamma_{22} \sim -0.6 \ldots 0$

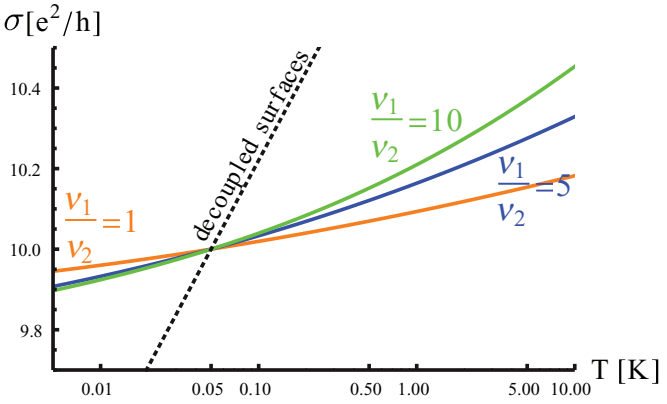

FIG. 17. (Color online) Theoretical prediction for the temperature dependence of the total conductivity in thin Bi_2Se_3 films.

hierarchy of length scales (lower part of the same table). One can see that all of the requirements of validity for our theory are fulfilled for length scales above $l_{\rm scr}$ [temperatures below $T_{\rm max} = 2.6 \ldots 1.9$ K, see condition (106c)].[127]

From the experimental data, the ratio of carrier densities is not known. Therefore we show in Fig. 17 the expected temperature dependence of total conductivity for various values of this ratio. Clearly, the behavior strongly differs from the case of decoupled surfaces (dashed line). First, the slope of $d\sigma/d\ln T$ is considerably smaller. Second, one observes a clear curvature of the dependence $\sigma(\ln T)$, which is a manifestation of the nonmonotonicity. (For the parameters used in the plot, the minimum of σ occurs at still lower temperatures.) This curvature is especially pronounced for strongly different surfaces.

It should be mentioned that the substrate used in Ref. 31 has a strongly temperature-dependent dielectric function ϵ_3 since $SrTiO_3$ approaches a ferroelectric transition at low temperatures. This could result in a temperature dependence of effective gate voltage and consequently of carrier density. The resulting classical temperature dependence of conductivity (and interaction constants) would mask the quantum effects described in our analysis. However, in the presence of the gating field, the temperature dependence of ϵ_3 saturates at low temperatures. This motivates the presentation in Fig. 17 where we assumed independent of temperature $\epsilon_3 = 1000$.

2. *Strained HgTe*

Another very promising 3D TI material is strained HgTe. The presence of Dirac-like surface states was experimentally confirmed by the odd series of QHE plateaus, as well as by ARPES.[25] While the transport experiment indicates dominant surface conduction, the extracted carrier density appears to be too large for a pure surface theory with linear spectrum, yielding the value of the chemical potential μ larger than the gap $\Delta_{\rm bulk}$, see Table II. (The role of the bulk conduction band as well as the parabolic bending of the dispersion was also discussed within an independent magnetooptical study by the same experimental group.[112]) Thus it remains to be clarified under what experimental conditions the strained HgTe sample is in the true TI regime (i.e., the bulk contribution to transport is negligible). Notwithstanding this point and motivated by the excellent surface transport data, we apply our theory to the HgTe experiment, see Fig. 18. In spite of the considerable thickness of the probe, the effect of intersurface interaction is clearly visible: the slope of $d\sigma/d\ln T$ is considerably smaller than it is expected for decoupled surfaces.

TABLE II. Typical experimental values for transport experiments on HgTe films of Refs. 25 and 111.

Fermi velocity	$v_F \sim 5 \times 10^5$ m/s
Bulk gap	$\Delta_{\rm bulk} \sim 0.022$ eV
Sample thickness	$d \sim 70$ nm
Dielectric properties	Coat: $\epsilon_1 \sim 1$
	3D TI (HgTe): $\epsilon_2 \sim 21$
	Substrate (CdTe): $\epsilon_3 \sim 10$
Carrier density	top surface: $n \sim 4.8 \times 10^{11}$ cm^{-2}
	bottom surface: $n \sim 3.7 \times 10^{11}$ cm^{-2}
Mobility	$\mu_{el} \sim 34000$ cm^2/V·s
Sheet resistance	$1/\sigma \sim 0.04\, h/e^2$ at $T = 50$ mK
Effective coupling	$\alpha \sim 0.21$
Chemical potential	top surface: $\mu_1 \sim 0.08$ eV
	bottom surface: $\mu_2 \sim 0.07$ eV
Penetration depth	$a \sim 15$ nm
Mean-free path	$l \sim 108$ nm
Diff. temperature	$T_{\rm Diff} \sim 18$ K
Screening length	top surface: $\kappa_1^{-1} \sim 19.53$ nm
	bottom surface: $\kappa_2^{-1} \sim 22.24$ nm
Bare interaction (RPA)	top surface: $\gamma_{11} \sim -0.893$
	bottom surface: $\gamma_{22} \sim -0.878$

C. Hallmarks of surface transport and interactions

We briefly summarize now our most salient predictions for experimental signatures of surface transport in 3D TI with an intersurface interactions.

(1) As already exploited in 3D TI experiments,[31] the magnetoconductance formula[100] for the total conductivity is

$$\delta\sigma(B) = -\frac{e^2}{2\pi h}\sum_{s=1,2}\left[\psi\left(\frac{1}{2}+\frac{B_\phi^{(s)}}{B}\right) - \ln\left(\frac{B_\phi^{(s)}}{B}\right)\right], \tag{116}$$

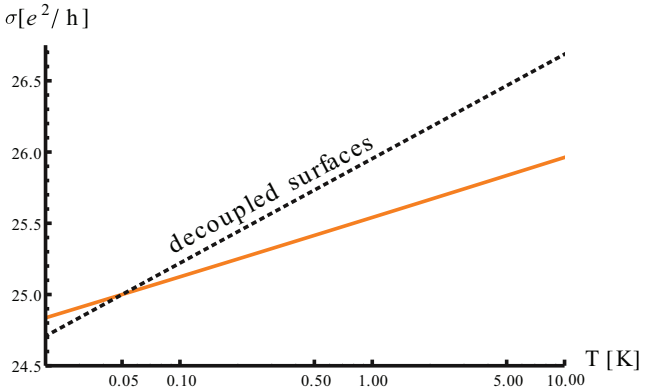

FIG. 18. (Color online) Theoretical prediction for the temperature dependence of the total conductivity in thin films of strained HgTe.

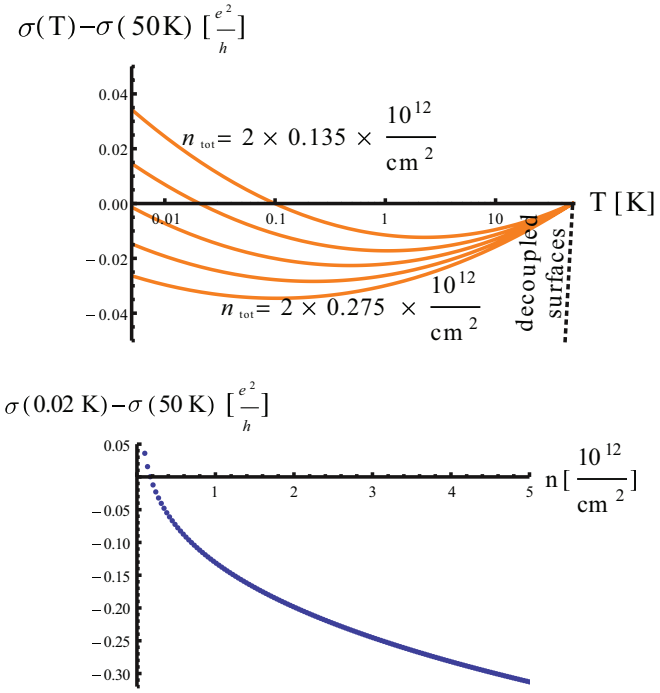

FIG. 19. (Color online) (Top) Conductivity corrections for low carrier concentration. The total electron concentration in units of 10^{12} cm^{-2} is equal to 0.55,0.48,0.41,0.34,0.27 from bottom to top. The characteristic nonmonotonous behavior is clearly seen; deviations from the behavior of decoupled surfaces are very strong. (Bottom) Carrier-density dependence of conductivity corrections. The nontrivial dependence is entirely due to the intersurface interaction: in the case of the decoupled surfaces, the conductivity correction would be constant as a function of density, $\sigma(0.02\text{K}) - \sigma(50\text{K}) \approx -2.49e^2/h$. We used the values of the parameters d, v_F, and α as in Table I for Bi_2Se_3. Further, we assumed the case of equal surfaces ($n_{\text{tot}} = 2n$) and $T_{\text{Diff}} = 1/2\tau = 50$K.

where the characteristic magnetic field $B_\phi^{(s)} = \hbar/[4eD_s^{(s)}\tau_\phi^{(s)}]$ is determined by the diffusion coefficient $D_s^{(s)}$ and the phase breaking time $\tau_\phi^{(s)}$ for the surface s. The function ψ denotes the digamma function here.

(2) The characteristic effect of intersurface interaction is the nonmonotonous temperature dependence of conductivity (see Fig. 19, top). It may happen that in experimentally accessible temperature window this effect manifests itself only as a deviation of the conductance slope

$$\delta\sigma(T) = \frac{e^2}{h} c \ln T \tag{117}$$

from the value $c = 1/\pi$ characteristic for two decoupled surfaces accompanied by some bending of the curve $\sigma(\ln T)$, see Figs. 17 and 18. The ultimate low-T behavior of the coupled system is always antilocalizing and following the universal law

$$\delta\sigma(T) = \frac{e^2}{\pi h}(1 - 2\ln 2)\ln T. \tag{118}$$

However, depending on the parameters, this asymptotics may become valid at very low temperatures only.

(3) The strength of intersurface interaction is governed by the parameters $\kappa_1 d$ and $\kappa_2 d$, where κ is the screening length. Therefore, in contrast to usual, single surface conductivity corrections, the predicted effect strongly depends on the carrier density (see Fig. 19, bottom). It is also possible to access the intersurface induced quantum corrections in the frequency dependence of the ac conductivity [by the simple replacement $T \to \omega$ in $\delta\sigma(T)$ if $\omega \gg T$].

VII. CONCLUSIONS

In this paper, we have investigated interference and interaction effects in the surface state conductivity of 3D topological insulator slabs. We have taken into account the electron-electron interaction within the top and bottom surfaces of a slab and between them. These two surfaces were in general assumed to be characterized by different carrier densities and scattering rates, and by asymmetric dielectric environment.

Our field-theoretical analysis was based on the interacting nonlinear σ model approach describing the system at length scales above the mean free path. We demonstrated how this effective theory can be obtained from the non-Abelian bosonization. In particular, we have shown that upon inclusion of potential disorder the Wess-Zumino term generates a local expression for the $\mathbb{Z}_2$ theta term. The appearance of this topological term is the hallmark of the Dirac surface states; it is absent in conventional 2D metals of the same symmetry class. We have further analyzed the **U**(1)-gauged σ model that describes a coupling to the external electromagnetic field. This has allowed us to connect the physical linear-response characteristics of the problem and the σ model coupling constants. We have also analyzed the effect of breaking of time-reversal symmetry, namely, the anomalous quantum Hall effect of Dirac electrons.

It is worth emphasizing that our theory treats the general situation of potentially strong interactions and thus went beyond perturbation theory. We have thus developed the Fermi liquid theory of the strongly correlated double layer system in the ballistic and diffusive regime.

We renormalized the interacting NLσM of the two surfaces in the one-loop approximation and obtained the RG equations, Eq. (94). This way we have determined the temperature (or else, frequency, or length scale) dependence of the conductivities of both surfaces. The RG is controlled by a large conductivity, $k_F l \gg 1$. Our calculations are exact in the singlet interaction amplitudes, while contributions due to a repulsive Cooper interaction are parametrically small and can be neglected.

Inspecting the RG equations, we showed that intersurface interaction is relevant in the RG sense and the limiting case of decoupled surfaces is therefore unstable. The rich flow diagram has been analyzed in detail. For fully decoupled surfaces the system flows into an intermediate-coupling fixed point ("interaction-induced criticality"). This point is, however, unstable with respect to the intersurface coupling. The flow is then towards a single attractive fixed point which is "supermetallic," $\sigma \to \infty$, and at which even originally different surfaces have the same transport properties, $\sigma_1 = \sigma_2$, see Figs. 11 and 16. Further, this fixed point is characterized by vanishing intrasurface and finite intersurface interactions. Typically, this fixed point is reached via a characteristic nonmonotonous temperature dependence of conductivity.

Our perturbative results are equally applicable to weak topological insulator[113,114] thin films and to nontopological double layer systems with spin-orbit interaction. For the latter type of structures, we have also discussed the difference compared to the strong TI films, which is in nonperturbative topological effects (see a comparison of the flow diagrams in Fig. 13). While in the TI case these effects lead to a topological protection of the surface states from Anderson localization, a conventional (nontopological) double layer system undergoes a metal-insulator transition, which is tuned by the ratio of interlayer distance and screening length. Finally, we have estimated parameters and presented explicit predictions for the temperature dependence of the conductivity for typical experimental setups based on Bi_2Se_3 and strained HgTe materials.

Before closing, we discuss perspectives for further research. First, experimental studies of temperature dependence of conductivity of 3D topological insulators for different positions of chemical potentials would be highly useful. A comparison of such experimental data with our theoretical predictions would allow one to judge whether the system is in the truly topological phase. Second, more work is needed on effects of local breaking of time-reversal symmetry in TI slabs. Third, it is known that Coulomb interaction in electronically decoupled double-layer systems induces a finite but typically small transconductance σ_{12}.[56,115–117] However, the side walls of 3D TI films connect the two major surfaces, which might be a serious obstacle for performing Coulomb drag experiments. Fourth, in view of recent experimental progress,[118] it would be interesting to perform an RG analysis for a superconducting counterpart of the system that we have explored, namely, for surface states of a 3D topological superconductor with spin-orbit interaction (class DIII).[119]

ACKNOWLEDGMENTS

We thank J. Smet, Y.Q. Li, A. Finkel'stein, M. Schütt, and U. Briskot for useful discussions. The work was supported by BMBF, DAAD, DFG SPP 1666, RAS programs "Quantum mesoscopic and disordered systems," "Quantum Physics of Condensed Matter," "Fundamentals of nanotechnology and nano materials," RFBR grant Nos. 11-02-12126 and 12-02-00579, Russian President Grant No. MK-4337.2013.02, Dynasty foundation, and by the Russian Ministry of Education and Science under Contract No. 8678.

APPENDIX A: NON-ABELIAN BOSONIZATION AND THE TOPOLOGICAL TERM

In this Appendix, we include more detailed calculations concerning non-Abelian bosonization, the gauged WZNW model and the topological term $S^{(\theta)}$. For brevity, we omit the surface index s in this appendix.

1. Gauged WZNW model

The Wiegmann-Polyakov formula[84] allows the inclusion of smooth $\mathfrak{o}(2_\tau \times 2N_M \times N_R)$ gauge fields A_μ. A generalization to potentially topological gauge potentials can be found in Refs. 86–89. The gauged WZNW model is given as[86–89]

$$S[O,A_\mu] = -\frac{1}{16\pi}\int_x \mathrm{tr}(O^T D_\mu O)(O^T D_\mu O) \tag{A1a}$$

$$+\frac{i\epsilon_{\mu\nu\rho}}{24\pi}\int_{x,w}\mathrm{tr}[(O^T D_\mu O)(O^T D_\nu O)(O^T D_\rho O)] \tag{A1b}$$

$$-\frac{i\epsilon_{\mu\nu\rho}}{16\pi}\int_{x,w}\mathrm{tr}[F_{\mu\nu}(O^T D_\rho O + D_\rho O O^T)], \tag{A1c}$$

where we introduced $D_\mu = \partial_\mu + [A_\mu,\cdot]$ and $F_{\mu\nu} = [D_\mu, D_\nu]$. In the main text, we were mostly interested in $\mathbf{U}(1)$ gauge fields $A_\mu = i\hat{A}^T\frac{1+\tau_y}{2} - i\hat{A}\frac{1-\tau_y}{2}$. (In this appendix, the electron charge is absorbed into the vector potential.)

To obtain the Wiegmann-Polyakov formula, one can use the following identity:[89]

$$\begin{aligned}(A1b) = {}& \frac{i}{24\pi}\Gamma[O] - \frac{i}{8\pi}\int_{x,w}\epsilon_{\mu\nu\rho}\partial_\mu \mathrm{tr}[O A_\nu O^T A_\rho \\ & + A_\nu(O^T\partial_\rho O + \partial_\rho O O^T)] \\ & + \frac{i\epsilon_{\mu\nu\rho}}{16\pi}\int_{x,w}\mathrm{tr}[F_{\mu\nu}(O^T D_\rho O + D_\rho O O^T)].\end{aligned} \tag{A2}$$

While the last integral in Eq. (A2) compensates the term (A1c), the total derivative term yields the Wiegmann-Polyakov formula provided A_μ is not singular:

$$\begin{aligned}S[O,A_\mu] = {}& S[O] + \frac{1}{8\pi}\int_x \mathrm{tr}[A_\mu(O\partial_\mu O^T + O^T\partial_\mu O) \\ & + A_\mu O^T A_\mu O - A_\mu^2 - i\epsilon_{\nu\rho}O A_\nu O^T A_\rho \\ & - i\epsilon_{\nu\rho}A_\nu(O^T\partial_\rho O + \partial_\rho O O^T)]\end{aligned} \tag{A3}$$

$$\begin{aligned}= {}& S[O] + \frac{1}{8\pi}\int_x \mathrm{tr}[A_-(O\partial_+ O^T) + A_+(O^T\partial_- O) \\ & + A_+ O^T A_- O - A_+ A_-].\end{aligned} \tag{A4}$$

Here we have introduced the (anti-)holomorphic combination of gauge potentials $A_\pm = A_x \pm i A_y$. In the case of topological gauge potentials, the integral over the total derivative yields also a contribution from the Dirac string.

Equation (A4) is a very powerful result. In particular, it justifies a posteriori the bosonization rules (25a) and (25b). Also, it follows immediately from expression (A3) that after disorder-induced symmetry breaking ($O \to Q = Q^T$) the gauge-field-dependent contributions from the topological term vanish.

Further, one can use Eq. (A4) to determine the prefactor of the kinetic term in the AII NLσM, Eq. (36). As explained in the main text, soft rotations $O_{\mathrm{soft}}^T O O_{\mathrm{soft}}$ of the WZNW fields O are not affected by disorder induced masses, see Eq. (30). The effective action for topologically trivial Goldstone modes contains

$$\begin{aligned}S_{\mathrm{eff,kin}}[\Phi_\mu] = {}& \frac{1}{8\pi}\int_x \langle \mathrm{tr}(\Phi_+ O^T \Phi_- O - \Phi_+\Phi_-)\rangle \\ & - \frac{1}{2}\left\langle\left[\int_x \mathrm{tr}(\Phi_+ j_- + \Phi_- j_+)\right]^2\right\rangle,\end{aligned} \tag{A5}$$

where $j_\pm$ are the (bosonic) currents, $\langle \dots \rangle$ denotes average with respect to the full bosonic theory (including the mass terms) and $\Phi_\pm = O_{\rm soft}\partial_\pm O_{\rm soft}^T$. To the leading order, the average can be calculated close to the saddle point. Exploiting the equivalence of bosonic and fermionic theories, one can equally evaluate $\langle \dots \rangle$ using the fermionic fields at SCBA level. At $|\mu|\tau \gg 1$, the major contribution comes from the second line of Eq. (A5), which, taking the vertex corrections into account, yields the correct prefactor (i.e., the conductivity) of the kinetic term in Eq. (36).

2. Instanton configuration

We define the following four-dimensional unit vector:

$$\begin{aligned}\underline{a} &\equiv (a_0,a_1,a_2,a_3)\\ &\equiv \frac{1}{|\vec{x}-\vec{x}'|^2+\lambda^2}[2\lambda(\vec{x}-\vec{x}'),|\vec{x}-\vec{x}'|^2-\lambda^2],\end{aligned}$$

where the $1+2$ vector $\vec{x}-\vec{x}' \equiv [(1-w)/w, \boldsymbol{x}-\boldsymbol{x}']$ contains the extension parameter and the real-space coordinates. It describes a topological excitation at position $(1,\boldsymbol{x}')$ in a three-dimensional base space. With the help of the vector $\underline{a}$ we can define the following extended field configuration:

$$\tilde{O}_{\rm inst} = \begin{pmatrix} -a_0 i\tau_y + a_3 & 0 & a_1 + a_2 i\tau_y & 0\\ 0 & \mathbf{1} & 0 & 0\\ a_1 - a_2 i\tau_y & 0 & -a_0 i\tau_y - a_3 & 0\\ 0 & 0 & 0 & -\mathbf{1}\end{pmatrix}. \tag{A6}$$

For $a_0 = 0$, i.e., on the physical space $w = 1$, $\tilde{O}_{\rm inst}$ is a symmetric matrix and characterizes the two-dimensional instanton. The choice of the extension is arbitrary, but, as has been stressed in the main text, the $\tilde{O}_{\rm inst}$ field has to leave the diffusive saddle-point manifold for some subinterval $w \in I \subseteq (0,1)$. For $w \to 0$, the extended field $\tilde{O}_{\rm inst}$ satisfies the boundary condition $\tilde{O}(\boldsymbol{x}, w = 0) = \Lambda = {\rm const}$.

We are now in the position to insert the instanton configuration into the WZNW term. After tracing out the matrix degrees of freedom this leads to

$$\begin{aligned} iS^{(\theta)} &= \frac{-i}{6\pi}\int_{x,w} \epsilon_{\mu\nu\lambda}(\epsilon_{abc}\, a_a\partial_\mu a_b\partial_\nu a_c\partial_\lambda a_0\\ &\quad -\epsilon_{abd}\, a_a\partial_\mu a_b\partial_\nu a_0\partial_\lambda a_d + \epsilon_{cda}\, a_a\partial_\mu a_0\partial_\nu a_c\partial_\lambda a_d\\ &\quad -\epsilon_{cdb}\, a_0\partial_\mu a_b\partial_\nu a_c\partial_\lambda a_d)\\ &= i\pi. \end{aligned} \tag{A7}$$

Here, the last line is obtained by a straightforward calculation. We have thus shown that the topological term distinguishes between the trivial and the nontrivial sectors as it acquires on them the values 0 and $i\pi$ $(\mathrm{mod} 2\pi i)$, respectively.

APPENDIX B: EFFECT OF DIELECTRIC ENVIRONMENT ON COULOMB INTERACTION

1. Electrostatic potential and single-particle effects

As has been stated above, the experimental setup consists of a sandwich of (at least) three different dielectrics (see Fig. 1). We define the z axis to be perpendicular to the two surfaces. The sandwich consists of the coating material with a dielectric constant ϵ_1 (for $d/2 < z$), the topological insulator film with a dielectric constant ϵ_2 (for $-d/2 \leqslant z \leqslant d/2$), and the substrate with a dielectric constant ϵ_3 (for $z < -d/2$). Taking these different dielectric properties into account, we here present the expression for the Coulomb potential that generalizes Eq. (4).

By the method of mirror charges, one can derive[54–56] the electrostatic potential induced by a single point charge e located at $(\boldsymbol{x}_0,z_0) = (0,0,z_0)$ inside the middle region of the sandwich $(z,z_0 \in [-\frac{d}{2},\frac{d}{2}])$:

$$\begin{aligned}\Phi(\boldsymbol{x},z,z_0) = \frac{e}{\epsilon_2}\Bigg\{&\frac{1}{\sqrt{\boldsymbol{x}^2+(z-z_0)^2}} + r_{23}^{-1}\mathcal{F}[\boldsymbol{x},d+(z+z_0)]\\ &+ r_{21}^{-1}\mathcal{F}[\boldsymbol{x},d-(z+z_0)]\\ &+ \mathcal{F}(\boldsymbol{x},z-z_0) + \mathcal{F}[\boldsymbol{x},-(z-z_0)]\Bigg\},\end{aligned} \tag{B1}$$

where

$$\mathcal{F}(\boldsymbol{x},z) = \sum_{k=1}^{\infty}\frac{(r_{21}r_{23})^k}{\sqrt{\boldsymbol{x}^2+(z-2dk)^2}} \tag{B2}$$

and the ratios

$$r_{21} \equiv \frac{\epsilon_2-\epsilon_1}{\epsilon_2+\epsilon_1} \quad \text{and} \quad r_{23} \equiv \frac{\epsilon_2-\epsilon_3}{\epsilon_2+\epsilon_3}$$

were introduced. If one of these ratios vanishes, the textbook limit of two dielectric half-planes follows. A Fourier transformation of the $\boldsymbol{x}$ coordinates yields

$$\begin{aligned}\Phi(\boldsymbol{q},z,z_0) = \frac{2\pi e}{q\epsilon_2}\Bigg(&e^{-|z-z_0|q} + \frac{e^{-2dq}}{1-r_{21}r_{23}e^{-2dq}}\\ &\times\{r_{21}e^{(d+z+z_0)q} + r_{23}e^{(d-z-z_0)q}\\ &+ 2r_{21}r_{23}\cosh[(z-z_0)q]\}\Bigg).\end{aligned} \tag{B3}$$

We consider now 3D TI surface states: the charges are located at a typical distance $a \sim v_F/\Delta_{\rm bulk}$ (the penetration depth) from the boundaries $z = \pm\frac{d}{2}$. The consequences of the general expression (B3) on the 3D TI surface states are twofold.

First, there is a single-particle effect, stemming from the interaction of the charged particles with their own mirror charges. The associated electrostatic energy is incorporated in the chemical potential in the main text and can be expressed as

$$\begin{aligned}\Delta\mu_1 &= \frac{e}{2}\Phi^{\rm reg}\left(0,\frac{d}{2}-a,\frac{d}{2}-a\right)\\ &= \frac{e^2}{4\epsilon_2}\left[\frac{r_{21}}{a} - \frac{r_{21}+r_{21}^{-1}+2}{d}\ln(1-r_{21}r_{23})\right].\end{aligned} \tag{B4}$$

The analogous shift of the chemical potential at the second surface $\Delta\mu_2$ is easily obtained by interchanging $r_{21} \leftrightarrow r_{23}$. The superscript "reg" indicates that self-interaction of the charges is subtracted. In the second term, we used the approximation $a \ll d$. The first term, i.e., the interaction with the nearest mirror charge, is typically the dominating contribution $\Delta\mu_1 \approx \alpha_2 r_{21}/4\Delta_{\rm bulk}$.

Second, the electrostatic energy associated with two-particle interaction is the quantity $\underline{U}_0$ entering $S_{\rm int}$ in Eq. (14). This leads to the interaction parameters analyzed below.

2. Interaction parameters

The interaction parameters are obtained by placing a test charge into Eq. (B3). We will present this effective Coulomb interaction in the surface space. The terms induced by intersurface interaction contain a factor $\exp(-qd)$ (q takes values between the IR and UV cutoffs, $q \in [L_E^{-1}, l^{-1}]$). As a result we have to distinguish between the following two cases.

In the first case, the momenta are large ($qd \gg 1$) throughout our RG-procedure if $dL^{-1} > 1$ or for part of it if $d \in [l, L]$. Then the two surfaces become decoupled and

$$\underline{U}_0 = \frac{2\pi}{q} \begin{pmatrix} \frac{2}{\epsilon_2+\epsilon_1} & 0 \\ 0 & \frac{2}{\epsilon_2+\epsilon_3} \end{pmatrix}. \tag{B5}$$

(Here and in all subsequent appendices, we drop the electron charge, it is formally included into a redefinition of $\epsilon_1, \epsilon_2, \epsilon_3$.)

In the second case, the momenta are small, $qd \ll 1$. As we shall be interested in the low-energy theory, we keep only the Fourier transformed terms that are not vanishing in the limit of small transferred momentum $qd \to 0$. All others are irrelevant in the RG sense. This way we obtain the true long-range Coulomb part

$$\underline{U}_C = \frac{2}{\epsilon_1+\epsilon_3} \frac{2\pi}{q} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \tag{B6}$$

As expected, it does no longer depend on ϵ_2. The limit we considered is the large-distance behavior in which the dominant part of the electric field lines lives in the dielectrics surrounding the film.

There are other contributions that do not vanish in the $qd \to 0$ limit. These are short-range interaction amplitudes introduced by the finite thickness of the film:

$$\underline{F}^{(d)} = -\frac{2\pi}{\epsilon_2} d \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - \frac{4\pi}{\epsilon_1+\epsilon_3} d \left[F_{\text{symm}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \underline{F}_M \right]. \tag{B7}$$

Here, we have defined the scalar

$$F_{\text{symm}} = (\epsilon_2 - \epsilon_1)(\epsilon_2 - \epsilon_3) \left[\frac{1}{2\epsilon_2^2} + \frac{1}{\epsilon_2(\epsilon_1+\epsilon_3)} \right] \tag{B8}$$

and the matrix

$$\underline{F}_M = \frac{1}{2\epsilon_2^2} \begin{pmatrix} (\epsilon_2+\epsilon_1)(\epsilon_2-\epsilon_3) & \epsilon_2^2 - \epsilon_1\epsilon_3 \\ \epsilon_2^2 - \epsilon_1\epsilon_3 & (\epsilon_2-\epsilon_1)(\epsilon_2+\epsilon_3) \end{pmatrix}, \tag{B9}$$

which both vanish in the limit of $\epsilon_1 = \epsilon_2 = \epsilon_3$. In summary, for coupled surfaces, we can write $\underline{U}_0 = \underline{U}_C + \underline{F}^{(d)}$.

The derivation of the above equations includes some subtleties. First, we derived the electric field configuration for a single point charge. Thus, in particular, we did not consider the metallic surfaces between the dielectrics. As in the theory of conventional metals, their effect will be incorporated in the field theoretical description of the model (see Appendix C). Second, we used the potential (B3) derived for charged particles at position z, z_0 and then moved them on the surface between the dielectrics from inside of the TI film ($z_0 = \pm d/2 \mp a \approx \pm d/2$ and equally for z). This requires that the (macroscopic) electrostatic theory of continuous, homogeneous dielectrics can be applied to electrons located at a distance a from the boundary. This is justified, as we are interested in the long-range behavior of the electric field. Furthermore, for Bi_2Se_3 it is known that a is of the order of a few nanometers,[52,120] hence one order of magnitude larger than the atomic scale.

APPENDIX C: CLEAN FERMI LIQUID

In this appendix, we present the formal resummation of scattering amplitudes following Refs. 79–81. We first consider the short-range (one-Coulomb-line-irreducible) part of the singlet channel [see also Eq. (43)],

$$\Gamma^{1-2}_{ss'} = \Gamma^1_{ss'} - \Gamma^2_{ss}\delta_{ss'}\,, \tag{C1}$$

and include the long-range, one-Coulomb-line-reducible, diagrams (Γ^0) later on.

1. Resummation of interaction amplitudes

The first step is to single out the subset of particle-hole-section irreducible diagrams I^{1-2}. The total interaction amplitude as a matrix in the surface space and in $2+1$-momentum space is given by the Dyson equation

$$\underline{\Gamma}^{1-2}(K) = \underline{I}^{1-2} - \underline{I}^{1-2}\underline{R}(K)\underline{\Gamma}^{1-2}(K). \tag{C2}$$

(Matrix multiplication includes momentum integral $\int_p$ and a Matsubara sum $T\sum_n$.)

The matrix

$$[\underline{R}(K)]_{PP',ss'} = \delta_{ss'}\delta_{PP'}R_{s,P}(K), \tag{C3}$$

$$R_{s,P}(K) \equiv G_s(P)\,G_s(P+K) \tag{C4}$$

describes particle-hole bubbles and in the singlet channel. This matrix is diagonal in both $2+1$ momentum and surface space; as we explained in the main text, it is sufficient to keep only intrasurface bubbles in the assumed case of uncorrelated disorder. In the presence of generic interaction, the quantity $\underline{R}_{s,P}(K)$ can be represented as

$$R_{s,P}(K) = R^{\omega}_{s,P} + \Delta_{s,P}(K) \tag{C5}$$

$$= R^{q}_{s,P} + \tilde{\Delta}_{s,P}(K). \tag{C6}$$

Here, $R^{\omega}_{s,P}$ ($R^{q}_{s,P}$) are called regular (static) part of the bubble. The ω and q limits are defined in the main text [see Eqs. (48) and (49)]. The singular (dynamic) part of the particle-hole bubble is

$$\Delta_{s,P}(K) = \beta \frac{-i\boldsymbol{v}_s^F \cdot \boldsymbol{q}}{\omega_m + i\boldsymbol{v}_s^F \cdot \boldsymbol{q}} \delta_P^{(s)},$$

$$\tilde{\Delta}_{s,P}(K) = \beta \frac{\omega_m}{\omega_m + i\boldsymbol{v}_s^F \cdot \boldsymbol{q}} \delta_P^{(s)}.$$

[We have absorbed the Fermi liquid (FL) residues into a redefinition of the scattering amplitudes. In our notation, $\delta_P^{(s)}$ restricts the momentum integration and Matsubara summation over P to the Fermi surface (see Ref. 79).] From these definitions and Eq. (C2), we obtain the relations

$$\underline{\Gamma}^{1-2}(K) = \underline{\Gamma}^{1-2,\omega} - \underline{\Gamma}^{1-2}(K)\,\underline{\Delta}(K)\,\underline{\Gamma}^{1-2,\omega} \tag{C7a}$$

and

$$\underline{\Gamma}^{1-2}(K) = \underline{\Gamma}^{1-2,q} - \underline{\Gamma}^{1-2}(K)\,\underline{\tilde{\Delta}}(K)\,\underline{\Gamma}^{1-2,q}. \tag{C7b}$$

This formal (re-)expression of the general scattering amplitude will be used to calculate the polarization operator in the next section.

2. Definitions

In order to introduce the long-range Coulomb interaction and to describe its screening, we define the following quantities. The bare triangular vertices are obtained in response to an external scalar potential $\phi^{(s)}(\omega_m,\boldsymbol{q})$:

$$\underline{\boldsymbol{v}}_0^{(1)} = (1,0) \quad \text{and} \quad \underline{\boldsymbol{v}}_0^{(2)} = (0,1)\,. \tag{C8}$$

We used the approximation $\langle \mu_s, \boldsymbol{p}|\mu_s, \boldsymbol{p}+\boldsymbol{q}\rangle \approx 1$. In our notation, bold, italic, underlined quantities are vectors in surface space.

The triangular vertex $\underline{\boldsymbol{T}}^{(s)}$ renormalized by interaction satisfies

$$\underline{\boldsymbol{T}}^{(s)}(K) = \underline{\boldsymbol{v}}_0^{(s)} - \underline{\boldsymbol{v}}_0^{(s)}\underline{R}(K)\underline{\Gamma}^{1-2}(K)\,. \tag{C9}$$

The polarization operator is a matrix in the surface space and can be written as

$$\begin{aligned}\Pi^{ss'}(K) = {}& \underline{\boldsymbol{v}}_0^{(s)}\underline{R}(K)\left[\underline{\boldsymbol{v}}_0^{(s')}\right]^T \\ & - \underline{\boldsymbol{v}}_0^{(s)}\underline{R}(K)\underline{\Gamma}^{1-2}(K)\underline{R}(K)\left[\underline{\boldsymbol{v}}_0^{(s')}\right]^T,\end{aligned} \tag{C10}$$

which transforms into

$$\begin{aligned}\Pi^{ss'}(K) &= \Pi^{ss',q} + \underline{\boldsymbol{T}}^{(s),q}\tilde{\underline{\Delta}}(K)[\underline{\boldsymbol{T}}^{(s'),q}]^T \\ &\quad -\underline{\boldsymbol{T}}^{(s),q}\tilde{\underline{\Delta}}(K)\underline{\Gamma}^{1-2}(K)\tilde{\underline{\Delta}}(K)[\underline{\boldsymbol{T}}^{(s'),q}]^T \qquad \text{(C11a)}\\ &= \Pi^{ss',\omega} + \underline{\boldsymbol{T}}^{(s),\omega}\underline{\Delta}(K)[\underline{\boldsymbol{T}}^{(s'),\omega}]^T \\ &\quad -\underline{\boldsymbol{T}}^{(s),\omega}\underline{\Delta}(K)\underline{\Gamma}^{1-2}(K)\underline{\Delta}(K)[\underline{\boldsymbol{T}}^{(s'),\omega}]^T. \qquad \text{(C11b)}\end{aligned}$$

We will show below that these equations combined with Ward identities can be used to derive the ω and q limits of the polarization operator.

3. Ward identities

We will first investigate the Ward identities, which are due to invariance under separate **U**(1) rotation of the fermionic fields. Following the standard procedure, we obtain

$$\left(\frac{\partial G_1^{-1}}{\partial p_0},0\right) = \underline{\boldsymbol{T}}^{(1),\omega} \quad \text{and} \quad \left(0,\frac{\partial G_2^{-1}}{\partial p_0}\right) = \underline{\boldsymbol{T}}^{(2),\omega}. \tag{C12}$$

Next, we exploit that constant external fields can be reabsorbed into a redefinition of the chemical potentials. This leads to

$$\left(\frac{\partial G_1^{-1}}{\partial \mu_s},\frac{\partial G_2^{-1}}{\partial \mu_s}\right) = \underline{\boldsymbol{T}}^{(s),q}. \tag{C13}$$

We insert this into the ω and q limits of the polarization operator and obtain

$$\Pi^{ss',\omega} = 0 \quad \text{and} \quad \Pi^{ss',q} = -\frac{\partial N_s}{\partial \mu_{s'}} = -\frac{\partial N_{s'}}{\partial \mu_s}. \tag{C14}$$

The Ward identities (C13) and (C14) have very profound consequences. They relate the static triangular vertex and the static polarization operator to derivatives of physical observables with respect to the chemical potential. It is explained in the main text that for this reason they are not renormalized in the diffusive RG.[40]

4. Screening of the Coulomb interaction

We consider the singular part of the Coulomb interaction [see Eq. (B6)], i.e.,

$$\underline{U}_0 = \frac{2\pi}{\epsilon_{\text{eff}} q}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}, \tag{C15}$$

where $\epsilon_{\text{eff}} = (\epsilon_1+\epsilon_3)/2$ for the most general situation of a dielectric sandwich structure. This matrix has zero determinant, $\det \underline{U}_0 = 0$.

The RPA-screened Coulomb interaction is defined as

$$\underline{U}_{\text{scr}}(\omega_m,\boldsymbol{q}) = (1-\underline{U}_0\underline{\Pi})^{-1}\underline{U}_0.$$

The static one-Coulomb-line-reducible singlet interaction amplitude is obtained by attaching the (q limit) triangular vertices to $\underline{U}_{\text{scr}}(\omega_m=0,\boldsymbol{q})$ from both sides (see Fig. 10 in the main text). From the definition in Sec. C 2 we know that $\Pi^{ss',q} = T^{s,q}\nu_{s'}$. (Note that none of these three quantities is renormalized during RG.) Therefore we obtain

$$\underline{\Gamma}^0 = -\underline{\left(\frac{1}{\nu}\right)}\underline{\Pi}^q\underline{U}_{\text{scr}}(\omega_m=0,\boldsymbol{q})\underline{\Pi}^q\underline{\left(\frac{1}{\nu}\right)}. \tag{C16}$$

By means of the orthogonal matrix

$$O = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}, \tag{C17}$$

we can rotate $\underline{U}_{\text{scr}}(\omega_m=0,\boldsymbol{q})$ into the basis where $\underline{U}_0$ is diagonal:

$$\begin{aligned}&O^T\underline{U}_{\text{scr}}(\omega_m=0,\boldsymbol{q})O \\ &= \left[1-\begin{pmatrix}\frac{4\pi}{\epsilon_{\text{eff}}q} & 0\\ 0 & 0\end{pmatrix}O^T\underline{\Pi}^q O\right]^{-1}\begin{pmatrix}\frac{4\pi}{\epsilon_{\text{eff}}q} & 0\\ 0 & 0\end{pmatrix} \\ &= \frac{\frac{4\pi}{\epsilon_{\text{eff}}}}{q-\frac{2\pi}{\epsilon_{\text{eff}}}(\Pi_{11}^q+\Pi_{22}^q+2\Pi_{12}^q)}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}.\end{aligned} \tag{C18}$$

The denominator in the last line of Eq. (C18) defines the coupled surface screening length [analogously to Eqs. (7) and (8)].

In the considered parameter range, we can take the q-limit under the following condition: $|\frac{2\pi}{\epsilon_{\text{eff}}}(\Pi_{11}^q+\Pi_{22}^q+2\Pi_{12}^q)| \gg q$. Then we obtain

$$\underline{U}_{\text{scr}}^q = -O\begin{pmatrix}[\hat{e}_1^T O^T\underline{\Pi}^q O\hat{e}_1]^{-1} & 0\\ 0 & 0\end{pmatrix}O^T. \tag{C19}$$

The q limit of Eq. (C16) is

$$\underline{\Gamma}^{0,q} = \underline{\left(\frac{1}{\nu}\right)}\underline{\Pi}^q O\hat{e}_1\otimes\hat{e}_1^T O^T\underline{\Pi}^q\underline{\left(\frac{1}{\nu}\right)}\frac{1}{\hat{e}_1^T O^T\underline{\Pi}^q O\hat{e}_1}. \tag{C20}$$

We multiply by $\underline{\nu}O\hat{e}_1$ from the right side and find

$$\left[-\underline{\left(\frac{1}{\nu}\right)}\underline{\Pi}^q + \underline{\Gamma}^{0,q}\underline{\nu}\right]O\hat{e}_1 = 0. \tag{C21}$$

This matrix equation implies that the surface-space matrix in brackets has to be of zero determinant.

Alternatively, using the q-limit of Eqs. (C11b) and (C16), we can express the bare total interaction amplitude $\underline{\Gamma}^{\rho} \equiv \underline{\Gamma}^{0} + \underline{\Gamma}^{1-2}$ as

$$\underline{\nu\Gamma^{\rho}\nu} = -\underline{\nu} - \frac{\det \underline{\Pi}^{q}}{\Pi_{11}^{q} + \Pi_{22}^{q} + 2\Pi_{12}^{q}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \quad \text{(C22)}$$

From Eq. (C22), the following statement immediately follows:

$$\det[\underline{\nu} + \underline{\nu\Gamma^{\rho}\nu}] = 0. \quad \text{(C23)}$$

This relationship is equivalent to Eq. (C21).

5. Total density-density response

Here we analyze the one-Coulomb-line-reducible (which we will also term "total") density-density response $\underline{\Pi}^{\mathrm{RPA}}$. It is defined as

$$\underline{\Pi}^{\mathrm{RPA}}(K) = \underline{\Pi}(K) + \underline{\Pi}(K)\,\underline{U}_0(\boldsymbol{q})\,\underline{\Pi}^{\mathrm{RPA}}(K). \quad \text{(C24)}$$

Equation (C24) implies that Π^{RPA} is obtained by a resummation of the RPA-type series, hence the corresponding superscript.

For the present case, we want to obtain $\underline{\Pi}^{\mathrm{RPA}}$ in the diffusive regime. The very idea of dirty FL is based on replacing dynamic section $\tilde{\Delta}_{s,P}$ according to the following prescription:

$$\frac{\omega_m}{\omega_m + i\boldsymbol{v}_s^F \cdot \boldsymbol{q}} \quad \to \quad \frac{\omega_m}{Z_s\omega_m + D_s\boldsymbol{q}^2}, \quad \text{(C25)}$$

with $Z_s = 1$ at the bare level. By using definitions (C11a) and (C16), the total density-density response can be written as

$$\underline{\Pi}^{\mathrm{RPA}}(K) = [\underline{\Pi}^{q} - \underline{\nu\Gamma^{0}\nu}][1 + \omega_m\underline{\Delta}^{\Gamma}[\underline{\Pi}^{q} - \underline{\nu\Gamma^{0}\nu}]], \quad \text{(C26)}$$

where

$$\Delta^{\Gamma} \equiv \Delta^{\Gamma}(\omega_m, \boldsymbol{q}) = [\underline{\nu D}\boldsymbol{q}^2 + (\underline{\nu Z} + \underline{\nu\Gamma^{0+1-2}\nu})\omega_m]^{-1}. \quad \text{(C27)}$$

These equations are used in the main text (see Sec. IV B) to provide a link between the bosonized NLσM and the dirty FL theory.

6. Bare NLσM coupling constants

According to Eqs. (C18) and (C22), the bare values of the interaction amplitudes are fully determined by ν_1, ν_2 and

$$\underline{\Pi}^{q} = -\underline{\nu}\left(1 + \underline{F\nu}\right)^{-1}, \quad \text{(C28)}$$

where

$$\underline{F} = \begin{pmatrix} F_{11} & F_{12} \\ F_{12} & F_{22} \end{pmatrix} \quad \text{(C29)}$$

are the FL constants in the density channel(s). It is convenient to express $\underline{\nu\Gamma^{\rho}\nu}$ in Eq. (C22) through $\underline{F}$ by means of the identity (C28):

$$\frac{\det \underline{\Pi}^{q}}{\Pi_{11}^{q} + \Pi_{22}^{q} + 2\Pi_{12}^{q}} = \frac{-1}{1/\nu_1 + 1/\nu_2 + F_{11} + F_{22} - 2F_{12}}. \quad \text{(C30)}$$

In Appendix B, we derived the general expression for FL constants $\underline{F}^{(d)} \equiv \frac{2\pi}{\epsilon_2} d\,\underline{f}$ induced by the finite thickness of the topological insulator film. Assuming that there is no additional short-range interaction, one can deduce the bare value of interaction constants for the NLσM. This is equivalent to the RPA estimate (valid if $\alpha \ll 1$).

In the following, we consider two limits. As in the main text, the inverse single surface screening length is denoted by $\kappa_s = 2\pi\nu_s/\epsilon_2$. The first limit is the case of equal surfaces $\nu_1 = \nu_2$ in a symmetric setup $\epsilon_1 = \epsilon_2 = \epsilon_3$. Then the effective FL amplitude is

$$\underline{F}^{(d)} = -\frac{2\pi d}{\epsilon_2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad \text{(C31)}$$

The bare value of the interaction constant is

$$\gamma_{11} = \gamma_{22} = \frac{[\underline{\nu\Gamma^{\rho}\nu}]_{11}}{\nu_1} = -\frac{1}{2}\left(1 + \frac{\kappa d}{1 + \kappa d}\right). \quad \text{(C32)}$$

Note that in the limit $\kappa d \to \infty$ ($\kappa d \to 0$), the bare value of $\gamma_{11} = \gamma_{22}$ is equal to -1 ($-1/2$).

The second limit is the experimentally relevant situation with $\epsilon_1 \ll \epsilon_2, \epsilon_3$. In this limit, we find

$$\underline{F}^{(d)} = \frac{4\pi d}{\epsilon_2} \begin{pmatrix} 1 - (\epsilon_2/\epsilon_3)^2 & -(\epsilon_2/\epsilon_3)^2 \\ -(\epsilon_2/\epsilon_3)^2 & -(\epsilon_2/\epsilon_3)^2 \end{pmatrix}. \quad \text{(C33)}$$

It follows from Eqs. (C30) and (C33) that the bare values for interaction constants are ϵ_3 independent:

$$\gamma_{11} = -1 + \frac{1}{1 + \frac{\kappa_1}{\kappa_2} + 2\kappa_1 d} \quad \text{(C34a)}$$

and

$$\gamma_{22} = -1 + \frac{1}{1 + \frac{\kappa_2}{\kappa_1} + 2\kappa_2 d}. \quad \text{(C34b)}$$

In view of Eq. (65) following from the $\mathcal{F}$ invariance, it is not surprising that the coupling constants are equal as long as $\nu_1 = \nu_2$ even in the case of asymmetric dielectric environment.

APPENDIX D: DETAILED DERIVATION OF RG EQUATIONS

In this section, we present the detailed derivation of the one-loop corrections to conductivity.

a. Correlator B_1

In the one-loop approximation, we can use $Q = \Lambda + \delta Q$ with $\delta Q = \left(\begin{smallmatrix} 0 & q \\ q^T & 0 \end{smallmatrix}\right)$. Then we directly single out the classical contribution in B_1, Eq. (70) and obtain

$$B_1^s = \sigma_s - \frac{\sigma_s}{4n}\sum_{\mu=0,2} \mathrm{tr}\left\langle I_n^{\alpha}\tilde{\tau}_{\mu}\delta Q\left[I_{-n}^{\alpha}\tilde{\tau}_{\mu}^{T}\delta Q - I_n^{\alpha}\tilde{\tau}_{\mu}\delta Q\right]\right\rangle. \quad \text{(D1)}$$

In addition, we write $q = \sum_{\nu=0}^{3} q^{(\nu)}\tilde{\tau}_{\nu}^{T}$. When performing the trace in τ space it turns out that the two diffuson contributions ($\nu = 0,2$) cancel out. This is a consequence of the opposite sign of τ_0 and τ_2 under transposition. The cooperons ($\nu = 1,3$) contribute only to the last term in Eq. (D1). Then we find

$$B_1^s = \sigma_s + \frac{\sigma_s}{4n}\sum_{\nu=1,3}\left\langle \mathrm{tr} I_n^{\alpha}\begin{pmatrix} 0 & q_{\nu} \\ q_{\nu}^T & 0 \end{pmatrix} I_n^{\alpha}\begin{pmatrix} 0 & q_{\nu} \\ q_{\nu}^T & 0 \end{pmatrix}\right\rangle$$
$$= \sigma_s + 2\int_{\boldsymbol{p}} D_s(\omega_n, \boldsymbol{p}). \quad \text{(D2)}$$

b. Correlator B_2

The second term B_2, Eq. (71), does not contribute on the classical level. Expanding to second order in q we obtain the tree level contribution, which also vanishes:

$$B_2^{ss'}\Big|_{\text{tree level}} = -\frac{\sigma_s\delta_{ss'}}{4}\int_{\boldsymbol{x}-\boldsymbol{x}'} e^{i\boldsymbol{p}(\boldsymbol{x}-\boldsymbol{x}')}\boldsymbol{p}^2 D^c_{ss}(\boldsymbol{p},\omega) = 0. \quad \text{(D3)}$$

The quartic order in q provides the one-loop corrections to the correlator B_2. We will first analyze the effect of diffusons. Exploiting the relation $\langle q^{(0)}q^{(0)}\rangle = \langle q^{(2)}q^{(2)}\rangle$, we can simplify the expression for B_2 ($\bigstar$ and $\blacklozenge$ denote Wick contractions):

$$\begin{aligned} B_2^{(ss')} = &-\frac{\sigma_s\sigma_{s'}}{8n}\int_{\boldsymbol{x}-\boldsymbol{x}'}\sum_{\mu=x,y} \\ &\times\big[\mathrm{tr}\big(I^\alpha_n q_0^{\blacklozenge}\partial_\mu q_0^{T\bigstar}\big)_{s,\boldsymbol{x}}\mathrm{tr}\big(I^\alpha_n q_0^{\blacklozenge}\partial'_\mu q_0^{T\bigstar}\big)_{s',\boldsymbol{x}'} \\ &+\mathrm{tr}\big(I^\alpha_n q_0^{T\blacklozenge}\partial_\mu q_0^{\bigstar}\big)_{s,\boldsymbol{x}}\mathrm{tr}\big(I^\alpha_n q_0^{T\blacklozenge}\partial'_\mu q_0^{\bigstar}\big)_{s',\boldsymbol{x}'} \\ &+2\mathrm{tr}\big(I^\alpha_n q_0^{T\bigstar}\partial_\mu q_0^{\blacklozenge}\big)_{s,\boldsymbol{x}}\mathrm{tr}\big(I^\alpha_n q_0^{\blacklozenge}\partial'_\mu q_0^{T\bigstar}\big)_{s',\boldsymbol{x}'}\big]. \end{aligned} \quad \text{(D4)}$$

The Wick contraction produces three types of terms for each of the three terms in Eq. (D4) [see Eq. (79)]. First, there is the interference term D_sD_s. It contains an additional sum over replicas and hence vanishes in the replica limit. Second, there can be a term $(\underline{D\Gamma D^c})_{ss'}(\underline{D\Gamma D^c})_{ss'}$. It vanishes due to its structure in the Matsubara space. The only remaining term is $(\underline{D\Gamma D^c})_{ss}D_s$, which yields

$$\begin{aligned} B_2^{(ss')} = &\frac{32\pi T\delta_{ss'}}{\sigma_s n}\int_{\boldsymbol{p}}\boldsymbol{p}^2\sum_{n_{12}=0}^{N_M} n_{12} \\ &\times[(\underline{D\Gamma D^c})_{ss}(\omega_{n_{12}},\boldsymbol{p})D_s(\omega_{n_{12}+n},\boldsymbol{p}) \\ &-(\underline{D\Gamma D^c})_{ss}(\omega_{n_{12}+n},\boldsymbol{p})D_s(\omega_{n_{12}+2n},\boldsymbol{p})]. \end{aligned} \quad \text{(D5)}$$

At this stage, we can send $N_M\to\infty$. Furthermore, note that because disorder is surface uncorrelated, there is no correction to the transconductance σ_{12}. Since we are interested in the zero temperature limit, Eq. (D5) becomes

$$B_2^{(ss')} = \frac{16\delta_{ss'}}{\sigma_s}\int_{\boldsymbol{p}}\boldsymbol{p}^2\int_0^\infty d\omega\,\left(\underline{D\Gamma D^c}\right)_{ss}(\omega,\boldsymbol{p})\,D_s(\omega,\boldsymbol{p}). \quad \text{(D6)}$$

We use the relation

$$\begin{aligned} &\left(\underline{D\Gamma D^c}\right)_{ss}(\omega,\boldsymbol{p})\,D_s(\omega,\boldsymbol{p}) \\ &\quad = \Gamma_{ss}D_s^2(\omega,\boldsymbol{p})\,\overline{D}_s(\omega,\boldsymbol{p}) - \frac{4\omega\Gamma_{12}^2 D_s(\omega,\boldsymbol{p})\,\overline{D}_s(\omega,\boldsymbol{p})}{\sigma_{(-s)}\det\{[\underline{D}^c(\omega,\boldsymbol{p})]^{-1}\}} \end{aligned} \quad \text{(D7)}$$

in order to split Eq. (D6) into the single surface and intersurface contributions. Here, $\sigma_{(-1)}=\sigma_2$, $\sigma_{(-2)}=\sigma_1$ and

$$\overline{D}_s(\omega,\boldsymbol{p}) = [\boldsymbol{p}^2 + L^{-2} + 4(z_s+\Gamma_{ss})\omega/\sigma_s]^{-1}. \quad \text{(D8)}$$

The single surface induced correction is given as

$$\begin{aligned} B_2^{ss'}\Big|_{\text{single}} &= \frac{16\delta_{ss'}}{\sigma_s}\int_{\boldsymbol{p}}\boldsymbol{p}^2\int_0^\infty d\omega\,\Gamma_{ss}D_s^2(\omega,\boldsymbol{p})\,\overline{D}_s(\omega,\boldsymbol{p}) \\ &= -4\delta_{ss'}f(\Gamma_{ss}/z_s)\int_{\boldsymbol{p}}\boldsymbol{p}^2 D_s^2(0,\boldsymbol{p}). \end{aligned} \quad \text{(D9)}$$

Here, we introduced the function

$$f(\boldsymbol{x}) = 1-(1+1/x)\ln(1+x). \quad \text{(D10)}$$

For the intersurface interaction induced term, we separate the poles of $\{\det[\underline{D}^c(\omega,\boldsymbol{p})]^{-1}\}$. It yields

$$\begin{aligned} B_2^{ss'}|_{\text{inter}} = &\frac{64\delta_{ss'}\Gamma_{12}^2}{\det(\underline{z}+\underline{\Gamma})}\int_{\boldsymbol{p}}\boldsymbol{p}^2 D_s(0,\boldsymbol{p})\int_0^\infty d\omega\,\omega\,D_s(\omega,\boldsymbol{p}) \\ &\times\frac{\overline{D}_s(\omega,\boldsymbol{p})}{d_+-d_-}\sum_{\varsigma=\pm}\frac{\varsigma}{d_\varsigma(\boldsymbol{p}^2+L^{-2})+4\omega} \\ = &\frac{2\sigma_s^2\Gamma_{12}^2\delta_{ss'}}{z_s(z_s+\Gamma_{ss})\det(\underline{z}+\underline{\Gamma})(d_+-d_-)} \\ &\times\left[\sum_{\varsigma=\pm}\varsigma f_2\left(\frac{\sigma_s}{z_s},\frac{\sigma_s}{z_s+\Gamma_{ss}},d_\varsigma\right)\right]\int_{\boldsymbol{p}}\boldsymbol{p}^2 D_s^2(0,\boldsymbol{p}), \end{aligned} \quad \text{(D11)}$$

where

$$d_\pm = \frac{(z_1\sigma_2+\sigma_1 z_2)}{2\det(\underline{z}+\underline{\Gamma})}\left[1\mp\sqrt{1-\frac{4\sigma_1\sigma_2\det(\underline{z}+\underline{\Gamma})}{(z_1\sigma_2+\sigma_1 z_2)^2}}\right] \quad \text{(D12)}$$

and

$$f_2(a,b,c) = 2\frac{(c-b)a\ln a+(a-c)b\ln b+(b-a)c\ln c}{(b-a)(c-a)(c-b)}. \quad \text{(D13)}$$

In the case of the long-range Coulomb interaction, the condition $\det(\underline{z}+\underline{\Gamma})=0$ holds. Therefore, d_- diverges and as a consequence $f_2(\frac{\sigma_s}{z_s},\frac{\sigma_s}{z_s+\Gamma_{ss}},d_-)\to 0$. The contribution due to d_+ is then, in the exemplary case $s=1$, given as

$$\begin{aligned} B_2^{11}|_{\text{inter}} = -4\left(1+\frac{\Gamma_{11}}{z_1}\right)\Bigg\{&\frac{\ln\left(1+\frac{\Gamma_{11}}{z_1}\right)}{\frac{\Gamma_{11}}{z_1}} \\ &-\frac{\ln\left[1+\frac{\Gamma_{11}}{z_1}+\frac{\sigma_1(z_2+\Gamma_{22})}{\sigma_2 z_1}\right]}{\frac{\Gamma_{11}}{z_1}+\frac{\sigma_1(z_2+\Gamma_{22})}{\sigma_2 z_1}}\Bigg\}\int_{\boldsymbol{p}}\boldsymbol{p}^2 D_s^2(0,\boldsymbol{p}). \end{aligned} \quad \text{(D14)}$$

Finally, we consider the effect of cooperons in B_2. Due to the absence of interaction amplitudes in the cooper channel, all contributions are of the type D_sD_s and, in analogy with the corresponding diffuson terms, vanish in the replica limit.

APPENDIX E: STABILITY OF THE FIXED PLANE OF EQUAL SURFACES

We discuss here the stability of the fixed plane of identical surfaces with respect to small perturbations. As anticipated, it hosts the overall attractive fixed point of the four-dimensional RG flow (see also Sec. V B2) and thus is itself attractive. However, the parameters describing the deviation from equal surfaces ($\delta t = t-1$ and $\delta\gamma=\gamma_{11}-\gamma_{22}$) flow towards zero in a quite nontrivial manner.

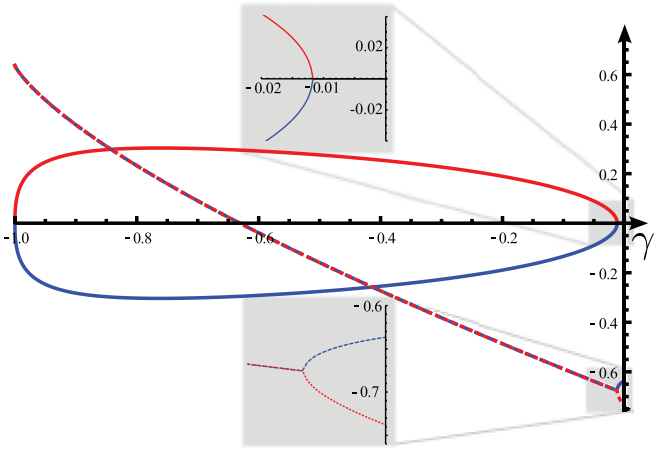

FIG. 20. (Color online) Eigenvalues of $\mathbf{M}(\gamma)$ (in units of $1/\sigma$) as a function of γ. Dashed lines: real part, solid lines: imaginary part.

From the general RG Equations (96), we obtain the equations for small deviations:

$$\frac{d}{dy}\begin{pmatrix}\delta t\\ \delta\gamma\end{pmatrix} = \mathbf{M}(\gamma)\begin{pmatrix}\delta t\\ \delta\gamma\end{pmatrix}, \tag{E1}$$

with the γ-dependent matrix

$$\mathbf{M}(\gamma) = -\frac{2}{\pi\sigma}\begin{pmatrix}\frac{(3+4\gamma)G(\gamma)}{(1+2\gamma)^2} & \frac{2G(\gamma)}{(1+2\gamma)^2}\\ -(\gamma+\gamma^2) & 1+2\gamma\end{pmatrix}, \tag{E2}$$

and $G(\gamma) = -1-2\gamma+(2+2\gamma)\ln(2+2\gamma)$. The eigenvalues of the matrix $\mathbf{M}(\gamma)$ are shown in Fig. 20. They turn out to be complex in most of the interval $\gamma \in [-1,0]$ (except for a narrow region of very small γ). This implies a curious oscillatory scale dependence of the difference of conductivities $\delta t = 2(\sigma_1-\sigma_2)/\sigma$. Although the fixed plane of equal surfaces is repulsive in the regime $\gamma < \gamma_* \approx -0.64$, one should keep in mind that γ itself is subjected to renormalization, flowing towards zero and therefore, the plane of identical surfaces becomes ultimately attractive.

APPENDIX F: RG FLOW FOR EXTERNALLY SCREENED INTERACTION

If the single-layer screening length κ_s^{-1} and the typical length scale L_E (e.g., the thermal length) exceed the distance to the electrostatic gates, the external screening of interactions can no longer be neglected. Effectively, the interactions become short ranged. This implies the breakdown of $\mathcal{F}$ invariance. As a consequence, the relations for NLσM parameters $\det(\underline{z}+\underline{\Gamma}) = 0$ and $(z_1+\Gamma_{11})/(z_2+\Gamma_{22}) = 1$ (derived in the case of long-range interaction in Sec. III G 2 and Appendix C 4) are no longer true. Note that the invariance under renormalization of $(\underline{z}+\underline{\Gamma})$ is not a consequence of $\mathcal{F}$ invariance and still holds.

Here, we present general RG equations that allow us to describe the crossover between the cases of long-range Coulomb interaction and of no interaction:

$$\frac{d\sigma_1}{dy} = \frac{2}{\pi}\left[\frac{1}{2} - f\left(\frac{\Gamma_{11}}{z_1}\right) - \frac{\sigma_1^2\Gamma_{12}^2\sum_{\varsigma=\pm}\varsigma f_2\left(\frac{\sigma_1}{z_1},\frac{\sigma_1}{z_1+\Gamma_{11}},d_\varsigma\right)}{2z_1(z_1+\Gamma_{11})\det(\underline{z}+\underline{\Gamma})(d_- - d_+)}\right], \tag{F1a}$$

$$\frac{d\sigma_2}{dy} = \frac{2}{\pi}\left[\frac{1}{2} - f\left(\frac{\Gamma_{22}}{z_2}\right), - \frac{\sigma_2^2\Gamma_{12}^2\sum_{\varsigma=\pm}\varsigma f_2\left(\frac{\sigma_2}{z_2},\frac{\sigma_2}{z_2+\Gamma_{22}},d_\varsigma\right)}{2z_2(z_2+\Gamma_{22})\det(\underline{z}+\underline{\Gamma})(d_- - d_+)}\right], \tag{F1b}$$

$$\frac{dz_1}{dy} = -\frac{d\Gamma_{11}}{dy} = \frac{\Gamma_{11}}{\pi\sigma_1}, \tag{F1c}$$

$$\frac{dz_2}{dy} = -\frac{d\Gamma_{22}}{dy} = \frac{\Gamma_{22}}{\pi\sigma_2}. \tag{F1d}$$

In contrast to the Coulomb case [see Eq. (94)], these RG equations can not be expressed in terms of the parameter $\gamma_{ss} = \Gamma_{ss}/z_s$. Further, we emphasize that the RG equations for Γ_{ss} and z_s are exactly the same as in the Coulomb case. In particular, Γ_{12} is not renormalized, since the general arguments exposed in Sec. IV B hold also in the case of short ranged interactions. It is worthwhile to repeat that $0 \leqslant |\Gamma_{ss}| \leqslant z_s$ and typically $|\Gamma_{12}| \leqslant \max_{s=1,2}|\Gamma_{ss}|$.

For sufficiently strong interactions, the RG flow implies localizing behavior of the conductivities. However, as the RG flow predicts decreasing interaction amplitudes, the system undergoes a crossover to the free-electron weak-antilocalization effect. (Note that also Γ_{12}/z_s decreases.) Accordingly, similar to the case of Coulomb interaction, in the case of strong short-range interactions, we also predict a nonmonotonic conductivity behavior. The quantitative difference is the steeper antilocalizing slope in the final stage of the flow.

[1] B. A. Bernevig and S.-C. Zhang, Phys. Rev. Lett. **96**, 106802 (2006).
[2] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, Science **314**, 1757 (2006).
[3] L. Fu, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **98**, 106803 (2007).
[4] J. E. Moore and L. Balents, Phys. Rev. B **75**, 121306 (2007).
[5] R. Roy, Phys. Rev. B **79**, 195322 (2009).
[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] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, and M. Z. Cava, R. J. Hasan, Nature (London) **452**, 970 (2008).
[8] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).
[9] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).
[10] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, Phys. Rev. B **78**, 195125 (2008).
[11] A. Kitaev, AIP Conf. Proc. **1134**, 22 (2009).
[12] V. Gurarie, Phys. Rev. B **83**, 085426 (2011).
[13] K. v. Klitzing, G. Dorda, and M. Pepper, Phys. Rev. Lett. **45**, 494 (1980).

[14]D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, Phys. Rev. Lett. **49**, 405 (1982).
[15]D. Culcer, Physica E **44**, 860 (2012).
[16]H. Steinberg, D. R. Gardner, Y. S. Lee, and P. Jarillo-Herrero, Nano Lett. **10**, 5032 (2010).
[17]J. G. Checkelsky, Y. S. Hor, R. J. Cava, and N. P. Ong, Phys. Rev. Lett. **106**, 196801 (2011).
[18]D. Kim, S. Cho, N. P. Butch, P. Syers, K. Kirshenbaum, S. Adam, J. Paglione, and M. S. Fuhrer, Nat. Phys. **8**, 460 (2012).
[19]S. S. Hong, J. J. Cha, D. Kong, and Y. Cui, Nat. Comm. **3**, 757 (2012).
[20]D. Kim, Q. Li, P. Syers, N. P. Butch, J. Paglione, S. Das Sarma, and M. S. Fuhrer, Phys. Rev. Lett. **109**, 166801 (2012).
[21]P. Cheng, C. Song, T. Zhang, Y. Zhang, Y. Wang, J.-F. Jia, J. Wang, Y. Wang, B.-F. Zhu, X. Chen *et al.* Phys. Rev. Lett. **105**, 076801 (2010).
[22]T. Hanaguri, K. Igarashi, M. Kawamura, H. Takagi, and T. Sasagawa, Phys. Rev. B **82**, 081305 (2010).
[23]J. G. Analytis, R. D. McDonald, S. C. Riggs, J.-H. Chu, G. S. Boebinger, and I. R. Fisher, Nat. Phys. **6**, 960 (2010).
[24]B. Sacépé, J. B. Oostinga, J. Li, A. Ubaldini, N. J. G. Couto, E. Giannini, and A. F. Morpurgo, Nat. Comm. **2**, 575 (2011).
[25]C. Brüne, C. X. Liu, E. G. Novik, E. M. Hankiewicz, H. Buhmann, Y. L. Chen, X. L. Qi, Z. X. Shen, S. C. Zhang, and L. W. Molenkamp, Phys. Rev. Lett. **106**, 126803 (2011).
[26]H. Peng, K. Lai, D. Kong, S. Meister, Y. Chen, X.-L. Qi, S.-C. Zhang, Z.-X. Shen, and Yi Cui, Nat. Mater. **9**, 225 (2010).
[27]F. Xiu, L. He, Y. Wang, L. Cheng, L.-T. Chang, M. Lang, G. Huang, X. Kou, Y. Zhou, X. Jiang, Z. Chen, J. Zou, A. Shailos, and K. L. Wang, Nat. Nanotechnology **6**, 216 (2011).
[28]J. Dufouleur, L. Veyrat, A. Teichgraber, S. Neuhaus, C. Nowka, S. Hampel, J. Cayssol, J. Schumann, B. Eichler, O. G. Schmidt, B. Buchner, and R. Giraud, Phys. Rev. Lett. **110**, 186806 (2013).
[29]H. Steinberg, J. B. Laloë, V. Fatemi, J. S. Moodera, and P. Jarillo-Herrero, Phys. Rev. B **84**, 233101 (2011).
[30]G. Zhang, H. Qin, J. Chen, X. He, L. Lu, Y. Li, and K. Wu, Adv. Funct. Mater. **21**, 2351 (2011).
[31]J. Chen, X. Y. He, K. H. Wu, Z. Q. Ji, L. Lu, J. R. Shi, J. H. Smet, and Y. Q. Li, Phys. Rev. B **83**, 241304 (2011).
[32]J. Wang, A. M. DaSilva, C.-Z. Chang, K. He, J. K. Jain, N. Samarth, X.-C. Ma, Q.-K. Xue, and Moses H. W. Chan, Phys. Rev. B **83**, 245438 (2011).
[33]M. Liu, C.-Z. Chang, Z. Zhang, Y. Zhang, W. Ruan, K. He, L.-L. Wang, X. Chen, J.-F. Jia, S.-C. Zhang *et al.*, Phys. Rev. B **83**, 165440 (2011).
[34]B. L. Altshuler and A. G. Aronov, *in Electron-Electron Interactions in Disordered Conductors*, edited by A. J. Efros and M. Pollack (North-Holland, Amsterdam, 1985).
[35]P. M. Ostrovsky, I. V. Gornyi, and A. D. Mirlin, Phys. Rev. Lett. **105**, 036803 (2010).
[36]I. S. Burmistrov, I. V. Gornyi, and K. S. Tikhonov, Phys. Rev. B **84**, 075338 (2011).
[37]G. M. Minkov, A. V. Germanenko, O. E. Rut, A. A. Sherstobitov, A. K. Bakarov, and D. V. Dmitriev, Phys. Rev. B **84**, 075337 (2011).
[38]A. M. Finkel'stein, Zh. Eksp. Teor. Fiz. **84**, 168 (1983).
[39]A. M.Finkel'stein, Z. Phys. B **56**, 189 (1984).
[40]A. M. Finkelstein, in *Soviet Scientific Reviews: Physics Reviews*, edited by I. M. Khalatnikov (Harwood Academic Publishers, 1990), Vol. 14, p. 1.
[41]D. Belitz and T. R. Kirkpatrick, Rev. Mod. Phys. **66**, 261 (1994).
[42]A. M. Finkel'stein, in *50 years of Anderson Localization*, edited by E. Abrahams (World Scientific, 2010), p. 385.
[43]C. Castellani, C. Di Castro, P. A. Lee, and M. Ma, Phys. Rev. B **30**, 527 (1984).
[44]A. Finkel'stein, Pis'ma Zh. Eksp. Teor. Fiz. **45**, 37 (1987) [JETP Lett. **45**, 46 (1987)].
[45]A. Finkel'stein, Physica B **197**, 636 (1994).
[46]M. B. A. M. M. Pruisken, Europhys. Lett. **31**, 543 (1995).
[47]A. Pruisken and I. Burmistrov, Ann. Phys. **322**, 1265 (2007).
[48]A. Punnoose and A. M. Finkel'stein, Science **310**, 289 (2005).
[49]L. Zheng and A. H. MacDonald, Phys. Rev. B **48**, 8203 (1993).
[50]A. Kamenev and Y. Oreg, Phys. Rev. B **52**, 7516 (1995).
[51]K. Flensberg, B. Y.-K. Hu, A.-P. Jauho, and J. M. Kinaret, Phys. Rev. B **52**, 14761 (1995).
[52]J. Linder, T. Yokoyama, and A. Sudbø, Phys. Rev. B **80**, 205401 (2009).
[53]A. A. Taskin, S. Sasaki, K. Segawa, and Y. Ando, Phys. Rev. Lett. **109**, 066803 (2012).
[54]R. E. V. Profumo, M. Polini, R. Asgari, R. Fazio, and A. H. MacDonald, Phys. Rev. B **82**, 085443, (2010).
[55]M. I. Katsnelson, Phys. Rev. B **84**, 041407(R) (2011).
[56]M. Carrega, T. Tudorovskiy, A. Principi, M. I. Katsnelson, and M. Polini, New J. Phys. **14**, 063033 (2012).
[57]F. Zhang, C. L. Kane, and E. J. Mele, Phys. Rev. B **86**, 081303 (2012).
[58]J. González, F. Guinea, and M. A. H. Vozmediano, Phys. Rev. B **63**, 134421 (2001).
[59]M. S. Foster and I. L. Aleiner, Phys. Rev. B **77**, 195413 (2008).
[60]D. E. Sheehy and J. Schmalian, Phys. Rev. Lett. **99**, 226803 (2007).
[61]I. L. Aleiner and K. B. Efetov, Phys. Rev. Lett. **97**, 236801 (2006).
[62]P. M. Ostrovsky, I. V. Gornyi, and A. D. Mirlin, Phys. Rev. B **74**, 235443 (2006).
[63]A. Schuessler, P. M. Ostrovsky, I. V. Gornyi, and A. D. Mirlin, Phys. Rev. B **79**, 075405 (2009).
[64]P. Adroguer, D. Carpentier, J. Cayssol, and E. Orignac, New J. Phys. **14**, 103027 (2012).
[65]M. Sitte, A. Rosch, and L. Fritz, arXiv:1305.1788.
[66]N. M. R. Peres, F. Guinea, and A. H. Castro Neto, Phys. Rev. B **72**, 174406 (2005).
[67]B. Seradjeh, J. E. Moore, and M. Franz, Phys. Rev. Lett. **103**, 066402 (2009).
[68]A. M. M. Pruisken, M. A. Baranov, and B. Skoric, Phys. Rev. B **60**, 16807 (1999).
[69]K. B. Efetov, A. I. Larkin, and D. E. Khmelnitskii, Sov. Phys. JETP **52**, 568 (1980).
[70]E. Witten, Commun. Math. Phys. **92**, 455 (1984).
[71]A. A. Nersesyan, A. M. Tsvelik, and F. Wenger, Phys. Rev. Lett. **72**, 2628 (1994).
[72]A. Nersesyan, A. Tsvelik, and F. Wenger, Nucl. Phys. B **438**, 561 (1995).
[73]A. Altland, B. Simons, and M. Zirnbauer, Phys. Rep. **359**, 283 (2002).
[74]A. Altland, Phys. Rev. Lett. **97**, 236802 (2006).
[75]E. J. König, P. M. Ostrovsky, I. V. Protopopov, and A. D. Mirlin, Phys. Rev. B **85**, 195130 (2012).
[76]S. Ryu, C. Mudry, H. Obuse, and A. Furusaki, Phys. Rev. Lett. **99**, 116601 (2007).
[77]P. M. Ostrovsky, I. V. Gornyi, and A. D. Mirlin, Phys. Rev. Lett. **98**, 256801 (2007).

[78]L. Fu and C. L. Kane, Phys. Rev. Lett. **109**, 246605 (2012).
[79]P. Nozieres and J. Luttinger, Phys. Rev. **127**, 1423 (1962).
[80]A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinskij, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, Englewood Cliffs, NJ, 1963).
[81]E. M. Lifshitz and L. P. Pitaevskij, *Statistical Physics, Part 2: Theory of Condensed State*, Course of Theoretical Physics Vol. 9 (Pergamon Press, Oxford, 1980).
[82]R. Nandkishore, J. Maciejko, D. A. Huse, and S. L. Sondhi, Phys. Rev. B **87**, 174511 (2013).
[83]I. S. Burmistrov, I. V. Gornyi, and A. D. Mirlin, Phys. Rev. Lett. **108**, 017002 (2012).
[84]A. Polyakov and P. Wiegmann, Phys. Lett. B **131**, 121 (1983).
[85]P. D. Vecchia, B. Durhuus, and J. Petersen, Phys. Lett. B **144**, 245 (1984).
[86]L. Faddeev, Lett. Math. Phys. **1**, 289 (1976).
[87]A. Gerasimov, arXiv:hep-th/9305090.
[88]A. Losev, G. Moore, N. Nekrasov, and S. Shatashvili, arXiv:hep-th/9511185v2.
[89]A. V. Smilga, Phys. Rev. D **54**, 7757 (1996).
[90]K. Nomura, S. Ryu, M. Koshino, C. Mudry, and A. Furusaki, Phys. Rev. Lett. **100**, 246806 (2008).
[91]X.-L. Qi, T. L. Hughes, and S.-C. Zhang, Phys. Rev. B **78**, 195424 (2008).
[92]X.-L. Qi, R. Li, J. Zang, and S.-C. Zhang, Science **323**, 1184 (2009).
[93]A. M. Essin, J. E. Moore, and D. Vanderbilt, Phys. Rev. Lett. **102**, 146805 (2009).
[94]D. A. Pesin and A. H. MacDonald, arXiv:1207.4444v1.
[95]P. M. Ostrovsky, I. V. Gornyi, and A. D. Mirlin, Phys. Rev. B **77**, 195430 (2008).
[96]A. M. M. Pruisken, in *The Quantum Hall Effect*, edited by R. E. Prange and S. M. Girvin (Springer-Verlag, Berlin, 1990).
[97]A. Pruisken and I. Burmistrov, Ann. Phys. **316**, 285 (2005).
[98]M. Bocquet, D. Serban, and M. Zirnbauer, Nucl. Phys. B **578**, 628 (2000).
[99]M. A. Baranov, A. M. M. Pruisken, and B. Škorić, Phys. Rev. B **60**, 16821 (1999).
[100]S. Hikami, A. I. Larkin, and Y. Nagaoka, Prog. Theor. Phys. **63**, 707 (1980).
[101]C. Castellani and C. Di Castro, Phys. Rev. B **34**, 5935 (1986).
[102]T. Ohtsuki, K. Slevin, and B. Kramer, Physica E **22**, 248 (2004).
[103]P. Markos and L. Schweitzer, J. Phys. A: Math. Gen. **39**, 3221 (2006).
[104]F. Evers and A. D. Mirlin, Rev. Mod. Phys. **80**, 1355 (2008).
[105]J. H. Bardarson, J. Tworzydło, P. W. Brouwer, and C. W. J. Beenakker, Phys. Rev. Lett. **99**, 106801 (2007).
[106]K. Nomura, M. Koshino, and S. Ryu, Phys. Rev. Lett. **99**, 146806 (2007).
[107]P. M. Ostrovsky, I. V. Gornyi, and A. D. Mirlin, Phys. Rev. B **86**, 125323 (2012).
[108]I. Garate and L. Glazman, Phys. Rev. B **86**, 035422 (2012).
[109]V. Krueckl and K. Richter, Semicond. Sci. Technol. **27**, 124006 (2012).
[110]Y. Li (private communication).
[111]J. Baars and F. Sorger, Solid State Commun. **10**, 875 (1972).
[112]J. N. Hancock, J. L. M. van Mechelen, A. B. Kuzmenko, D. van der Marel, C. Brüne, E. G. Novik, G. V. Astakhov, H. Buhmann, and L. W. Molenkamp, Phys. Rev. Lett. **107**, 136803 (2011).
[113]Z. Ringel, Y. E. Kraus, and A. Stern, Phys. Rev. B **86**, 045102 (2012).
[114]K. Kobayashi, T. Ohtsuki, and K.-I. Imura, Phys. Rev. Lett. **110**, 236803 (2013).
[115]P. M. Solomon, P. J. Price, D. J. Frank, and D. C. La Tulipe, Phys. Rev. Lett. **63**, 2508 (1989).
[116]T. J. Gramila, J. P. Eisenstein, A. H. MacDonald, L. N. Pfeiffer, and K. W. West, Phys. Rev. Lett. **66**, 1216 (1991).
[117]M. Schütt, P. M. Ostrovsky, M. Titov, I. V. Gornyi, B. N. Narozhny, and A. D. Mirlin, Phys. Rev. Lett. **110**, 026601 (2013).
[118]L. A. Wray, S. Xu, Y. Xia, D. Qian, A. V. Fedorov, H. Lin, A. Bansil, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nat. Phys. **6**, 855 (2010).
[119]Related problem on interaction effects on the surface of 3D topological superconductor of class CI was studied in M. S. Foster, E. A. Yuzbashyan, Phys. Rev. Lett. **109**, 246801 (2012).
[120]W. Zhang, R. Yu, H.-J. Zhang, X. Dai, and Z. Fang, New J. Phys. **12**, 065013 (2010).
[121]Closer than the typical length scale L_E of the system, see Eq. (3).
[122]The superconducting instability in a disordered system of Dirac fermions on a 3D TI surface (in the absence of density-density interactions) was recently addressed in Ref. 82.
[123]Instead of considering the renormalization of z_s, one can equivalently consider the renormalization of Γ_{ss}. It is governed by the interaction term $S^{\rm int}$ in Eq. (57). Within the background field method two types of contributions can arise. First, there is $\langle S^{\rm int}\rangle_{\rm fast}$. This term does not involve a frequency integration. Because disorder is uncorrelated between the surfaces, Γ_{11} and Γ_{22} are renormalized separately. This is described by Eq. (93). All possible further contributions at this order would arise from $\langle (S^{\rm int})^2\rangle_{\rm fast}$. This term generates so-called ring diagrams.[40] We have explicitly checked that the ring diagrams vanish in one-loop approximation.
[124]The $\mathbb{Z}_2$ theta term is also absent for the critical state separating 2D trivial and topological insulator. Such a state can be realized, in particular, on a surface of a weak 3D topological insulator. Despite the absence of theta term it is protected from Anderson localization due to topological reasons[78,107,113] (see Sec. III E2) and hence does not fall into our definition of non-topological symplectic metals.
[125]It is worthwhile to repeat that in the diffusive regime of typical experiments on thin films, the 3D TI bulk electrons are subjected to 2D diffusive motion.
[126]*A priori* σ_1^* and σ_2^* are different, although we cannot exclude a possibility that they might be equal.
[127]The only assumption that can not be directly verified on the basis of the quoted experimental data is the absence of complete intersurface correlations of disorder. We remind the reader that in the case of $\mu_1 = \mu_2$, a completely correlated disorder implies extra soft modes.[36] We see, however, no reason for such perfect correlations between impurities at opposite surfaces of a 3D TI film.