Interaction and disorder effects in 3D topological insulator thin films

E. J. König, P. M. Ostrovsky, I. V. Protopopov, I. V. Gornyi, I. S. Burmistrov, and A. D. Mirlin

1 Inst. 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.

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 sigma 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 non-monotonic temperature dependence of the conductivity. In the infrared (low-temperature) limit, the systems 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 $Z_2$ 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 non-topological counterpart in a double-well structure with strong spin-orbit interaction.

PACS numbers:

1. INTRODUCTION

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

In their bulk, topological insulators (TI) are electronic band insulators characterized by a topological invariant which accounts for the non-trivial 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. In retrospect we understand that the quantum Hall effect (QHE) 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 TKNN 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 $Z_2$ (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. The discriminating feature of all 3D TI is the massless Dirac states on the 2D boundary which were first spectroscopically detected in BiSb alloys and subsequently in many other materials. To present date, several experimental groups confirmed predominant surface state transport (for a review see Ref. 15), in particular elucidating the typical QHE-steps of Dirac electrons and analyzing weak antilocalization (WAL) corrections in the magnetoconductivity data.

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 induced by inter- and intrasurface interaction. 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 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, two of us were involved in the theoretical investigation of inter- and intrawell interaction effects in double quantum well heterostructures studied experimentally in Ref. 21. Let us point out key differences between the present paper and that work. First, in Ref. 20 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 well 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. 20, 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 non-linear sigma model (NL$\sigma$M) approach to capture the diffusive low-energy physics. Quantum corrections to the longitudinal conductivity $\sigma$ 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 NL$\sigma$M was originally developed by Finkel'stein in the eighties (for review articles see Refs. 24–26). In addition to perturbative RG treatment (which can also be performed diagrammatically) 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, the integer QHE, and the metal-insulator transition in Si MOSFETs.

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 non-monotonic 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 (Sec. III), we use the non-Abelian bosonization technique to map the fermionic theory on the $(U(1))$-gauged, interacting NL$\sigma$M with $Z_2$ topological term. Here we also discuss the Fermi liquid treatment of generally strong electron-electron interactions. Next, we renormalize the NL$\sigma$M 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 $\alpha$ 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 $\mu_1$ and $\mu_2$ 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 $\mu_1$ and $\mu_2$ lie well within the bulk gap $\Delta_{\text{bulk}}$. 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 section V B 1.

If the electrostatic gates are present and too close 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 gates and concentrate on the limit of long-range Coulomb interaction. In addition we derive general RG equations (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 is 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,

$$E \ll \min_{s=1,2} \frac{h}{\tau_s}. \quad (1)$$

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

$$\frac{h}{\tau_s} \ll |\mu_s|. \quad (2)$$

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

$$l \ll L_E, \quad (3)$$

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 $l_{scr}$, the (maximal) mean free path $l$ and the experimentally tunable scale $L_E$.

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

$$U_0 (r) = \frac{e^2}{\epsilon} \left( \frac{1}{\sqrt{r^2 + d^2}} - \frac{1}{r} \right). \quad (4)$$

The two dimensional vector $r$ connects the two dimensional positions of the particles, $r = |r|$, $e$ is the charge of the electrons, and $\epsilon$ denotes the effective dielectric constant.

Fourier transformation and RPA-screening leads to $U \equiv U(q) \equiv 2\pi e^2/\epsilon q$

$$U_{scr} (q) = \frac{U}{1 - (\Pi_1 + \Pi_2) U + U^2 \Pi_1 \Pi_2 \left(1 - e^{-2dq}\right)} \quad (5)$$

with

$$U = U \left(1 - \Pi_2 U \left(1 - e^{-2dq}\right) e^{-dq} \right) \left(1 - \Pi_1 U \left(1 - e^{-2dq}\right) \right).$$

Here $\Pi_s$ 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: $\Pi_s (\omega = 0, q) = -\nu_s$.

In the diffusive regime defined by the condition (3), the wavevector $q$ 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,

$$U_{scr} \approx \frac{2\pi e^2}{\epsilon} \left( \frac{1}{q + \kappa_s} \frac{0}{0 + \kappa_s} \right), \quad (6)$$

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 arises in the unitary limit when one can neglect $q$ as compared with $\kappa_s$ 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 section 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

$$U_{scr} \approx \frac{2\pi e^2}{\epsilon} \left( \frac{1}{q + \kappa_s + \kappa_2 + 2dq \kappa_2 (1 - qd)} \right) \times \left(1 + 2\kappa_2 d \right) \frac{1}{1 + 2\kappa_s d} \quad (7)$$
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 $\kappa^{-1}_1, \kappa^{-1}_2$ and $d$ the following scenarios are conceivable:

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

$$ (U_{scr})_{ss'}(q) \sim \frac{1}{q + \kappa_1 + \kappa_2} \Rightarrow l_{scr} = \frac{1}{\kappa_1 + \kappa_2}. \quad (8) $$

If in addition the condition $l_{scr} \ll l$ is fulfilled, the Coulomb interaction potential $(7)$ becomes “over-screened” ($q$-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 $U_{scr}$ 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 $\kappa_s^{-1} \ll \kappa$).

In this paper we derive the conductivity corrections in the unitarity limit of $q$-independent interaction, see Eqs. (94). As expected, in the limit of decoupled surfaces, $\kappa_s^{-1} \ll d$, they reproduce the previous result$^{19}$, 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 $d - \kappa^{-1}$. For simplicity, we 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.

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 $s$-th surface state and $v_F$ the Fermi velocity. Therefore

$$ \kappa_s^{-1} = \frac{1}{\alpha \kappa_1^{(s)}}. \quad (9) $$

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 $e/\epsilon v_F$ times the fine structure constant of quantum electrodynamics. Clearly, $\alpha$ 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 $\kappa_s^{-1}$ as a function of $k_F^{(s)}$. The actual value of $\kappa_s^{-1}$ for an exemplary case of Bi$_2$Se$_3$ (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$_2$Se$_3$ experiments$^{17,36–38}$ 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$^{39}$ and experimentally$^{40}$ 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 $d \sim 100$ nm.

On the other hand, surface electrons have a maximal Fermi wavevector of $k_F \sim 0.1/\AA$ 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 $\kappa_s^{-1}$ 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
by the following microscopic Matsubara action:

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

The notation \(\int_{\tau, x} = \int d^2x \int_0^\beta d\tau\) 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 \(\bar{\psi}(x, \tau) = (\psi^\dagger_1, \psi^\dagger_2, \psi^\dagger_3)\) and \(\psi(x, \tau) = (\psi_1^\dagger, \psi_2^\dagger, \psi_3^\dagger)^T\) describe the spinful (\(\uparrow, \downarrow\)) excitations living on surfaces \(s = 1\) and \(s = 2\). The one particle Hamiltonian which characterizes the surface \(s\) is

\[
(H_0 + H_{\text{dis}})_s = (V_s(x) - \mu_s) \otimes I_\tau + i(-)^s v_F^s \nabla \hat{\sigma} , \tag{12}
\]

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

\[
\langle V_s(x) V_s'(x') \rangle = \frac{\delta(x - x')}{\pi \nu_s \tau_s}. \tag{13}
\]

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.

- First, the model (and its analysis below) corresponds to the general case in which the chemical potentials \(\mu_1, \mu_2\) and hence the carrier densities of the two surfaces may differ.

- Second, since the disorder potentials are different for two surfaces, no inter-surface diffusion 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 inter-surface correlations of disorder. Indeed, any mismatch in chemical potentials and/or disorder configurations leads to an energy gap in the inter-surface 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 
\[ \left\langle |V_1(x) - V_2(x')|^2 \right\rangle \ll \sum_{s=1,2} (V_s(x)V_s(x')); \]
(ii) all other parameter regimes, when at least one of the conditions in (i) is not fulfilled.

Our model is designed for the case (ii), where the gap is comparable to the elastic scattering rate and inter-surface 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.

- Third, $\sigma$ in Eq. (12) in general does not describe the physical spin. For example, in Bi$_2$Se$_3$ structures the effective spin $\sigma$ 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. In this case also the Fermi velocity becomes anisotropic.
- Fourth, because of interaction effects, the true dispersion relation is not linear but contains logarithmic corrections (or more generally is subjected to “ballistic” RG) which leads to dependence of the Fermi velocity on the chemical potential. This is reflected in the notation $v_F^{(s)} \equiv v_F(\mu_s)$.
- Similarly, also the strength of the disorder may be substantially different for both surfaces, so that the (quantum) mean free times $\tau_s$ 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, $\tau_s$ acquire renormalization corrections, leading to a logarithmic dependence on $\mu_s$. In this case also the Fermi velocity becomes anisotropic.
- The (pseudo-)spin texture on the top and bottom surfaces is opposite (denoted by the factor $(-)^s$).
- Finally, in some materials (in particular, in Bi$_2$Te$_3$), 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, it has been shown that warping only influences the dephasing length (i.e., the lengthscale at which the RG flow is stopped).

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

\[
S_{\text{int}} = \frac{1}{2} \sum_{s,s'} \int_{\tau,x,x'} \rho_s(\tau, x) U_{0,s,s'}(|x-x'|) \rho_{s'}(\tau, x').
\]

For equal surfaces ($v_F^{(1)} = v_F^{(2)}$), a simple rescaling of equations (11) and (14) shows that the effective coupling to the Coulomb interaction is $\alpha$. 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 $\tau^{-1}_s$, see Sections III F 3, III F 4 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. Examples are the Stoner instability as well as more exotic phenomena such as topological exciton condensation, 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. SIGMA-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 diffusive and cooperon modes. In this Section we derive the effective field theory – diffusive non-linear $\sigma$ model – that describes the system in this regime.

#### A. Symmetries of the action

The structure of the effective low-energy theory, the diffusive NL$\sigma$M, 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 $U(1) \times U(1)$ transformations (global in space and time).

The presence of Coulomb interaction promotes the $U(1)$ symmetry in the total-density channel, $\rho_1 + \rho_2$, to transformations which are local in time but global in space. In other words, rotations of fermionic fields, $\bar{\psi}_s(\tau, x) \rightarrow \bar{\psi}_s(\tau, x) \exp[-i\chi(\tau)]$, $\psi_s(\tau, x) \rightarrow \exp[i\chi(\tau)]\psi_s(\tau, x)$, with equal phases $\chi(\tau) = \chi(\tau)$ leave the action (11) invariant. This is a special case of “$F$-invariance” and has important consequences for the present problem. The $F$-invariance (it is intimately linked to gauge invariance) generally states that in each channel with long-range interaction, time-dependent but
spatially constant $U(1)$ rotations are symmetries of the action. In our problem, as it follows from the $q \to 0$ limit of the Coulomb interaction:
\[ U(q) \propto -q^2 1 \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \]
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 C4.

To make the time-reversal symmetry explicit, we define particle-hole bispinors by combining $\psi$ and $\bar{\psi}$ fields.\(^{25,56}\) In the momentum space the bispinors read
\[ \Phi_n(k) = \frac{1}{\sqrt{2}} \begin{pmatrix} \psi_n(-k)^T \\ i\sigma_y\psi_n(k) \end{pmatrix}, \]
and
\[ \tilde{\Phi}_n(k) = [C\Phi_n(-k)]^T \text{ with } C = i\sigma_y\tau_x, \]
where $n$ is the index associated to the fermionic Matsubara frequency $i\epsilon_n$, and $\tau$ matrices act in the particle-hole space. This allows us to rewrite the one-particle Hamiltonian as
\[ S_{\text{free}} = -\sum_n \int_k \bar{\Phi}_n(k)(i\epsilon_n - HT(-k))\Phi_n(k). \]

It is convenient to perform a rotation of bispinors
\[ \eta = \sqrt{\tau_x}\Phi, \]
where $\sqrt{\tau_x} = e^{-i\eta/(1 + i\tau_x)}\sqrt{2}$. The free action then takes the form
\[ S_{\text{free}} = -\sum_s \int_{\tau_+} \eta_s^T \left\{ i\dot{\eta} - V_s + \mu_s \right\}(i\sigma_y) \eta_s. \]

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

In order to perform the average over disorder, we replace the theory $N_R$ times. Furthermore, in order to implement the $U(1)$-gauge invariance in the framework of the NL$\sigma$M, we apply a double cutoff truncation procedure with $N_M \ll N_M'$ for the Matsubara frequencies.\(^{55}\) 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 NL$\sigma$M) degrees of freedom, respectively. As a consequence, $\eta$ becomes a $(2\tau \times 2\tau \times 2\tau \times 2N_M' \times N_R)$-dimensional Grassmannian vector field. Except for the frequency term, the free action (21) is manifestly invariant under global orthogonal rotations of the kind
\[ \eta_s \to (O_s \otimes I_{\tau})\eta_s \text{ with } O_s \in O(2\tau \times 2N_M' \times N_R). \]

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):
\[ \Sigma^s_n = \frac{-2i\sigma_y}{\pi\nu_s\tau_s} \langle \eta_x,s|\eta_x,s\rangle\text{SCBA}. \]

Here $\langle \ldots \rangle_{\text{SCBA}}$ denotes the self-consistent treatment, i.e. a shift $\mu_s \to \mu_s + \Sigma^s_n$ in the fermionic propagator. Equation (23) yields for the imaginary part of the self-energy $\text{Im}(\Sigma^s_n) = (i/2\tau_s)\text{sgn}(n)$. The quasiclassical Drude DC conductance of the non-interacting problem in the absence of a magnetic field is
\[ \sigma_s^D = 2\pi\nu_s D_s e^2/h, \]
with $D_s = (\nu_s^{(s)})^2\tau_s$. Note that the transport time is twice the quantum mean free time $\tau_s$. In the diagrammatic language, this is a consequence of vertex corrections.

### C. Fermionic currents and bosonization rules

To derive the NL$\sigma$M, we use the method of non-Abelian bosonization.\(^{57-61}\) An advantage of this approach is that non-trivial topological properties of the Dirac fermions are translated into the field theory in a particularly transparent way.

In the first step, the kinetic term (Sec. III D) is bosonized. Subsequently, we bosonize also the terms induced by the chemical potential, disorder and frequency (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 ($\eta_L \to O_L\eta_L$) and right ($\eta_R \to O_R\eta_R$) rotations define the left and right currents. The bosonization rules for these currents as well as for the mass term are
\[ j_+ = v_F\eta_1\eta_1^T \leftrightarrow \frac{1}{8\pi} (O\partial_t O^T), \]
\[ j_- = v_F\eta_2\eta_2^T \leftrightarrow \frac{1}{8\pi} (O^T\partial_- O), \]
\[ \eta_1\eta_1^T \leftrightarrow i\lambda O, \]
where $\partial_\pm = \partial_x \pm i\partial_y$. The energy scale $\lambda$ is of the order of the ultraviolet (UV) cutoff and is introduced here for dimensional reasons; see Sec. 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 $(2 \times 2 N_M \times N_R) \times (2 \times 2 N_M \times N_R)$ matrix field. Below we will need the following constant matrices in this space
\[
\Lambda_{nm}^{\tau_2;\alpha; \beta} = \text{sgn} (n) \delta_{\tau_2}^{\alpha} \delta_{\beta}^{\tau} \delta_{nm},
\]
\[
\delta_{nm}^{\tau_2;\alpha; \beta} = \delta_{\tau_2}^{\alpha} \delta_{\beta}^{\tau} \delta_{nm},
\]
\[
(\Gamma_{nm}^{\sigma})^{\tau_2;\alpha; \beta} = \delta_{\tau_2}^{\alpha} \delta_{\beta}^{\tau} \delta_{n-m},
\]
Here and throughout the paper we use a convention that $\alpha, \beta \in \{0, N_R\}$ denote replicas and $m, n \in \{-N_M, \ldots, N_M - 1\}$ Matsubara indices. The double cutoff regularization scheme\cite{55} prescribes that matrices $O$ have non-trivial 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 model manifold outside this low-energy region. As explained below, $O_0 = \Lambda$.

**D. Bosonization of the kinetic part**

The kinetic part of (21) is nothing but the Euclidean counterpart of the microscopic action (27) is the bosonized counterpart of the Wess-Zumino-Novikov-Witten (WZNW) action
\[
S_{WZNW} = \frac{1}{16\pi} \int \text{tr} \nabla O \nabla O^{-1} + \frac{i}{24\pi} \Gamma_{WZ},
\]
where $\Gamma_{WZ}$ is the Wess-Zumino (WZ) term
\[
\Gamma_{WZ} = \int \epsilon_{\mu \nu \rho} \text{tr} \left[ \left( \partial^{-1} \partial_{\mu} \partial_{\nu} \right) \left( \partial^{-1} \partial_{\rho} \partial_{\nu} \right) \right],
\]
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 $\partial (x, w)$ satisfying $\partial (x, w = 0) = \text{const}$ and $\partial (x, w = 1) = \partial (x)$. As a result the compactified two-dimensional coordinate space $\mathbb{R}^2 \cup \{ \infty \} \simeq S^2$ is promoted to the solid 3-ball $B^3$ (i.e., the “filled” sphere).

**E. Free NL$\sigma$M 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
\[
\delta S = 2 \int \text{tr} \left[ (i \dot{\epsilon} + \mu) \eta \eta^T \right] \leftrightarrow -2\lambda \int \text{tr} (\dot{\epsilon} - i \mu) O.
\]

Upon disorder averaging and bosonization, the term with random potential provides the following contribution to the field theory:
\[
\delta S_{\text{dis}} = -\frac{1}{\pi \nu} \lambda \Gamma (\eta \eta^T) + \frac{1}{\pi \nu} \int \text{tr} (\eta \eta^T)^2
\]
\[
\leftrightarrow \frac{\lambda^2}{2\pi \nu} \int \text{tr} (O^T - O)(O^T - O).
\]

As we see, disorder induces mass terms for $O$-matrices. Both mass terms in Eq. (30) are strictly non-negative.

**2. Free NL$\sigma$M with $Z_2$ topological term**

As we have just discussed, we keep only the soft modes
\[
Q = O^T \Lambda O \text{ with } O \in G.
\]
The subscript $\sigma$ will be omitted in the remainder. The NL$\sigma$M manifold $M = G/K$. We also rename the coupling constants according to the conventional notation of diffusive NL$\sigma$Ms and restore the surface index $s$,

$$S^{\text{free}} = \sum_s \int_x \frac{\sigma_s}{16} \text{tr} (\nabla Q_s)^2 - 2\pi T z_s \text{tr} [iQ_s] + iS^{(\theta)}_s. \quad (36)$$

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

The non-trivial second homotopy group of the NL$\sigma$M manifold $\pi_2(M) = Z_2$ 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 $S^{(\theta)}_s$ appearing in (36) only distinguishes between an even ($S^{(\theta)}_s = 0 \mod 2\pi$) and odd ($S^{(\theta)}_s = \pi \mod 2\pi$) number of instantons.

Such a $Z_2$ 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 NL$\sigma$M approach.

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

$$S^{(\theta)}_s = \frac{1}{24\pi} \Gamma_{WZ,s}\hat{\partial}_s(x,w=1)=Q_s(x)\partial_x (x). \quad (37)$$

Note that, since the second homotopy group of the NL$\sigma$M manifold is non-trivial, the definition of the WZ term requires that away from $w = 1$ the extended fields can take values in the big orthogonal group $G$.

To check that Eq. (37) is indeed the $Z_2$ theta-term, we proceed in the same way as was recently done for symmetry class CII.\textsuperscript{62} First of all, it is straightforward to check that $S^{(\theta)}_s$ is invariant under small variations of the sigma-model field, $Q_s \to Q'_s = Q_s + \delta Q_s$ ($Q'^2_s = Q^2_s = 1$). Thus, $S^{(\theta)}_s$ 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 $S^{(\theta)}_s$ also returns the correct value $S^{(\theta)}_s = \pi \mod 2\pi$ in the topologically non-trivial sector, it is sufficient to insert a single instanton into $S^{(\theta)}_s$. 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 NL$\sigma$M manifold for the larger orthogonal group. A direct calculation shows that the group volume covered while untwisting indeed yields the value $iS^{(\theta)}_s = i\pi$, see Appendix A 2).

There have been alternative derivations of the $Z_2$ term before\textsuperscript{63,64}. 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 response of the system to an external electric field requires coupling of the diffusive matter fields to $U(1)$ 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 Section III G 4 that such a procedure yields the correct linear response theory for the anomalous quantum Hall effect of Dirac fermions.

In addition to a non-trivial second homotopy group $\pi_2$, the sigma model manifold of the class AII possesses also a non-trivial first homotopy group, $\pi_1(M) = Z_2$. For this reason, the RG flow in 2D systems of class AII (as well as in other classes with a non-trivial $\pi_1$ group, namely AIII, BDI, CII, and DIII) is affected by vortices, as was shown in Ref. 62. In the case of AII (and DIII) class these are $Z_2$ vortices,\textsuperscript{62} i.e., a vortex is identical to an anti-vortex. In a recent work\textsuperscript{65} it was argued that such vortices are crucial for establishing localization in the class AII. Conversely, the robustness of a non-localized state on the surface of a weak topological insulator and of the critical state separating 2D trivial and topological insulator were explained by vanishing of the corresponding fugacity.

On the surface of a strong 3D TI, the effect of vortices is erased by the $Z_2$ topological term, in the same way as argued previously\textsuperscript{62} for the case of the symmetry class CII. Specifically, due to the $Z_2$ 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 NL$\sigma$M

In the previous subsection we have derived the diffusive non-linear sigma model for non-interacting 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:

$$S_{\text{int}} = \frac{T}{2} \sum_{m,\alpha;ss} \int x \text{tr} (I_m^\alpha \psi_s \bar{\psi}_s) U^q_{ss'} (I_m^\alpha \psi_{s'} \bar{\psi}_{s'})$$  \hspace{1cm} (38)

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

$$\text{tr} I_m^\alpha \psi_s \bar{\psi}_s = \text{tr} I_m^\alpha (1 - \tau_y) \eta_s,\eta_{s'}^T - \text{tr} I_m^\alpha (1 - \tau_y) \eta_{s'},\eta_{s}^T$$
$$\leftrightarrow i\lambda \left[ \text{tr} I_m^\alpha (1 - \tau_y) (O_s + O_{s'}^T) \right].$$  \hspace{1cm} (39)

When disorder is introduced, the matrices $O$ become restricted to the sigma-model manifold $M$, and we obtain

$$S_{\text{int}} = -\lambda^2 ST \sum_{m,\alpha;ss} \int x \text{tr} [J_m^\alpha Q_s] U^q_{ss'} \text{tr} [J_m^\alpha Q_{s'}].$$  \hspace{1cm} (40)

Here we have defined $J_n^\alpha = I_n^\alpha + iT_n^\alpha$. 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

$$P_s = |\mu_s, p\rangle \langle \mu_s, p|$$

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

$$S_0^{(s)} = -\sum_s \int dp \bar{\zeta}_s(p) [i\dot{\zeta} + \text{sgn}(\mu_s) (|\mu_s| - v_F p_s |p|)] \zeta_s(p),$$  \hspace{1cm} (42)

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

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:

$$S_{\text{int}} = -\frac{T}{2} \int_{P_1, P_2, K} \sum_{s_1 s_2} \left[ \mathcal{O}_{0+1}^A + \mathcal{O}_2^A + \mathcal{O}_c^A \right]$$  \hspace{1cm} (43)

with

$$\mathcal{O}_{0+1}^A = \sum_{s_1 s_2} \left[ \bar{\zeta}_{s_1}^\alpha (P_1) \zeta_{s_1}^\alpha (P_1 + K) \right]$$
$$\times \Gamma_{s_1, s_2, \tilde{p}_1, \tilde{p}_2}^{0, 1, q} \zeta_{s_2}^\alpha (P_2) \zeta_{s_2}^\alpha (P_2 - K),$$  \hspace{1cm} (44)

$$\mathcal{O}_2^A = \sum_{s_1 s_2} \left[ \bar{\zeta}_{s_1}^\alpha (P_2) \zeta_{s_1}^\alpha (P_1 + K) \right]$$
$$\times \Gamma_{s_1, s_2, \tilde{p}_1, \tilde{p}_2}^{2, q} \zeta_{s_2}^\alpha (P_1) \zeta_{s_2}^\alpha (P_2 - K),$$  \hspace{1cm} (45)

and

$$\mathcal{O}_c^A = \sum_{s_1 s_2} \left[ \bar{\zeta}_{s_1}^\alpha (P_2) \zeta_{s_1}^\alpha (-P_1 - K) \right]$$
$$\times \Gamma_{s_1, s_2, \tilde{p}_1, \tilde{p}_2}^{c, q} \zeta_{s_2}^\alpha (-P_2 + K) \zeta_{s_2}^\alpha (P_1).$$  \hspace{1cm} (46)

Here the capital letters denote 2+1 momenta. The smallness of $K = (\omega_m, q)$ means that the following conditions hold $(\omega_m, q) \ll \left( |\mu_s|, p_F \right)$ 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 $\Gamma_{\tilde{p}_1, \tilde{p}_2}$).

We refer to the three scattering channels as small angle scattering channel ($\Gamma^{0+1}$), large angle scattering channel ($\Gamma^2$), and the Cooper channel ($\Gamma^c$). The quantities entering Eq. (43) are the static limit of the corresponding scattering amplitude, $\Gamma(\omega_m = 0, q)$. They already include static screening and do not acquire any tree-level corrections due to disorder.\textsuperscript{24,26} Exemplary diagrams are
The irreducible part $\Gamma^1$ also includes the short range interaction induced by the finite thickness of the 3D TI film (see Appendix B and C).

For the short-range interaction amplitudes ($\Gamma^1$, $\Gamma^2$, $\Gamma^c$), the static limit coincides with the “q-limit” $\Gamma^0 = \lim_{q \to 0} \Gamma(\omega_m = 0, q)$, see also Appendix C. It should be kept in mind that for the one-Coulomb-line-reducible part $\Gamma^0$ (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 $\Gamma^0$ 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. In order to find the conventional pole structure of the FL Green’s functions for the case $\text{sgn}(\mu_s) = -1$ one needs to transpose the bilinear form in action (42) and swap the notation $\zeta_s(\epsilon_n) \rightarrow \bar{\zeta}_s(\epsilon_n)$. If $\text{sgn}(\mu_1\mu_2) = -1$, this interchange of notations obviously happens in only one surface. In this case, the large-angle scattering amplitude $\Gamma^1_{12}$ and the Cooper-channel amplitude $\Gamma^c_{12}$ are interchanged. Even though this procedure illustrates the analogy between exciton condensation (divergence in $\Gamma^1_{12}$) and Cooper instability (divergence in $\Gamma^c_{12}$), in the following we choose to keep our original notation of $\zeta_s$ 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 (see Appendix C.) It is valid down to energy scales $\sim \tau_{1,2}^{-1}$ and therefore constitutes the starting point for the effective diffusive theory at lower energies, $T \ll \tau_{1,2}^{-1}$.

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

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.

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 $\alpha_s \in (0, 1)$. As usual in the context of disordered FLs, 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:

$$\Pi^c_{s_1,s_2} \equiv \lim_{\omega_m \to 0} \Pi_{s_1,s_2} (\omega_m, q = 0) = 0$$

and

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

Since these identities reflect the gauge invariance, they can not be altered during the RG procedure. Thus, the

FIG. 9: An example of contribution to a scattering amplitude in the Cooper channel.
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 \( \Gamma^1, \Gamma^2 \) and \( \Gamma^c \). For electrons in metals, one has also to include the long-range Coulomb interaction. Following Ref. 66, the associated scattering amplitude \( \Gamma^0 \) 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 NL \( \sigma \) theory (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. In particular, only the zeroth angular harmonic of the scattering amplitudes survives in the diffusive limit.

The scattering amplitude \( \Gamma^2_{12} \) (as well as \( \Gamma^c_{12} \)) 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 \( \Gamma^c_{12} \) and \( \Gamma^c_{12} \) any longer. As one can see from figures 6 - 8, the large angle scattering amplitudes \( \Gamma^c_{11} \) and \( \Gamma^c_{22} \) cannot be distinguished from the small angle scattering amplitudes \( \Gamma^c_{0+1} \) and \( \Gamma^c_{0+1} \), respectively. We incorporate the effect of \( \Gamma^c_{11} \) and \( \Gamma^c_{22} \) into the “singlet channel”, which has the following matrix structure in the surface space

\[
\Gamma^p = \begin{pmatrix}
\Gamma^{0+1-2}_{11} & \Gamma^{0+1}_{12} \\
\Gamma^{0+1}_{12} & \Gamma^{0+1-2}_{22}
\end{pmatrix}.
\]

Here we used

\[
\Gamma^{0+1-2} = \Gamma^{0+1} - \Gamma^2.
\]

The intrasurface Cooper channel interaction \( \Gamma^c_{11} \) 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. 24 and supplementary material of Ref. 19). Consequently we do not consider the superconductive instability in this work. For the opposite case of attraction in the Cooper channel Coulomb interaction suppresses the transition temperature \( T_c \). The difference between Coulomb and short-range repulsive interaction was addressed in Ref. 69.

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 interaction. 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 \( Z_2 \) 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 \( Z_2 \) theta term is a consequence of the chiral anomaly and thus the only term determined by energies far from \( \mu \). However, it is well known that anomalies in quantum field theories are insensitive to interactions. Hence, the \( Z_2 \) term in the diffusive NL \( \sigma \) 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 NL \( \sigma \) 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,

\[
\Gamma^p \rightarrow -\Gamma^p
\]

in Eq. (40).

7. **Bare value of scattering amplitudes**

According to the formal FL treatment (Appendix C4), the singlet-channel interaction amplitude is given by

\[
\nu \Gamma^p \nu = -\nu \frac{\det \Pi^q}{\Pi^q_{11} + \Pi^q_{22} + 2\Pi^q_{12}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},
\]

where \( \Pi^q \) is the renormalized polarization operator. The value of the renormalized polarization operator is obtained by means of the FL renormalization of the static polarization operator (see Sec. III F 7).
where \((\nu)_{ss'} = \nu_s \delta_{ss'}\) and
\[
\Pi^a = -\nu - \nu \left( \begin{array}{ccc}
\Gamma_{11}^{1-2} & \Gamma_{12}^{1-2} \\
\Gamma_{12}^{1-2} & \Gamma_{22}^{1-2}
\end{array} \right) \nu. \tag{53}
\]

Here \(\Gamma^{1-2} = \Gamma^1 - \Gamma^2\). The remarkably simple matrix structure of \(\nu + \nu \Gamma^a \nu\) is actually due to the presence of the long-range Coulomb interaction. This fact will be explained by means of \(F\)-invariance in section 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:
\[
S = \sum_s \left[ S_s^{(\text{kin})} + i S_s^{(\theta)} \right] + S^{(\eta + \text{int})}. \tag{54}
\]

It contains the kinetic term
\[
S_s^{(\text{kin})} = \frac{\sigma_s}{16} \int_x (\nabla Q_s)^2 \tag{55}
\]
and the \(Z_2\) theta term
\[
S_s^{(\theta)} = \frac{1}{24\pi} \Gamma^s \delta_{s,l=1} = Q_s(x) = Q_T(x) \tag{56}
\]
for each of the surfaces, as well as the frequency and interaction terms,
\[
S^{(\eta + \text{int})} = -\pi T \left[ \sum_{s} 2z_s \text{tr}\hat{\eta} Q_s \\
- \sum_{ss',n,\alpha} \text{tr} \left[ J_{n,s}^\alpha Q_s \right] \Gamma_{ss'} \text{tr} \left[ J_{n,s}^\alpha Q_{s'} \right] \right]. \tag{57}
\]

Here we have introduced the notation
\[
\Gamma_{ss'} = \frac{8}{\pi} \lambda_s \Gamma_{ss'} \lambda_{s'}. \tag{58}
\]

G. Inclusion of scalar and vector potentials into the NLσM

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

1. Electromagnetic gauge invariance

We include the scalar potential \(\Phi_s\) 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 \(\psi\) 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 \(\psi\)-fields imply the following rotation of bispinors:
\[
\eta_s(x) \rightarrow W_s \eta_s(x), \tag{59}
\]
where
\[
W_s = \left[ e^{-i\hat{x} \tau^1 + \tau^2} \frac{1 + \tau_1}{2} + e^{i\hat{x} \tau^1 - \tau_2} \frac{1 - \tau_1}{2} \right] \tag{60}
\]
and we use the following convention for hatted matrices: \(\hat{a} \equiv \sum_{n,\alpha} \alpha_n^\alpha \Gamma_n^\alpha\). Let us recall that the \(\eta_s\) 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. \(F\)-algebra and \(F\)-invariance

As a direct consequence of (59), \(Q\)-matrices transform under a gauge transformation \(\chi_s\) in the following way:
\[
Q_s \rightarrow W_s Q_s W_s^T. \tag{61}
\]

Under such rotations, in the limit \(N'_M, N_M \rightarrow \infty, N_M/N'_M \rightarrow 0\), the frequency term acquires the correction
\[
\delta_s \text{tr} \hat{\eta} Q_s = 2 \sum_{n,\alpha} \left[ i n \chi^\alpha_{s,n} \text{tr} J_n^\alpha Q_s - n^2 \chi^\alpha_{s,n} \chi^\alpha_{s,-n} \right], \tag{62}
\]
while the factors entering the interaction term vary as follows:
\[
\delta_s \text{tr} J_{n,s}^\alpha Q_s = -i 2 n \chi^\alpha_{s,n}. \tag{63}
\]

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 \(\text{without} \) inclusion of gauge potentials (“\(F\)-invariance”). This symmetry has to be preserved on NLσM level, implying that
\[
(z + \Gamma) \left( \begin{array}{c} 1 \\ 1 \end{array} \right) = 0. \tag{64}
\]

Here \((z)_{ss'} = z_s \delta_{ss'}\). Since the intersurface interaction is symmetric, \(\Gamma_{12} = \Gamma_{21}\), Eq. (64) yields
\[
(z + \Gamma) = \text{const.} \times \left( \begin{array}{cc} 1 & -1 \\ -1 & 1 \end{array} \right). \tag{65}
\]

This relation is consistent with Eq. (52). However, contrary to Eq. (52), the relation (65) is manifestly imposed by the symmetry (“\(F\)-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 \( \partial_\mu Q_s \rightarrow D_\mu Q_s \), with the long derivative \( D_\mu \) of the form

\[
D_\mu Q_s \equiv \partial_\mu Q_s + \sum_{n,\alpha} i A^\alpha_{\mu, s, - n} \left[ J^\mu_n - (J^\mu_n)^T, Q_s \right].
\]

(66)

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

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

\[
S^\Phi = -2 \sum_{n,\alpha, ss'} \Phi^\alpha_{n, ss'} (z + \Gamma)_{ss'} \text{tr} J^\alpha_n Q_s + 1 \pi T \sum_{n,\alpha, ss'} \Phi^\alpha_{n, ss'} \text{tr} \overline{\Phi}^\alpha_{n, ss'}.
\]

(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,

\[
\Pi^\text{RPA} (\omega_n, q) = -2 \pi (z + \Gamma)_{ss'} \text{tr} J^\alpha_n Q_s + 4T \sum_{s_1, s_2} (z + \Gamma)_{ss_1} \text{tr} J^\alpha_n Q_s (q) \times
\]

\[ \times \text{tr} J^\alpha_n Q_s (q) \times \text{tr} J^\alpha_n Q_s (q) \times \text{tr} J^\alpha_n Q_s (q).
\]

(68)

Here \((...)\) 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 \( \omega_n \):

\[
\sigma_{ss'} (\omega_n) = B_1^{(ss')} (\omega_n) \delta_{ss'} + B_2^{(ss')}.
\]

(69)

Here we introduced two correlators:

\[
B_1^{(ss')} (\omega_n) = \frac{\alpha}{8n} \left\{ \text{tr} \left[ J^\alpha_n - (J^\alpha_n)^T, Q_s \right] \left[ J^\alpha_n - (J^\alpha_n)^T, Q_s \right] \right\},
\]

(70)

and

\[
B_2^{(ss')} = \frac{\alpha \sigma}{128n} \sum_{x, y, \mu, \nu} \left\{ \text{tr} \left[ J^\alpha_n - (J^\alpha_n)^T, Q_s \right] \partial_\mu Q_s \right\} \times
\]

\[ \times \text{tr} \left[ J^\alpha_n - (J^\alpha_n)^T, Q_s \right] \partial_\nu Q_s \right\}.
\]

(71)

Substituting the saddle-point value \( Q_s = \Lambda \), we obtain the classical value \( \sigma_{ss'} (\omega_n) = \alpha \delta_{ss'} \). 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 \( Z_2 \) theta term, i.e., the WZW-term, Eq. (56), also allows of inclusion of gauge potentials.70–75 However, the situation is more subtle here. Specifically, it turns out that the contribution of non-singular gauge potentials to the topological term \( S^{(\theta)} \) vanishes. We explicitly show this in Appendix A1.

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 with quantized transverse conductance

\[
\sigma_{xy} = g \left( n \pm \frac{1}{2} \right) \frac{e^2}{h}, n \in \mathbb{Z},
\]

(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.77–80 Theoretically, the anomalous quantum Hall effect was explained and discussed in a previous work by three of the authors.81 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,82 which assumes the following form upon inclusion of the vector potential:55

\[
S^{\text{QHE}} = \frac{\vartheta}{16\pi} \int_x \epsilon_{\mu\nu} \text{tr} Q_U \partial_{\nu} Q_U \partial_\mu Q_U + \frac{\vartheta}{4\pi} \int_x \epsilon_{\mu\nu} \text{tr} \partial_{\nu} A_{\mu, Q_U} \]

\[ + \frac{\vartheta}{4\pi} \int_x \epsilon_{\mu\nu} \sum_{n,\alpha} n A^\alpha_{\mu, n} A^\alpha_{\mu, - n}.
\]

(73a)

Here \( Q_U = U^{-1} A_U U \) with \( U \in U (2N_M \times N_R) \), \( \epsilon_{\mu\nu} = -\epsilon_{\nu\mu} \) is the 2D antisymmetric symbol \((\epsilon_{xy} \equiv 1)\), and \( \vartheta \) is the theta angle of the Pruisken’s NLσM. We emphasize, that the last two terms (Eqs. (73b) and (73c)) determine the effective electromagnetic response and thus prescribe the relation between the physical observable \( \sigma_{xy} \) (in units of \( e^2/h \)) and the theta angle \( \vartheta \). In particular, \( \vartheta /2\pi \) is identified as the bare value of the Hall conductance.83

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:

$$\mathcal{M} \rightarrow U(2N_M N_R) / U(N_M N_R) \times U(N_M N_R).$$  (74)

The $\mathbb{Z}_2$ theta term becomes the Pruisken’s theta term\cite{84} (recall $\theta = \pi \bmod 2\pi$)

$$S^{(\theta)}_U = \frac{\theta}{16\pi} \int d^2 x \epsilon_{\mu\nu} \partial_\mu Q_U \partial_\nu Q_U.$$  (75)

We emphasize that together with the gauged kinetic term $S^{(\theta)}_U$ is the complete gauged theory, no extra terms of the type (73b) and (73c) appear. Being topological, the Pruisken’s theta term is invariant under smooth $U(1)$ rotations. Recall that exactly the terms (73b) and (73c) provided a link between $\theta$ and $\sigma_{xy}$ 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 non-zero. 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 $\theta + \bar{\theta}$. On the other hand, only $\theta$ is related with the bare value of $\sigma_{xy}$. Then standard arguments for the quantization of the Hall conductivity\cite{31} leads to the result (72) for the anomalous QHE.

IV. ONE-LOOP RG

In the preceding section we have derived the diffusive NLσM, Eqs. (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

$$S_L = -\sum_{s=1,2} \frac{\sigma_s L^{-2}}{8} \int d^2 x \text{tr} \Lambda Q_s.$$  (76)

The connection between the running length scale $L$ and the physical regulators temperature or frequency was analyzed in Ref. [85]. 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 $O(2 \tau \times N_M \times N_R) \times O(2 \tau \times N_M \times N_R)$-symmetry of the $Q$-matrix (35) and to ensure the renormalizability of the theory.

A. Diffusive propagators

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

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

anticommute with $\Lambda$. Further, we define a set of real matrices in the particle-hole space: $\tilde{\tau}_\mu \equiv 2^{-1/2} (1, \tau_x, i\tau_y, \tau_z)$. This allows us to introduce the fields $q^{(\mu)} \equiv \text{tr}^\tau q^\tau_{\mu}$, where $\text{tr}^\tau$ 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 $q^{(2)}$ describe cooperons. Their propagator is unaffected by interaction (since we have discarded the interaction in the Cooper channel),

$$\left< q^{(\mu)}_s(p) q^{(\nu)}_{s'}(-p) \right> = \frac{4}{\sigma_s} D_s (\omega_{n_{12}}, p) \delta_{ss'} \times \delta_{\mu\nu} \delta_{n_{11}} \delta_{n_{22}} \delta_{\alpha_1\beta_1} \delta_{\alpha_2\beta_2} (\delta_{\mu 1} + \delta_{\mu 3}),$$  (77)

where

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

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

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

$$\left< q^{(\mu)}_s(p) q^{(\nu)}_{s'}(-p) \delta_{n_{11}} \right> = \frac{4}{\sigma_s} D_s (\omega_{n_{12}}, p) \times \frac{8\pi T}{\sigma_s} \delta_{\alpha_1\alpha_2} \left( \Gamma D^\tau (\omega_{n_{12}}, p) \right)_{ss'}. $$  (79)

Here we have introduced

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

The bare action contains, aside from the mass $L^{-1}$, seven running coupling constants: $\sigma_1$, $\sigma_2$, $z_1$, $z_2$, $\Gamma_{11}$, $\Gamma_{22}$ and $\Gamma_{12}$. 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:

$$\Pi^{RPA}_{ss} (\omega, p) = -\frac{2}{\pi} \left[ \hat{z} + \hat{\Gamma} \right] (1 - 4\omega \sigma^{-1} D^\nu \omega) \left[ \hat{z} + \hat{\Gamma} \right]$$

(81)

where $\sigma_{ss'} = \sigma \delta_{ss'}$. 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:

$$\Pi^{RPA} = \left[ \Pi^0 - \nu \Gamma^0 p^2 \right] (1 + \omega \Delta^\nu (\omega, p) \left[ \Pi^0 - \nu \Gamma^0 p^2 \right])$$

(82)

where

$$\Delta^\nu (\omega, p) = \left[ \nu D^p p^2 + \omega (\nu + \nu \Gamma^0 p^2 \mathbf{q}) \right]^{-1}.$$

(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 NL$\sigma$M coupling constants with physical FL parameters:

$$\frac{2}{\pi} (\hat{z} + \hat{\Gamma}) = -\Pi^0 + \nu \Gamma^0 \nu.$$

(84)

Next, from comparison of momentum dependence in Eqs. (81) and (82), we find the Einstein relation: $\sigma_4 = 2\pi \nu D_4$. Accordingly, $\sigma$ measures the conductance in units of $e^2/h$, consistently with what has been found in Secs. III B and III G 3.

In view of gauge invariance (Sec. III F 4), the static polarization operator entering Eq. (84) is nothing but the compressibility

$$\Pi^{\nu_{ss'} \nu_{ss'}} = -\frac{\partial N_s}{\partial \mu_{ss'}}.$$

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. Since $\nu \Gamma^0 p^2$ only depends on $\Pi^0$ (see Appendix C 4) it is not renormalized as well. Therefore, the right-hand side of (84) is not renormalized and hence neither is its left-hand-side, i.e., $\hat{z} + \hat{\Gamma}$. This matrix constraint yields three RG invariants: $z_1 + \Gamma_{11}$, $z_2 + \Gamma_{22}$, and $\Gamma_{12}$. Thus, only four out of seven NL$\sigma$M 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 $\Pi^0 = -\nu - \nu \Gamma^{1 - 2 \nu}$ and using the definition of $\nu$ in Sec. III E 1 one can find the following relations for the bare values

$$\frac{4\lambda_s}{\pi} \equiv \frac{2}{\pi} z_s = \nu_s.$$

(85)

Equivalently, the same relationship between $\lambda_s$ and $\nu_s$ can be obtained by comparing the bare definition of $\Gamma$ [Eq. (58)] with the right hand side of (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 $B_1^{(s)}$, Eq. (70). The one-loop correction is determined by the expansion to second order in $q^{(s)}$. 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

$$B_1^{(s)} = \sigma_s + 2 \int D \omega \langle \omega, n, p \rangle.$$  

(86)

We evaluate this term in the announced regularization scheme:

$$B_1^{(s)} = \sigma_s + 2 \Pi^{(2+\epsilon)} \left[ \sigma_s + \frac{1}{2\pi} \left( -\frac{2}{\epsilon} + 2 \ln \frac{L}{l} + \text{const.} \right) \right].$$

(88)

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

$$I_1^{(D)} = \int_0^{2\pi} \frac{d^D p}{(2\pi)^D} \frac{1}{p^2 + L^2} = \frac{(2\pi)^D}{(4\pi)^D} \Gamma \left( 1 - \frac{D}{2} \right) = \frac{1}{4\pi} \frac{2 + 2 \ln \frac{l}{L} + \ln \frac{4\pi - \gamma + O(\epsilon)}{2}}{2}.$$
where $\gamma \approx 0.577$ is the Euler-Mascheroni constant.

The logarithmic term in Eq. (88) is nothing but the well-known weak-antilocalization effect.\(^{86}\)

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 $\sigma_{ss}$. The correlator $B_2^{(ss')}$ can be recast into the form (see Appendix D)

\[
B_2^{(ss')} = \frac{16\delta_{ss'}}{n_{ss}} \int_p p^2 \sum_{\omega_m > 0} \omega_m \times \left[ \left(D\bar{D}\hat{D}^c\right)_{ss'}(\omega_m, p) D_s(\omega_{m+n}, p) \right. \\
- \left. \left(D\bar{D}\hat{D}^c\right)_{ss}(\omega_{m+n}, p) D_s(\omega_{m+2n}, p) \right]
\]  

(89)

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

\[
B_2^{(ss')} = -4\delta_{ss'} \left( 1 - \frac{1 + \gamma_{ss}}{\gamma_{ss}} \ln (1 + \gamma_{ss}) \right) \times \\
+ \left( 1 + \gamma_{ss} \right) \left( \frac{\ln (1 + \gamma_{ss})}{\gamma_{ss}} - \frac{\ln (1 + \gamma_{11})}{\gamma_{11}} \right) \int_2^{(2+\epsilon)} \\
- \frac{\delta_{ss'}}{\pi} \left( 1 - \frac{1 + \gamma_{ss}}{\gamma_{ss}} \ln (1 + \gamma_{ss}) \right) \times \\
\times \left[ \frac{2}{\epsilon} + 2 \ln L/l + \text{const} \right]. 
\]  

(90)

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

\[
\zeta^{(D)} = \int_2^{D-2} \frac{dDp}{(2\pi)^D} \frac{p^2}{(p^2 + L^{-2})^2} = \left( \frac{4\pi^2}{D} \frac{D-1}{2} \right) \Gamma \left( 1 - \frac{D}{2} \right) \frac{D=2+\epsilon}{\pi} \left[ \frac{2}{\epsilon} + 2 \ln L/l + \ln 4\pi - 1 - \gamma + O(\epsilon) \right]. 
\]

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.\(^{87}\) This is because the scale (e.g., temperature) dependence of the total thermodynamic potential $\Omega$ 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.\(^{85}\)

\[
z_1' + z_2' = \frac{1}{2\pi i \bar{n} \lambda} \frac{\partial \Omega}{\partial T}. 
\]

(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

\[
(z_1' + z_2') = (z_1 + z_2) + 2 \sum_{s=1,2} \Gamma_{ss} \int_p D_s(0, p). 
\]

(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):

\[
z_s' = z_s + 2\Gamma_{ss} \int_p D_s(0, p) \\
= z_s + 2\frac{\Gamma_{ss}}{\sigma_s} \left[ 1^{(2+\epsilon)} \right] \\
= z_s + 1 \frac{\Gamma_{ss}}{2\pi \sigma_s} \left[ \frac{2}{\epsilon} + 2 \ln L/l + \text{const} \right]. 
\]

(93)

We have directly proven this assumption of separate $z_s$ renormalization by the background field method.\(^{107}\)

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:

\[
\frac{d\sigma_1}{dy} = -\frac{2}{\pi} F \left( \gamma_{11}, \sigma_1/\sigma_2 \right), 
\]

(94a)

\[
\frac{d\sigma_2}{dy} = -\frac{2}{\pi} F \left( \gamma_{22}, \sigma_2/\sigma_1 \right), 
\]

(94b)

\[
\frac{d\gamma_{11}}{dy} = -\gamma_{11} (1 + \gamma_{11}), 
\]

(94c)

\[
\frac{d\gamma_{22}}{dy} = -\gamma_{22} (1 + \gamma_{22}), 
\]

(94d)

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

\[
F(\gamma, x) = \frac{1}{2} - \frac{1 + \gamma}{x [1 + \gamma (1 + 1/\pi)]} \ln [(1 + x) (1 + \gamma)]. 
\]

(95)
We recall that \( \Gamma_{12}, \gamma_{11}, \) and \( \gamma_{22}, \) are not renormalized. We mention that the mass \( L^{-1} \) acquires a quantum correction\(^{85}\) but it does not affect the one-loop renormalization of the other parameters \( \sigma_s, \gamma_z \) 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:

\[
\begin{align}
\frac{d\sigma}{dy} &= -\frac{2}{\pi} \left\{ 1 - \frac{1}{t} \frac{1 + \gamma_{11}}{1 + \gamma_{11} + \frac{1}{2}} \ln \left[ (1 + t)(1 + \gamma_{11}) \right] - t \frac{1 + \gamma_{22}}{1 + \gamma_{22} + (1 + t)} \ln \left[ \frac{1 + \frac{1}{t}}{1 + \gamma_{22}} \right] \right\}, \\
\frac{dt}{dy} &= -\frac{2}{\pi} \frac{1 + t}{\sigma} \left\{ 1 - \frac{1}{t} \frac{1 + \gamma_{11}}{1 + \gamma_{11} + \frac{1}{2}} \ln \left[ (1 + t)(1 + \gamma_{11}) \right] + t^2 \frac{1 + \gamma_{22}}{1 + \gamma_{22} + (1 + t)} \ln \left[ \frac{1 + \frac{1}{t}}{1 + \gamma_{22}} \right] \right\}, \\
\frac{d\gamma_{11}}{dy} &= -\left( 1 + \frac{1}{t} \frac{\gamma_{11}(1 + \gamma_{11})}{\pi \sigma} \right), \\
\frac{d\gamma_{22}}{dy} &= -(1 + t) \frac{\gamma_{22}(1 + \gamma_{22})}{\pi \sigma}.
\end{align}
\]

\[\text{V. ANALYSIS OF THE RG EQUATIONS}\]

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) \( \delta\sigma^\text{WAL} = (1/\pi) \ln L/l \) 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.

Equations (94) which determine the flow of the coupling constants \( \sigma_1, \sigma_2, \gamma_{11} \) and \( \gamma_{22} \) imply 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.

\[\text{A. Two equal surfaces}\]

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 \( \sigma \) and \( \gamma \) are

\[
\begin{align}
\frac{d\sigma}{dy} &= -\frac{2}{\pi} \left[ 1 - \frac{2 + 2\gamma}{1 + 2\gamma} \ln (2 + 2\gamma) \right], \\
\frac{d\gamma}{dy} &= -\frac{2\gamma(1 + \gamma)}{\pi \sigma}.
\end{align}
\]

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 (\( \epsilon_1 = \epsilon_3 \)).

\[\text{1. Flow Diagram within the fixed plane}\]

The RG flow within the \( \sigma-\gamma \) 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. 19 for a single surface of 3D TI. In this limit the total correction to the conductivity is negative and obeys the universal law

\[
\delta\sigma_{\gamma = -1} = 2 \times \frac{2}{\pi} \left( \frac{1}{\text{WAL}} - \frac{1}{\text{AA}} \right) \ln L/l = -\frac{2}{\pi} \ln L/l.
\]

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:

\[ \delta \sigma_{\gamma=0} = 2 \times \frac{2}{\pi} \left( \frac{1}{2} - \left[ 1 - \ln 2 \right] \right) \ln L/l. \]  \hspace{1cm} (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 \( \gamma_{12} \) is weaker than or equal to the intrasurface interaction \( \gamma \). This implies that its bare value \( \gamma_0 \) 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 \( \alpha \) we can approximate \( \gamma_0 \) by its RPA value:

\[ \gamma_0 = -\frac{1}{2} - \frac{1}{2} \frac{\kappa d}{1 + \kappa d}. \]  \hspace{1cm} (100)

Here \( d \) is the system thickness and \( \kappa = 2 \pi \frac{\alpha^2}{\nu} \) 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:

\[ \delta \sigma_{\gamma=-1/2} = \frac{2}{\pi} \left( 2 \times \left( 1/2 - 1 \right) \right) \ln L/l = 0, \]

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 consequence of intersurface interaction is the non-monotonic 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 \( y_\ast \) at which the conductivity reaches its minimum can be extracted from Eqs. (97) and is expressed by the integral

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

where \( f(x) = \text{Li}_2(-x) - \text{Li}_2(-1 + 2x) \), \( \text{Li}_2 \) is the dilogarithm, and \( \gamma_\ast = -1/2 \).

Numerical integration of (101) yields the crossover length scale or temperature \( y_\ast = \ln L_\ast/l = 1/2 \ln T_0/T_\ast \). Its dependence on the bare values \( \sigma_0 \) and \( \gamma_0 \) is plotted in Fig. 12. Using Eq. (100) one can also investigate the dependence of \( y_\ast \) on \( \kappa d \) instead of \( \gamma_0 \) (see inset in Fig.12).

3. Role of topology: Dirac electrons vs. 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 E 2, in the diffusive NLσM of Dirac electrons, the \( Z_2 \) theta term reflects the topological protection from Anderson localization. This term is absent in the case of non-topological symplectic metals (NTSM) such as electrons with quadratic dispersion subjected to strong spin-orbit coupling.\(^{108}\)

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 non-perturbative 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.\(^{19}\) For NTSM localizing AA corrections overcome the WAL effect and the system always flows towards localization (Fig. 13, left). In contrast, for TI the topological protection implies \( d\sigma/dy > 0 \) for small \( \sigma \) and hence an attractive fixed point at \( \sigma \sim 1 \) (Fig. 13, 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 (non-interacting) WAL effect. Its contribution \( 2 \times (1/\pi) \ln L/l \) is independent of \( \sigma \) only for \( \sigma \gg 1 \). For NTSM it decreases with decreasing \( \sigma \) and eventually becomes negative at the metal-insulator transition (MIT) point \( \sigma_{MIT} \approx 2 \times 1.42 e^2/h \).\(^{88-90}\) (As explained above, Sec. III E 2, in a recent investigation\(^{65}\) the crucial role of \( Z_2 \) 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 \( \sigma_{MIT} \sim 1 \). 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,\(^{64}\) i.e., the beta function \( d\sigma/dy \) bends up when \( \sigma \to 0 \). There is a numerical evidence\(^{91,92}\) that in a non-interacting 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 non-monotonic conductivity behavior leads to the metallic state. As the horizontal position in the phase diagram is controlled by the parameter \( k_d \), 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 \( \gamma \).

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 \( \gamma_{11} \) and \( \gamma_{22} \), Eqs. (96c) and (96d), determines four fixed planes of the RG flow:

- \( \gamma_{11} = -1 = \gamma_{22} \). Repulsive fixed plane of two decoupled surfaces with only intrasurface Coulomb interaction. This problem has been studied in Ref. 19.
- \( \gamma_{11} = 0, \gamma_{22} = -1 \) or vice versa. Fixed plane describing a 3D TI film with strongly different surface population. This case is analyzed in Sec. V B 1 below.
- \( \gamma_{11} = 0 = \gamma_{22} \). Attractive fixed plane. Intrasurface interaction has died out and only intersurface interaction survived. This case is analyzed in Sec. V B 2 below.

![FIG. 13: Comparison between expected RG-flow for a double layer system of NTSM (left) and the coupled surfaces of a thin 3D TI film (right).](image-url)
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 $\gamma_{11}$ and $\gamma_{22}$ 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 finite-range interaction (and thus the whole crossover between the problem with Coulomb interaction and the non-interacting system) is discussed in Appendix F.

1. Strongly different surface population

We investigate here the fixed plane of Eqs. (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 $\gamma$-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$_2$Se$_3$. 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 non-trivial regime. A situation as depicted in Fig. 14 is then believed to arise in a certain range of gate voltages: one of the two surfaces (here surface 1) is separated by a depletion region from a relatively thick bulk-surface layer.

Recently, disorder-induced interference corrections for 3D TI bulk electrons have been investigated theoretically. While at small length scales additional symmetries of the Hamiltonian provide non-trivial 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 NL$_\sigma$M, 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 Eqs. (96), we find that the RG equations in this fixed plane are as follows:

$$\frac{d\sigma}{dy} = -\frac{2}{\pi} \left\{ \frac{1}{1 - \frac{1}{\kappa_1} \ln(1 + t)} \right\}, \quad (102a)$$
$$\frac{dt}{dy} = -\frac{2 t + 1}{\pi \sigma} \left\{ \frac{1}{\kappa_2} \left\{ 1 - \frac{1}{\kappa_1} \ln \left( 1 + \frac{\sigma_1}{\sigma_2} \right) \right\} \right\}, \quad (102b)$$

They can equivalently be written in terms of conductivities $\sigma_1$ and $\sigma_2$:

$$\frac{d\sigma_1}{dy} = -\frac{2}{\pi} \left\{ \frac{1}{\kappa_2} \left\{ 1 - \frac{\sigma_2}{\sigma_1} \ln \left( 1 + \frac{\sigma_1}{\sigma_2} \right) \right\} \right\}, \quad (103a)$$
$$\frac{d\sigma_2}{dy} = -\frac{1}{\pi}. \quad (103b)$$

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 $D_s = \sigma_s/4\kappa_s$ 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 $t = \sigma_1/\sigma_2$ 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 $\sigma_1$ is non-monotonous; 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 $\sigma_s$. We now discuss the topological effects at small values of conductivities. In the limit $\gamma_{11} = 0$, $\gamma_{22} = -1$ the
renormalization of \( \sigma_2 \) 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 NL\( \sigma \)M 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 \( \sigma_2 \) is topologically protected and flows towards \( \sigma_1^* \) of the order of the quantum of conductance (“interaction-induced criticality”). Before this happens, the flow of \( \sigma_1 \) 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 \( \sigma \to \sigma_1^* > 0 \). 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

\[
\frac{d\sigma}{dy} = -2 \left\{ 1 - \frac{1}{t} \ln [1 + t] - \frac{1}{t} \left[ 1 + \frac{1}{t} \right] \right\},
\]

\[
\frac{dt}{dy} = -2 \frac{1 + t}{\sigma} \left\{ -2 - \frac{1}{t} \ln [1 + t] + t^2 \ln \left[ 1 + \frac{1}{t} \right] \right\},
\]

or, equivalently,

\[
\frac{d\sigma_1}{dy} = -2 \left\{ 1 - \frac{\sigma_2}{\sigma_1} \ln \left[ 1 + \frac{\sigma_1}{\sigma_2} \right] \right\},
\]

\[
\frac{d\sigma_2}{dy} = -2 \left\{ 1 - \frac{\sigma_1}{\sigma_2} \ln \left[ 1 + \frac{\sigma_2}{\sigma_1} \right] \right\}.
\]

Even though the single-surface conductivities \( \sigma_s \) display non-monotonic 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 \( \gamma_{ss} \) close to \( -1 \) the corresponding conductivity \( \sigma_s \) is first subjected to localizing quantum corrections and will thus show a non-monotonic 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 non-monotonous 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:

\[ l \ll L_E, \quad (106a) \]

\[ d \ll l. \quad (106b) \]

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 \( s = 2 \):

\[ l_{\text{scr}} \ll L_E. \quad (106c) \]

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

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

\[ a \ll d. \quad (106d) \]

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, \( l_T = \min_{s=1,2} \sqrt{D_\perp/kT} \) the smaller thermal length, \( d \) the sample thickness, \( a \) the penetration depth, \( \kappa_s \) the inverse Thomas-Fermi screening length for the surface \( s \) and \( l_{\text{scr}} \) 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 \( \kappa^{-1}_s \). As discussed above (Sec. V), this condition implies that the bare values of interaction \( \gamma_{11} \) and \( \gamma_{22} \) 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

\[ \nu_s = \sqrt{\frac{n_s}{\pi v_F^2}}, \quad \kappa_s = \frac{2\pi e^2}{\epsilon_2} \nu_s = \frac{2\pi \alpha}{\pi} \sqrt{\frac{n_s}{\pi}} \quad (107) \]

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

\[ \nu_1^2 + \nu_2^2 = \frac{n_{\text{tot}}}{\pi v_F^2}, \quad \kappa_1^2 + \kappa_2^2 = \frac{2\pi \alpha}{\pi} n_{\text{tot}}. \quad (108) \]

The mean free path can be expressed as

\[ l = v_F \tau_T = \frac{\sigma}{\pi v_F (\nu_1 + \nu_2)}. \quad (109) \]

The thermal length in the diffusive regime is given by

\[ l_T = \sqrt{\frac{D}{kT}} = \sqrt{\frac{v_F l}{2kT}} = \sqrt{\frac{\sigma}{kT 2\pi (\nu_1 + \nu_2)}} \quad (110) \]

Hence, the condition (106a) is fulfilled for temperatures

\[ kT \ll kT_{\text{Diff}}, \quad (111) \]

where

\[ kT_{\text{Diff}} = \frac{v_F^2}{2l} = \frac{1}{\sigma [\epsilon^2/\hbar]} \left( \frac{v_F^2}{2} \right) 2\pi (\nu_1 + \nu_2) \quad (112) \]

is the temperature scale at which the diffusion sets in.

In order to obtain \( l_{\text{scr}} \) 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

\[ l_{\text{scr}} = \frac{\epsilon_1 + \epsilon_3}{2\epsilon_2} \frac{1}{\kappa_1 + \kappa_2}. \quad (113) \]

When deriving Eq. (113), we assumed for simplicity that \( \epsilon_2 \ll \epsilon_1 + \epsilon_3 \). Regarding the experimental setups discussed in Sec. VI B, this condition is well fulfilled for Bi\(_2\)Se\(_3\) 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

\[ \frac{v_{F,\perp} p_{\perp}}{\Delta_{\text{bulk}}} \sim 1, \quad (114) \]

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

\[ a \sim \frac{v_F}{\Delta_{\text{bulk}}}. \quad (115) \]

We are now going to consider two exemplary materials for 3D TIs: Bi\(_2\)Se\(_3\) 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\(_2\)Se\(_3\)

Bi\(_2\)Se\(_3\) 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. [17,97]) is summarized in the upper part of Table I. Using Eqs. (107) – (115) we can estimate the 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 \( l_{\text{scr}} \) [temperatures below \( T_{\text{max}} = 2.6 \ldots 1.9 \text{K} \), see condition (106c)].\(^{111}\)

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 \( d\sigma /dT \) is considerably smaller. Second, one observes a clear curvature of the dependence \( \sigma(T) \) which is a manifestation of the non-monotonicity. (For the parameters used in the plot the minimum of \( \sigma \) occurs at still lower temperatures.) This curvature is especially pronounced for strongly different surfaces.

It should be mentioned that the substrate used in Ref. 17 has a strongly temperature-dependent dielectric function \( \epsilon_3 \) 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 \( \epsilon_3 \) 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\(^{16}\). While the transport experiment indicates dominant surface conduction, the extracted carrier density appears to be too large for a pure

| TABLE I: Experimental values of sample parameters at the point of the minimal carrier density and associated length scales for transport experiments on Bi\(_2\)Se\(_3\) films of Refs. [17,97]. The dots “…” separate values for the symmetric \( (n_1 = n_2) \) and totally asymmetric \( (n_1 = n_{\text{tot}}, n_2 = 0) \) cases. The bare interaction amplitudes are estimated in the random phase approximation (RPA). |
|-----------------|-----------------|
| Fermi velocity | \( v_F \approx 5 \times 10^5 \text{m/s} \) |
| Bulk gap | \( \Delta_{\text{bulk}} \approx 0.3 \text{eV} \) |
| Sample thickness | \( d \approx 10 \text{nm} \) |
| Dielectric properties | Coat: \( \epsilon_1 \approx 1 \) \( \quad \) 3D TI (Bi\(_2\)Se\(_3\)): \( \epsilon_2 \approx 100 \) \( \quad \) Substrate (SrTiO\(_3\)): \( \epsilon_3 \approx 10^4 \) |
| Carrier density | \( n_{\text{tot}} \approx 3 \times 10^{12} \text{cm}^{-2} \) |
| Mobility | \( \mu_{\text{el}} \approx 100 \ldots 1000 \text{cm}^2/\text{V}\text{s} \) |
| Sheet resistance | \( 1/\sigma \approx 0.097 \, \text{h}/e^2 \) at \( T \approx 50 \text{mK} \) |
| Effective coupling | \( \alpha \approx 4 \times 10^{-2} \) |
| Chemical potential | \( \rho_1^2 + \rho_2^2 = (0.2 \text{eV})^2 \) |
| Penetration depth | \( a \approx 1 \text{nm} \) |
| Mean free path | \( l \approx 24 \ldots 34 \text{nm} \) |
| Diff. temperature | \( T_{\text{Diff}} \approx 80 \ldots 57 \text{K} \) |
| Screening length | \( \kappa_1^2 + \kappa_2^2 \approx (37 \text{nm})^{-2} \) |
| Scr. length (total) | \( l_{\text{scr}} \approx 132 \ldots 186 \text{nm}, \text{ for } \epsilon_3 = 10^4 \) |
| Bare interaction (RPA) | top surface: \( \gamma_{11} \approx -0.6 \ldots -1 \) \( \quad \) bottom surface: \( \gamma_{22} \approx -0.6 \ldots 0 \) |

| TABLE II: Typical experimental values for transport experiments on HgTe films of Refs. [16,98]. |
|-----------------|-----------------|
| Fermi velocity | \( v_F \approx 5 \times 10^5 \text{m/s} \) |
| Bulk gap | \( \Delta_{\text{bulk}} \approx 0.622 \text{eV} \) |
| Sample thickness | \( d \approx 70 \text{nm} \) |
| Dielectric properties | Coat: \( \epsilon_1 \approx 1 \) \( \quad \) 3D TI (HgTe): \( \epsilon_2 \approx 21 \) \( \quad \) Substrate (CdTe): \( \epsilon_3 \approx 10 \) |
| Carrier density | top surface: \( n \approx 4.8 \times 10^{11} \text{cm}^{-2} \) \( \quad \) bottom surface: \( n \approx 3.7 \times 10^{11} \text{cm}^{-2} \) |
| Mobility | \( \mu_{\text{el}} \approx 34000 \text{cm}^2/\text{V}\text{s} \) |
| Sheet resistance | \( 1/\sigma \approx 0.014 \, \text{h}/e^2 \) at \( T \approx 50 \text{mK} \) |
| Effective coupling | \( \alpha \approx 0.21 \) |
| Chemical potential | top surface: \( \mu_1 \approx 0.08 \text{eV} \) \( \quad \) bottom surface: \( \mu_2 \approx 0.07 \text{eV} \) |
| Penetration depth | \( a \approx 15 \text{nm} \) |
| Mean free path | \( l \approx 108 \text{nm} \) |
| Diff. temperature | \( T_{\text{Diff}} \approx 18 \text{K} \) |
| Screening length | top surface: \( \kappa_{11} \approx 19.53 \text{nm} \) \( \quad \) bottom surface: \( \kappa_{22} \approx 22.24 \text{nm} \) |
| Bare interaction (RPA) | top surface: \( \gamma_{11} \approx -0.893 \) \( \quad \) bottom surface: \( \gamma_{22} \approx -0.878 \) |
surface theory with linear spectrum, yielding the value of the chemical potential $\mu$ larger than the gap $\Delta_{\text{bulk}}$, 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 magneto-optical study by the same experimental group.) 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.

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 interaction.

- As already exploited in 3D TI experiments, the magnetoconductance formula for the total conductivity is

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

where the characteristic magnetic field $B^{(s)}_\phi = h/(4eD^{(s)}_s \tau^{(s)}_\phi)$ is determined by the diffusion coefficient $D^{(s)}_s$ and the phase breaking time $\tau^{(s)}_\phi$ for the surface $s$. The function $\psi$ denotes the digamma function here.

- The characteristic effect of intersurface interaction is the non-monotonous 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

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

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

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

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

- The strength of intersurface interaction is governed by the parameters $\kappa_1 d$ and $\kappa_2 d$, where $\kappa$ 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 non-linear sigma 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 $Z_2$ 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 $SU(2)$-gauged sigma 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 sigma-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
The total electron concentration in units of $10^{12}\text{cm}^{-2}$ is equal to 0.55, 0.48, 0.41, 0.34, 0.27 from bottom to top. The characteristic non-monotonous behavior is clearly seen; deviations from the behavior of decoupled surfaces are very strong. Bottom: Carrier-density dependence of conductivity corrections. The non-trivial 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 $\alpha$ as in Table I for Bi$_2$Se$_3$. Further, we assumed the case of equal surfaces ($n_{tot} = 2n$) and $T_{\text{Diff}} = 1/2\tau = 50\text{K}$.

We renormalized the interacting NL$\sigma$M 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, $k_Fl \gg 1$. 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 vanish-

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig19}
\caption{Top: Conductivity corrections for low carrier concentration. The total electron concentration in units of $10^{12}\text{cm}^{-2}$ is equal to 0.55, 0.48, 0.41, 0.34, 0.27 from bottom to top. The characteristic non-monotonous behavior is clearly seen; deviations from the behavior of decoupled surfaces are very strong. Bottom: Carrier-density dependence of conductivity corrections. The non-trivial 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 $\alpha$ as in Table I for Bi$_2$Se$_3$. Further, we assumed the case of equal surfaces ($n_{tot} = 2n$) and $T_{\text{Diff}} = 1/2\tau = 50\text{K}$.}
\end{figure}

Finally, we have estimated parameters and presented explicit predictions for the temperature dependence of the conductivity for typical experimental setups based on Bi$_2$Se$_3$ 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 $\sigma_{12}$. 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, 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).

**VIII. ACKNOWLEDGEMENTS**

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 grants No. 11-02-12126 and No. 12-02-00579, Russian President grant No. MK-4337.2013.02, 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 $S^{(0)}$. For brevity, we omit the surface index $s$ in this appendix.

1. Gauged WZNW model

The Wiegmann-Polyakov formula\(^7\) allows the inclusion of smooth $\Phi(r = 2\times 2N_M \times N_R)$ gauge fields $A_\mu$. A generalization to potentially topological gauge potentials can be found in Refs. 72–75. The gauged WZNW model is given as\(^8\)

\[ S[O,A_\mu] = \int \frac{i}{16\pi} \left( O^T D_\mu O \right) \left( (O^T D_\mu O) \right) \tag{A1a} \]

\[ + \frac{i\epsilon_{\mu\nu}}{24\pi} \int \left( \left[ (O^T D_\mu O) \left( (O^T D_\nu O) \right) \right] + A_\mu \left( O\partial_\nu O + \partial_\nu OO^T \right) \right) \tag{A1b} \]

\[- \frac{i\epsilon_{\mu\nu}}{16\pi} \int \left( \left[ F_{\mu\nu} \left( O^T D_\rho O + D_\rho OO^T \right) \right] \right) \tag{A1c} \]

where $D_\mu = \partial_\mu + [A_\mu,\cdot]$ and $F_{\mu\nu} = [D_\mu, D_\nu]$. In the main text, we were mostly interested in the $U(1)$ gauge fields $A_\mu = i\tilde{A}_\mu = i\tilde{A}_\mu^{1-\tau_2} - i\tilde{A}_\mu^{1-\tau_2}$. (In this appendix, the electron charge is absorbed into the vector potential.)

To obtain the Wiegmann-Polyakov formula, one can use the following identity\(^75\)

\[ (A1b) = \frac{i}{24\pi} \Gamma[O] \]

\[ - \frac{i}{8\pi} \int \epsilon_{\mu\nu\rho\sigma} \partial_\mu \text{tr} \left[ OA_\rho O^T A_\sigma \right] + A_\mu \left( O^T \partial_\nu O + \partial_\nu OO^T \right) \]

\[ + \frac{i\epsilon_{\mu\nu}}{16\pi} \int \left( \left[ F_{\mu\nu} \left( O^T D_\rho O + D_\rho OO^T \right) \right] \right) \tag{A2} \]

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

\[ S[O,A_\mu] = S[O] + \]

\[ + \frac{1}{8\pi} \int [A_\mu \left( O\partial_\nu O + O^T \partial_\nu O \right) + A\mu \left( O\partial_\nu O + O^T \partial_\nu O \right) - i\epsilon_{\mu\nu\rho\sigma} A_\rho \left( O^T \partial_\nu O + \partial_\nu OO^T \right)] \]

\[ = S[O] + \]

\[ + \frac{1}{8\pi} \int \left[ A_+ \left( O\partial_\nu O + O^T \partial_\nu O \right) + A_+ \left( O^T \partial_\nu O + \partial_\nu OO^T \right) \right] \tag{A3} \]

Here we have introduced the (anti-)holomorphic combination of gauge potentials $A_\pm = A_x \pm iA_y$. 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 $(O \rightarrow Q = Q^T)$ 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 $O_{soft}^T O_{soft}$ of the WZNW fields $O$ are not affected by disorder induced masses, Eq. (30). The effective action for topologically trivial Goldstone modes contains

\[ S_{eff,kin} [\Phi_\mu] = \frac{1}{8\pi} \int \left( \text{tr} \left[ \Phi_\mu O^T \Phi_\mu O - \Phi_\mu \Phi_\mu \right] \right) \]

\[ - \frac{1}{2} \left( \int \left( \text{tr} \left[ \Phi_\mu \Phi_\mu \right] \right) \right)^2 \tag{A5} \]

where $j_\pm$ are the (bosonic) currents, $\langle \ldots \rangle$ denotes average with respect to the full bosonic theory (including the mass terms) and $\Phi_\pm = O_{soft} \Phi_\pm O_{soft}^T$. 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 $\langle \ldots \rangle$ 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

\[ a \equiv \left( a_0, a_1, a_2, a_3 \right) \]

\[ \equiv \frac{1}{\sqrt{(2\lambda(\tilde{x}^2 - \tilde{x}'^2), |\tilde{x} - \tilde{x}'|^2 - \lambda^2)}} \]

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

\[ \hat{O}_{\text{inst}} = \left( \begin{array}{c} a_0 i \tau_y + a_3 0 a_1 + a_2 i \tau_y 0 \\ 0 1 0 0 \\ a_1 - a_2 i \tau_y 0 a_0 i \tau_y - a_3 0 \\ 0 0 0 -1 \end{array} \right) \tag{A6} \]

For $a_0 = 0$, i.e. on the physical space $w = 1$, $\hat{O}_{\text{inst}}$ 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 $\tilde{O}_{\text{inst}}$ 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 $\tilde{O}_{\text{inst}}$ satisfies the boundary condition $\tilde{O}(x, w = 0) = \Lambda = \text{const.}$.

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

$$i S^{(d)} = \frac{-i}{6\pi} \int_{x, w} \epsilon_{\mu \nu \lambda} \left( \epsilon_{abc} a_{\mu} \partial_{\nu} a_{b} \partial_{\lambda} a_{c} - \epsilon_{\mu \nu \lambda} \epsilon_{a b c} \partial_{\mu} a_{b} \partial_{\nu} a_{c} \partial_{\lambda} a_{0} - \epsilon_{\nu \lambda \rho} \epsilon_{a b c} \partial_{\nu} a_{b} \partial_{\lambda} a_{c} \partial_{\rho} a_{0} + \epsilon_{\mu \lambda \rho} \epsilon_{a b c} \partial_{\mu} a_{b} \partial_{\rho} a_{c} \partial_{\nu} a_{0} \right)$$

$$= i \pi. \quad (A7)$$

Here the last line is obtained by a straightforward calculation. We have thus shown that the topological term distinguishes between the trivial and the non-trivial sectors as it acquires on them the values 0 and $i \pi \mod 2\pi i$, 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 figure 1). We define the $z$-axis to be perpendicular to the two surfaces. The sandwich consists of the coating material with a dielectric constant $\epsilon_1$ (for $d/2 < z$), the topological insulator film with a dielectric constant $\epsilon_2$ (for $-d/2 \leq z \leq d/2$), and the substrate with a dielectric constant $\epsilon_3$ (for $z < -d/2$). Taking these different dielectric properties into account, we here present the expression for the Coulomb potential which generalizes Eq. (4).

By the method of mirror charges, one can derive\textsuperscript{41-43} the electrostatic potential induced by a single point charge $e$ located at $(x_0, z_0) = (0, 0, z_0)$ inside the middle region of the sandwich ($z, z_0 \in [-\frac{d}{2}, \frac{d}{2}]$):

$$\Phi(x, z, z_0) = \frac{e}{2\epsilon_2} \left[ \frac{1}{\sqrt{x^2 + (z - z_0)^2}} + r_2^{-1} F(x, d + (z + z_0)) + r_2^{-1} F(x, d - (z + z_0)) + F(x, z - z_0) + F(x, -(z - z_0)) \right], \quad (B1)$$

where

$$\mathcal{F}(x, z) = \sum_{k=1}^{\infty} \frac{(r_2 r_3)^k}{\sqrt{x^2 + (z - 2dk)^2}}, \quad (B2)$$

and the ratios

$$r_{21} = \frac{\epsilon_2 - \epsilon_1}{\epsilon_2 + \epsilon_1}; \quad r_{23} = \frac{\epsilon_2 - \epsilon_3}{\epsilon_2 + \epsilon_3}$$

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

$$\Phi(q, z, z_0) = \frac{2\pi e}{q \epsilon_2} \left[ e^{-|z - z_0|q} + e^{-2dq} \frac{1 - r_{21} r_{23} e^{-2dq}}{r_{21} e^{(d + z + z_0)q} + r_{23} e^{(d - z - z_0)q} + 2r_{21} r_{23} \cosh((z - z_0)q)} \right]. \quad (B3)$$

We consider now 3D TI surface states: the charges are located at a typical distance $a \approx v_F/\Delta_{\text{bulk}}$ (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

$$\Delta \mu_1 = \frac{e}{2} \Phi_{\text{reg}} \left( 0, \frac{d}{2} - a, \frac{d}{2} - a \right)$$

$$= \frac{e^2}{4 \epsilon_2} \left[ \frac{r_{21}}{a} - \frac{r_{21} + r_{21}^{-1} + 2 \ln(1 - r_{21} r_{23})}{d} \right]. \quad (B4)$$

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 $\text{reg}$ indicates that selfinteraction 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 $\Delta \mu_1 \approx \alpha_2 r_{21}/4 \Delta_{\text{bulk}}$.

Second, the electrostatic energy associated with two-particle interaction is the quantity $U_0$ entering $S_{\text{int}}$ 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 $\exp(-q d)$ ($q$ 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 \(dL^{-1} \gg 1\) or for part of it if \(d \in [l, L]\). Then the two surfaces become decoupled and
\[
U'_0 = \frac{2\pi}{q} \left( \begin{array}{cc} 2 & 0 \\ \epsilon_2 + \epsilon_1 & \epsilon_2 + \epsilon_3 \end{array} \right).
\] (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 which 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
\[
U_{LC} = \frac{2}{\epsilon_1 + \epsilon_3} \frac{2\pi}{q} \left( \begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array} \right).
\] (B6)

As expected, it does no longer depend on \(\epsilon_2\). 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 which do not vanish in the \(qd \to 0\) limit. These are short range interaction amplitudes introduced by the finite thickness of the film:
\[
F^{(d)} = -\frac{2\pi}{\epsilon_2} d \left( \begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right) - \frac{4\pi}{\epsilon_1 + \epsilon_3} d \left[ F_{\text{symm}} \left( \begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array} \right) + F_{\text{M}} \right].
\] (B7)

Here we have defined the scalar
\[
F_{\text{symm}} = (\epsilon_2 - \epsilon_1)(\epsilon_2 - \epsilon_3) \left\{ \frac{1}{2\epsilon_2} + \frac{1}{\epsilon_2(\epsilon_1 + \epsilon_3)} \right\}
\] (B8)
and the matrix
\[
F_{\text{M}} = \frac{1}{2\epsilon_2} \left( \begin{array}{cc} \epsilon_2 - \epsilon_1 & \epsilon_2 - \epsilon_3 \\ \epsilon_2 - \epsilon_1 & \epsilon_2 - \epsilon_3 \end{array} \right)
\] (B9)
which both vanish in the limit of \(\epsilon_1 = \epsilon_2 = \epsilon_3\). In summary, for coupled surface we can write \(U'_0 = U_{LC} + F^{(d)}\).

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 (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 \((z_0 = \pm d/2 + a \approx \pm d/2\) 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 \(\text{Bi}_2\text{Se}_3\) it is known that \(a\) is of the order of a few nanometers, 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 references 66–68. We first consider the short range (one-Coulomb-line-irreducible) part of the singlet channel (see also Eq. (43))
\[
\Gamma^{1-2}_{ss'} = \Gamma^1_{ss'} - \Gamma^2_{ss'} \delta_{ss'},
\] (C1)
and include the long-range, one-Coulomb-line-reducible, diagrams \((\Gamma^0)\) 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
\[
\Gamma^{1-2}(K) = I^{1-2} - I^{1-2} R(K) \Gamma^{1-2}(K)
\] (C2)
(Matrix multiplication includes momentum integral \(\int_p\) and a Matsubara sum \(T \sum_n\).)

The matrix
\[
[R(K)]_{pp',ss'} = \delta_{ss'} \delta_{pp'} R_{s,p}(K),
\] (C3)
\[
R_{s,p}(K) \equiv G_s(P) G_s(P + K)
\] (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 \(R_{s,p}(K)\) can be represented as
\[
R_{s,p}(K) = R^{\omega}_{s,p} + \Delta_{s,p}(K)
\] (C5)
\[
= R^{\omega}_{s,p} + \tilde{\Delta}_{s,p}(K).
\] (C6)
Here \(R^{\omega}_{s,p}\) (\(R^0_{s,p}\)) are called regular (static) part of the bubble. The \(\omega\)- and \(q\)-limits are defined in the main text (see Eqs. (48) and (49)). The singular (dynamic) part of the particle-hole bubble is
\[
\Delta_{s,p}(K) = \beta - \frac{-iv^F_{s,p}}{\omega_m + iv^F_{s,p}} \delta^{(s)},
\] \[
\tilde{\Delta}_{s,p}(K) = \beta - \frac{-iv^F_{s,p}}{\omega_m + iv^F_{s,p}} \delta^{(s)}.
\]
(We have absorbed the Fermi liquid (FL) residues into a redefinition of the scattering amplitudes.) From these definitions and Eq. (C2) we obtain the relations
\[
\Gamma^{1-2}(K) = \Gamma^{1-2,\omega} - \Gamma^{1-2}(K) \Delta(K) \Gamma^{1-2,\omega},
\] (C7a)
We used the approximation $\langle \phi^{(s)}(\omega_m, q) \rangle \approx \Gamma^{1-2}_0(\omega_m, q) - \Gamma^{1-2}_0(K) \tilde{A}(K) \Gamma^{1-2}_0(K)$. (C7b)

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

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 $\phi^{(s)}(\omega_m, q)$:

$$v^{(1)}_0 = (1, 0)$$ and $v^{(2)}_0 = (0, 1).$ (C8)

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

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

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

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

$$\Pi^{ss'}(K) = -v^{(s)}_0 R(K) \Gamma^{1-2}(K) R(K) [v^{(s')}_0]^T,$$ (C10)

which transforms into

$$\Pi^{ss'}(K) = \Pi^{ss', q} + T^{(s), q} \tilde{A}(K) [T^{(s'), q}]^T - T^{(s), q} \tilde{A}(K) \Gamma^{1-2}(K) R(K) [v^{(s')}_0]^T.$$ (C11a)

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

$$\left( \frac{\partial G^{-1}_{1, s}}{\partial \mu_s}, 0 \right) = T^{(1), \omega} \text{ and } \left( 0, \frac{\partial G^{-1}_2}{\partial \mu_s} \right) = T^{(2), \omega}. $$ (C12)

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

$$\left( \frac{\partial G^{-1}_{1, s}}{\partial \mu_s}, \frac{\partial G^{-1}_2}{\partial \mu_s} \right) = T^{(s), q}. $$ (C13)

We insert this into the $\omega$- and $q$- limits of the polarization operator and obtain

$$\Pi^{ss', \omega} = 0 \text{ and } \Pi^{ss', q} = -\frac{\partial N_{\omega}}{\partial \mu_s} = -\frac{\partial N_{\omega}}{\partial \mu_s}. $$ (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. 24

4. Screening of the Coulomb interaction

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

$$U_0 = \frac{2\pi}{\epsilon_{eff}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$ (C15)

where $\epsilon_{eff} = (\epsilon_1 + \epsilon_3)/2$ for the most general situation of a dielectric sandwich structure. This matrix has zero determinant, det $U_0 = 0$.

The RPA-screened Coulomb interaction is defined as

$$U_{scr}(\omega_m, q) = (1 - U_0 \Pi)^{-1} U_0.$$ (C16)

By means of the orthogonal matrix

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

we can rotate $U_{scr}(\omega_m = 0, q)$ into the basis where $U_0$ is diagonal:

$$O^T U_{scr}(\omega_m = 0, q) O = \begin{pmatrix} 1 - \frac{4\pi}{\epsilon_{eff}} & 0 \\ 0 & 0 \end{pmatrix} O^T \Pi^O O^{-1} \times \begin{pmatrix} \frac{4\pi}{\epsilon_{eff}} & 0 \\ 0 & 0 \end{pmatrix} = q - \frac{2\pi}{\epsilon_{eff}} (\Pi^{21}_{11} + \Pi^{22}_{22} + 2\Pi^{23}_{12}) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. $$ (C18)

The denominator in the last line of Eq. (C18) defines the coupled surface screening length [analogously to Eqs. (7), (8)].
In the considered parameter range we can take the $q$-limit under the following condition: 

$$\left| \frac{2\pi}{\epsilon_{\text{eff}}} (\Pi_{11}^q + \Pi_{22}^q + 2\Pi_{12}^q) \right| \gg q.$$ 

Then we obtain

$$U_{\text{cr}}^q = -O \left[ \left[ \frac{\epsilon_1^T}{\epsilon_1^T} O^T \Pi^q O \epsilon_1 \right]^{-1} 0 0 \right] O^T. \quad (C19)$$

The $q$-limit of Eq. (C16) is

$$\Gamma_0^q = \left( \frac{1}{\nu} \right)^2 \Pi^q O \epsilon_1 \otimes \epsilon_1^T O^T \Pi^q \left( \frac{1}{\nu} \right)^2 \epsilon_1^T O^T \Pi^q O \epsilon_1. \quad (C20)$$

We multiply by $\nu O \epsilon_1$ from the right side and find

$$\left[ -\left( \frac{1}{\nu} \right)^2 \Pi^q + \Gamma_0^q + \nu \right] O \epsilon_1 = 0. \quad (C21)$$

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

Alternatively, using the $q$-limit of Eq. (C11b) and (C16), we can express the bare total interaction amplitude $\Gamma^q = \Gamma^0 + \Gamma^1$ as

$$\nu \Gamma^q \nu = -\nu - \frac{\det \Pi^q \Pi_{11}^q + \Pi_{22}^q + 2\Pi_{12}^q}{1 - 1 1} \nu \Gamma^q \nu. \quad (C22)$$

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

$$\det \left[ \nu + \nu \Gamma^q \nu \right] = 0 \quad (C23)$$

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 $\Pi^{\text{RPA}}$. It is defined as

$$\Pi^{\text{RPA}} (K) = \Pi (K) + \Pi (K) U_{\text{cr}}^q (q) \Pi^{\text{RPA}} (K). \quad (C24)$$

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

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

$$\frac{\omega_m}{\omega_m + \nu \nu^F \nu} \rightarrow \frac{\omega_m}{Z_s \omega_m + D_s q^2}, \quad (C25)$$

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

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

where

$$\Delta^{-1} = \Delta^{-1} (\omega_m, q) = \left[ \nu D q^2 + (\nu Z + \Pi^{\nu+1-2} Z_{\nu}) \omega_m \right]^{-1}. \quad (C27)$$

These equations are used in the main text (Sec. IV B) to provide a link between the bosonized NL$\sigma$M and the dirty FL theory.

6. Bare NL$\sigma$M coupling constants

According to Eqs. (C18) and (C22) the bare values of the interaction amplitudes are fully determined by $\nu_1$, $\nu_2$ and

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

where

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

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

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

In appendix B we derived the general expression for FL constants $F^{(d)} = \frac{\nu_0^d}{2 \epsilon_2^d} f_d$ 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 NL$\sigma$M. 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

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

The bare value of the interaction constant is

$$\gamma_{11} = \gamma_{22} = \frac{\nu \nu^0 \nu_{11}}{\nu_1} = -\frac{1}{2} \left( 1 + \frac{\kappa d}{1 + \kappa d} \right). \quad (C32)$$

Note that in the limit $\kappa d \rightarrow \infty \ (\kappa d \rightarrow 0)$ 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

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

It follows from Eqs. (C30) and (C33) that the bare values for interaction constants are $\epsilon_3$-independent:

$$\gamma_{11} = -1 + \frac{1}{1 + \frac{\kappa_1}{\epsilon_2} + 2\kappa_1 d}. \quad (C34a)$$
and

\[ \gamma_{22} = -1 + \frac{1}{1 + \frac{1}{\kappa_2} + 2k_2 q^2}. \]  
(C34b)

In view of Eq. (65) following from the 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 \( \delta Q = \begin{pmatrix} 0 & q \\ q^T & 0 \end{pmatrix} \). Then we directly single out the classical contribution in \( B_1 \), Eq. (70) and obtain

\[ B_1^s = \sigma_s - \frac{\sigma_s}{4n} \sum_{\mu=0,2} \text{tr} \left[ I_n^\alpha \tilde{\tau}_\mu \delta Q \left[ I_n^{\alpha'} \tilde{\tau}^T_\mu \delta Q - I_n^{\alpha'} \tilde{\tau}^T_\mu \delta Q \right] \right]. \]  
(D1)

In addition, we write \( q = \sum_{\nu=0}^3 q^{(\nu)} \tilde{\tau}^T_\nu \). When performing the trace in \( \tau \)-space it turns out that the two diffusion contributions (\( \nu = 0, 2 \)) cancel up. This is a consequence of the opposite sign of \( \tau_0 \) and \( \tau_2 \) under transposition. The cooperons (\( \nu = 1, 3 \)) contribute only to the last term in Eq. (D1). Then we find

\[ B_1^s = \sigma_s + \frac{\sigma_s}{4n} \sum_{\nu=1,3} \left( \text{tr} I_n^\alpha \left( \begin{array}{cc} 0 & q^{(\nu)} \\ q^{(\nu)} & 0 \end{array} \right) I_n^\alpha \left( \begin{array}{cc} 0 & q^{(\nu)} \\ q^{(\nu)} & 0 \end{array} \right) \right) \]

\[ = \sigma_s + 2 \int \sum_{n=0}^{N_M} D_s(\omega_n, p). \]  
(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:

\[ B_2^{ss'} \bigg|_{\text{tree level}} = -\frac{\sigma_s \delta_{ss'}}{4} \int e^{ip(z-x')} p^2 D_{ss'}(p, \omega) = 0. \]  
(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 \( B_2 \) (\( \star \) and \( \delta \) denote Wick contractions):

\[ B_2^{ss'} = -\frac{\sigma_s \sigma_{s'}}{8n} \int x-x' \sum_{\mu=x,y} \left[ \text{tr} \left( I_n^{\alpha} \tilde{\tau}_\mu q^{(0)}_0 \right)_{s,x} \text{tr} \left( I_n^{\alpha} \tilde{\tau}_\mu q^{(0)}_0 \right)_{s',x'} + \text{tr} \left( I_n^{\alpha} \tilde{\tau}_\mu q^{(0)}_0 \right)_{s,z} \text{tr} \left( I_n^{\alpha} \tilde{\tau}_\mu q^{(0)}_0 \right)_{s',z'} + 2 \text{tr} \left( I_n^{\alpha} \tilde{\tau}_\mu q^{(0)}_0 \right)_{s,x} \text{tr} \left( I_n^{\alpha} \tilde{\tau}_\mu q^{(0)}_0 \right)_{s',x'} \right]. \]  
(D4)

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

\[ B_2^{ss'} = \frac{32\pi T \delta_{ss'}}{\sigma_s n} \int p^2 \sum_{n=0}^{N_M} \frac{n_{12}}{n_{12}} \left[ (D \bar{G} D^c)_{ss'} (\omega_{n+2}; p) D_s (\omega_{n+2}, p) - (D \bar{G} D^c)_{ss'} (\omega_{n+2}, p) D_s (\omega_{n+2}; p) \right]. \]  
(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 \( \sigma_{s'} \). Since we are interested in the zero temperature limit, Eq. (D5) becomes

\[ B_2^{ss'} = \frac{16 \delta_{ss'}}{\sigma_s} \int p^2 \int_0^\infty d\omega \left( D \bar{G} D^c \right)_{ss'} (\omega, p) D_s (\omega, p). \]  
(D6)

We use the relation

\[ (D \bar{G} D^c)_{ss'} (\omega, p) D_s (\omega, p) = \Gamma_{ss} D^2_s (\omega, p) \bar{D}_s (\omega, p) \frac{4\omega \Gamma_{12} D_s (\omega, p) \bar{D}_s (\omega, p)}{\sigma_{(-s)} \det \left( (D^c (\omega, p))^{-1} \right)} \]  
(D7)

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

\[ \bar{D}_s (\omega, p) = \left[ p^2 + L^{-2} + 4 (\zeta_s + \Gamma_{ss}) \omega / \sigma_s \right]^{-1}. \]  
(D8)

The single surface induced correction is given as

\[ B_2^{ss'} \bigg|_{\text{single}} = \frac{16 \delta_{ss'}}{\sigma_s} \int p^2 \int_0^\infty d\omega \Gamma_{ss} D^2_s (\omega, p) \bar{D}_s (\omega, p) \]

\[ = -4 \delta_{ss'} f \left( \zeta_s / \zeta_s \right) \int p^2 D^2_s (0, p). \]  
(D9)
Here we introduced the function

\[ f(x) = 1 - (1 + 1/x) \ln(1+x). \]  

(D10)

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

\[
B_{2s's'}^\text{inter} \bigg|_{\text{inter}} = \frac{64\delta_{ss'}^2\Gamma_1^2}{\det(z + \Gamma)} \int_{p} p^2 D_s(0, p) \int_{0}^{\infty} d\omega \omega D_s(\omega, p) \times \frac{D_s(\omega, p)}{d_+ - d_-} \sum_{\zeta=\pm} \zeta \left[ 2\sigma^2_1\Gamma_1^2 \delta_s \right] \left[ \sum_{\varsigma=\pm} \varsigma \right] \times f_2 \left( \frac{\sigma_s}{z_s}, \frac{\sigma_s}{z_s + \Gamma}, d_+ \right) \int_{p} p^2 D_s^2(0, p),
\]

where

\[
d_\pm = \left( z_1 \sigma_2 + \sigma_1 z_2 \right) \frac{2}{2 \det(z + \Gamma)} \left[ 1 + \frac{1}{\left( z_2 + \Gamma \right)^2} \left( z_1 \sigma_2 + \sigma_1 z_2 \right)^2 \right] \]

and

\[
f_2(a, b, c) = 2 \frac{(c - b) \ln a + (a - c) \ln b + (b - a)c \ln c}{(b - a)(c - a)(c - b)}. \]

(D13)

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

\[
B_{211}^\text{inter} = -4 \left[ 1 + \frac{\Gamma_{11}}{z_{11}} \right] \ln \left( \frac{1 + \frac{\Gamma_{11}}{z_{11}}}{\Gamma_{11}} \right) \left( \frac{1 + \sigma_1 z_2}{\sigma_2 z_1} \right) \frac{\sigma_1 (z_2 + \Gamma z_2)}{\sigma_2 z_1} \int_{p} p^2 D_s^2(0, p).
\]

(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 \( D_s D_s \) and, in analogy with the corresponding diffusion 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. VB2) 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.

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

\[
\frac{d}{dy} \left( \begin{array}{c} \delta t \\ \delta \gamma \end{array} \right) = M(\gamma) \left( \begin{array}{c} \delta t \\ \delta \gamma \end{array} \right),
\]

(E1)

with the \( \gamma \)-dependent matrix

\[
M(\gamma) = -\frac{2}{\pi\sigma} \left( \begin{array}{cc} \frac{(3+4\gamma)G(\gamma)}{(1+2\gamma)^2} & \frac{2G(\gamma)}{1+2\gamma} \\ \frac{(\gamma+\gamma^2)}{(1+2\gamma)^2} & -\gamma + \gamma^2 \end{array} \right),
\]

(E2)

and \( G(\gamma) = -1 - 2\gamma + (2 + 2\gamma) \ln(2 + 2\gamma) \). The eigenvalues of the matrix \( M(\gamma) \) 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 \( \gamma \)). 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 \( \gamma < \gamma_r \approx -0.64 \) one should keep in mind that \( \gamma \) 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 \( \kappa_0^{-1} \) 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 \( F \)-invariance. As a consequence, the relations for NL\( \sigma \)M parameters \( \det(z + \Gamma) = 0 \) and
In contrast to the Coulomb case (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 $\Gamma_{ss}$ and $z_s$ are exactly the same as in the Coulomb case. In particular, $\Gamma_{12}$ 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 $0 \leq |\Gamma_{ss}| \leq z_s$ and typically $|\Gamma_{12}| \leq \max_{s=1,2} |\Gamma_{ss}|$.

For sufficiently strong interactions, the RG flow implies localizing behaviour 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 $\Gamma_{12}/z_s$ decreases.) Accordingly, similar to the case of Coulomb interaction, in the case of strong short range interactions we also predict a non-monotonic conductivity behaviour. The quantitative difference is the steeper antilocalizing slope in the final stage of the flow.

\begin{align}
\frac{d\sigma_1}{dy} &= \frac{2}{\pi} \left[ \frac{1}{2} - f \left( \frac{\Gamma_{11}}{z_1} \right) \right. \\
&\quad \left. - \frac{\bar{\sigma}_1^2 \Gamma_{12}^2}{2z_1 (z_1 + \Gamma_{11})} \sum_{\varsigma = \pm} f_2 \left( \frac{\sigma_{\varsigma}}{z_1 + \Gamma_{11}}, \frac{\bar{\sigma}_1}{z_1 + \Gamma_{11}} \right) \right] \tag{F1a}
\end{align}

\begin{align}
\frac{d\sigma_2}{dy} &= \frac{2}{\pi} \left[ \frac{1}{2} - f \left( \frac{\Gamma_{22}}{z_2} \right) \right. \\
&\quad \left. - \frac{\bar{\sigma}_2^2 \Gamma_{12}^2}{2z_2 (z_2 + \Gamma_{22})} \sum_{\varsigma = \pm} f_2 \left( \frac{\sigma_{\varsigma}}{z_2 + \Gamma_{22}}, \frac{\bar{\sigma}_2}{z_2 + \Gamma_{22}} \right) \right] \tag{F1b}
\end{align}

\begin{align}
\frac{dz_1}{dy} &= -\frac{d\Gamma_{11}}{dy} = \frac{\Gamma_{11}}{\pi z_1}, \tag{F1c}
\end{align}

\begin{align}
\frac{dz_2}{dy} &= -\frac{d\Gamma_{22}}{dy} = \frac{\Gamma_{22}}{\pi z_2}. \tag{F1d}
\end{align}
H. Buhmann, and L. W. Molenkamp, Phys. Rev. Lett. 107, 136803 (2011).

Z. Ringel, Y. E. Kraus, and A. Stern, Phys. Rev. B 86, 045102 (2012).

P. M. Solomon, P. J. Price, D. J. Frank, and D. C. La Tulipe, Phys. Rev. Lett. 63, 2508 (1989).

T. J. Gramila, J. P. Eisenstein, A. H. MacDonald, L. N. Pfeiffer, and K. W. West, Phys. Rev. Lett. 66, 1216 (1991).

M. Schütt, P. M. Ostrovsky, M. Titov, I. V. Gornyi, B. N. Narozhny, and A. D. Mirlin, Phys. Rev. Lett. 110, 026601 (2013).

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).

W. Zhang, R. Yu, H.-J. Zhang, X. Dai, and Z. Fang, New J. Phys. 12, 065013 (2010).

Closer than the typical length scale $L_E$ of the system, see Eq. (3).

Instead of considering the renormalization of $z_s$ one can equivalently consider the renormalization of $\Gamma_{ss}$. It is governed by the interaction term $S^{\text{int}}$ in Eq. (57). Within the background field method two types of contributions can arise. First, there is $(S^{\text{int}})^{\text{fast}}$. This term does not involve a frequency integration. Because disorder is uncorrelated between the surfaces, $\Gamma_{11}$ and $\Gamma_{22}$ are renormalized separately. This is described by Eq. (93). All possible further contributions at this order would arise from $\left\langle (S^{\text{int}})^2 \right\rangle^{\text{fast}}$. This term generates so-called ring diagrams. We have explicitly checked that the ring diagrams vanish in one-loop approximation.

The $Z_2$ 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 (see Sec. III E 2) and hence do not fall into our definition of non-topological symplectic metals.

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. A priori $\sigma_1^*$ and $\sigma_2^*$ are different although we cannot exclude a possibility that they might be equal.

The only assumption that cannot 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. We see however no reason for such perfect correlations between impurities at opposite surfaces of a 3D TI film.