Two-dimensional transport and screening in topological insulator surface states

S. Adam,[1] E. H. Hwang,[2] and S. Das Sarma[2]

[1]*Center for Nanoscale Science and Technology, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA*
[2]*Condensed Matter Theory Center, Department of Physics, University of Maryland, College Park, Maryland 20742-4111, USA*

We study disorder effects on the surface states of the topological insulator Bi_2Se_3 close to the topologically protected crossing point. Close to charge neutrality, local fluctuations in carrier density arising from the random charged disorder in the environment result in electron and hole puddles that dominate the electronic properties of these materials. By calculating the polarizability of the surface state using the random-phase approximation, and determining the characteristics of puddles using the self-consistent approximation, we find that band asymmetry plays a crucial role in determining experimentally measured quantities, including the conductivity and the puddle autocorrelation length.

Topological insulators (TIs) are a new class of materials[1–4] that are typically distinguished by their robust metallic surface state encapsulating a nonconducting bulk. From this perspective, a major shortcoming of the early experiments was that carrier doping levels were sufficiently high, resulting in the bulk bands being metallic, as opposed to insulating. This made it almost impossible to definitively separate the properties of the surface state from that of the bulk, and the system did not behave as a TI because the bulk is conducting rather than insulating. This leads to the problem that while spectroscopic measurements such as angle-resolved photoemission spectroscopy (ARPES) demonstrate the clear existence of the expected surface topological bands, transport measurements have been difficult to interpret due to the coexistence of both bulk and surface conduction.

It is, therefore, encouraging that several recent experimental studies[5–8] report the direct observation of two-dimensional (2D) surface states in transport measurement. These reports claim to observe the electronic properties of the surface state with energy close to the topologically protected band-crossing point (also called the Dirac point). Our previous work on the electronic properties of graphene[9–12] would lead us to expect that close to the Dirac point the energy landscape and the spatial electronic structure would become highly inhomogeneous, breaking the surface into puddles of electrons and holes. This physically intuitive picture of inhomogeneous electron and hole puddles has been highly successful in understanding both transport experiments[13,14] and scanning probe experiments[15] in graphene, where the charge inhomogeneity results in a low-density conductivity plateau with a nonvanishing minimum conductivity at the Dirac point. Since topological insulators and graphene have different electronic structure and screening properties, for the case of these topological surface states, it is not *a priori* clear what will be the energy scale of the electron-hole puddles and the corresponding signature of such inhomogeneity in transport and scanning probe microscopy.

To understand the role of a locally fluctuating carrier density on the electronic screening and conductivity of the 2D surface states in these new 3D TI materials, we perform an effective medium analysis where the local conductivity is determined using a Boltzmann transport theory in the presence of disorder, and where the effective disorder strength itself is calculated self-consistently using the random-phase approximation (RPA) screening theory. Our main finding is that band asymmetry, which can almost always be ignored when studying graphene, plays a crucial role in TI experiments. One striking example is that in some cases, the TI surface conductivity does not even have a minimum conductivity in the vicinity of the band-crossing point, making the physics very different from that of graphene. Comparing our theory with the existing Bi_2Se_3 experiments,[5–8] we also conclude that the current systems have very large background disorder making it difficult to access the physics of the Dirac point.

The starting point for our calculation is to make some reasonable approximation for the band structure of the topological surface state. One approach[16,17] would be to perform *ab initio* calculations and fit that data to the most general model Hamiltonian allowed by symmetry. However, we find significant discrepancy between the electronic structure calculations and the photoemission (ARPES) experiments. As a result, a reasonable comparison with experiment would require that we fine tune a model Hamiltonian with 12 parameters. While this procedure could be done, it would be unnecessarily cumbersome and obscure any physical insight. Instead, we follow the minimal model proposed earlier[18] in the literature mimicking the full *ab initio* band structure,

$$\mathcal{H}(\mathbf{k}) = \frac{\hbar^2 k^2}{2m^*} + \hbar v_{\text{F}}(k_x \sigma_y - k_y \sigma_x), \qquad (1)$$

where (σ_x, σ_y) is a 2D vector of Pauli matrices, $\mathbf{k} = (k_x, k_y)$ is the 2D wave vector, v_{F} is the Fermi velocity of the Dirac bands, and effective mass m^* characterizes the degree of asymmetry between the electron and hole bands. Estimates for the values of these two parameters for Bi_2Se_3 vary widely in the experimental literature. For example, values for v_{F} vary from 2×10^5 m/s (Ref. 19) to 6.4×10^5 m/s (Ref. 5), and measured values for m^* vary from $0.11m_e$ (Ref. 20) (m_e is the electron mass) to $0.32m_e$ (Ref. 21). This situation should be contrasted with graphene, where v_{F} is the single band parameter, and most experimental reports agree on its value to within 5%.[9] We use

the two-parameter (m^* and v_F) model of Eq. (1) in the current work.

Since we are concerned with the screening properties of electrons, it is useful to define an interaction parameter $r_s = e^2/(\kappa \hbar v_F)$, where we reiterate that throughout this work, v_F is the parameter in Eq. (1) characterizing the Dirac-like bands; only at low carrier density does it coincide with the Fermi velocity. Here, κ is approximately half of the dielectric constant of the bulk Bi_2Se_3 insulator [whose reported value varies from around 30 (Ref. 19) to about 55 (Ref. 22)]. We certainly expect that as more experiments on TI materials become available, these parameters will become better known both for Bi_2Se_3 and other related materials. To make our theory more compact, we formulate everything in terms of r_s and a characteristic density $n_0 = (m^* v_F)^2/(4\pi\hbar^2)$, where reasonable values of r_s are in the range 0.05 to 0.5, and reasonable values of n_0 are from 10^{11} cm^{-2} to 3×10^{13} cm^{-2}. We note that n_0 is an important parameter characterizing the deviation of the system (for $n > n_0$) from purely Dirac-like behavior—in graphene, n_0 is very large.

The Thomas-Fermi screening theory for electrons specifies that all external potentials are screened by a surface 2D dielectric function $\epsilon(q) = 1 + q_{TF}/q$. For the Hamiltonian in Eq. (1), we find

$$q_{TF} = \frac{r_s k_F}{1 + \text{sgn}(n)\sqrt{|n|/n_0}} = \eta r_s k_F, \qquad (2)$$

where sgn is the signum function, and we use the convention that electrons have $\text{sgn}(n) = +1$ while holes have $\text{sgn}(n) = -1$. For both electrons and holes, we have $k_F = \sqrt{|4\pi n|}$. Note that q_{TF} diverges for $n \to -n_0$, implying perfect screening associated with the diverging density of states in Eq. (1) arising from the quadratic dispersion and the band asymmetry. It turns out that the theory can be completely characterized by the two parameters r_s and n_0, rather than the three microscopic parameters m^*, v_F, and κ.

Our numerical analysis using the full random-phase approximation (RPA) shows that the Thomas-Fermi analysis is

accurate provided that we restrict the carrier density for holes to $|n| < |n_0|$. No such restriction is required for electrons. Within the Boltzmann transport approximation, the conductivity $\sigma = (k_F \ell)(e^2/2h)$, where $\ell = v_F \tau/\eta$ is the mean free path. The scattering time is calculated within the Born approximation as

$$\frac{\hbar}{\tau} = \pi n_{imp} \sum_{\mathbf{k'}} \left| \frac{v(q)}{\epsilon(q)} \right|^2 \sin^2[\theta_{kk'}]\delta(\varepsilon_k - \varepsilon_{k'}), \qquad (3)$$

where n_{imp} is the surface density of random charged impurities, ε_k is the carrier energy, $\theta_{kk'}$ is the scattering angle between wave vectors $\mathbf{k}$ and $\mathbf{k'}$, and $v(q)/\epsilon(q)$ is the Fourier transform of the screened impurity potential. For the purpose of this calculation, we assume that the dominant scatterers are long-ranged Coulomb impurities (with an average 2D density of n_{imp} placed an average distance of d away from the TI surface), although our formalism can easily be generalized to other types of impurities. We also ignore strong potential defects that couple the bulk and surface states, as was considered recently in Ref. 23. For these charged impurities, the conductivity can be calculated analytically, giving

$$\sigma_B[n, n_0, r_s] = \frac{1}{8} \frac{e^2}{h} \frac{n}{n_{imp}} \frac{1}{F_1[\eta r_s/2]},$$

$$\frac{F_1[x]}{x^2} = \frac{\pi}{4} + 3x - \frac{3x^2\pi}{2} + x(3x^2 - 2)\frac{\arccos[1/x]}{\sqrt{x^2 - 1}},$$

$$\eta[n/n_0] = \frac{\sqrt{n_0}}{\sqrt{n_0} + \text{sgn}(n)\sqrt{|n|}}. \qquad (4)$$

The results for the Boltzmann transport theory are shown in Fig. 1. One immediately observes that the asymmetry in the conductivity is quite pronounced compared to a linear Dirac dispersion ($m^* = \infty$; also shown). The electron branch has a much larger conductivity with $\sigma_B(n)$ being superlinear, while the hole branch has a much lower sublinear $\sigma_B(n)$. This pronounced asymmetry between electron and hole transport, following directly from the band asymmetry of Eq. (1), is a characteristic feature of 2D TI transport.[24]
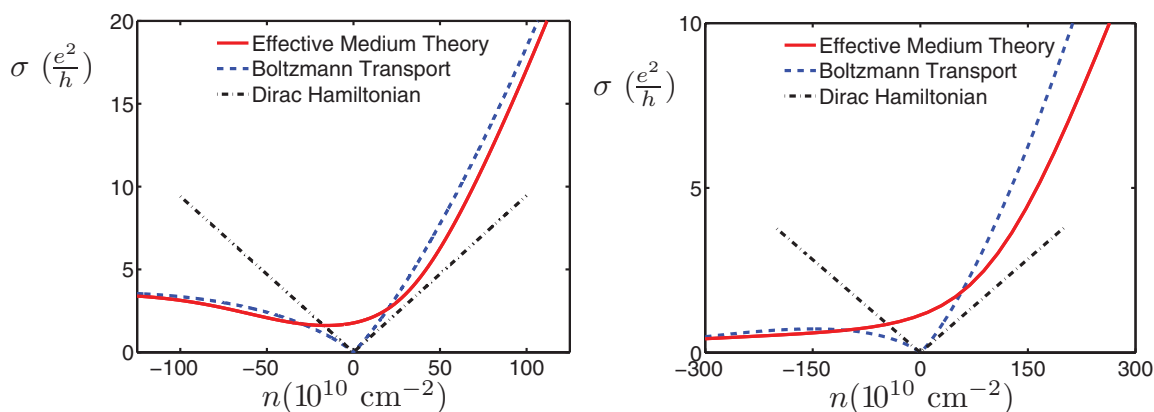

FIG. 1. (Color online) Theoretical calculations for the dependence of conductivity on carrier density using parameters $r_s = 0.1$ and $n_0 = 2 \times 10^{13}$ cm^{-2}. The left panel has impurity concentration $n_{imp} = 10^{13}$ cm^{-2}, and the right panel is dirtier with $n_{imp} = 5 \times 10^{13}$ cm^{-2}. Dashed blue lines show the Boltzmann transport result described in Eq. (4), and dash-dotted black lines are for Dirac bands, i.e., $m^* \to \infty$ in Eq. (1). The red solid line is the result of the effective-medium theory calculation obtained by solving Eq. (6) using $d = 0.1$ nm. The left panel has a conductivity minimum in the vicinity of the Dirac point, while for the right panel the conductivity monotonically increases as the carriers are tuned from holes (negative n) to electrons (positive n).
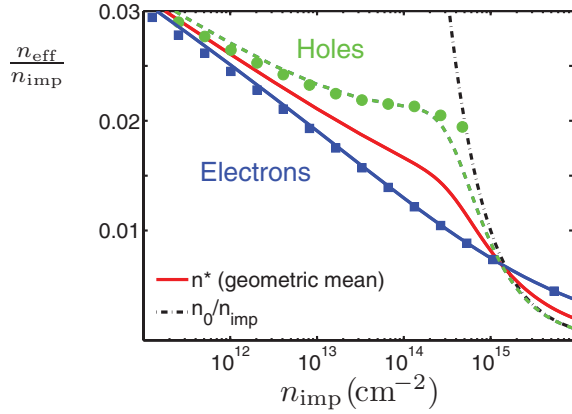
FIG. 2. (Color online) Theoretical calculations for disorder-induced carrier density n_{eff} as a function of impurity concentration n_{imp} using parameters $r_s = 0.1$, $n_0 = 10^{13}$ cm^{-2}, and $d = 0.1$ nm. The curves were calculated using the Thomas-Fermi screening theory, while the circles (holes) and squares (electrons) show the RPA. For electrons, the two approximations agree to within our numerical accuracy. For holes, they agree only for $n_{eff} < n_0$ (shown as a dotted line). One implication of this result is that the charge neutrality point will be distinct from the Dirac point.

At low carrier density, the disorder-induced fluctuations in carrier density become larger than the average carrier density. In particular, when the average carrier density vanishes with the chemical potential at the Dirac point, one might expect that the electronic properties of the system are determined by the typical carrier density inside the electron and hole puddles. For example, one could define a carrier density distribution function $P[n]$, and the condition of zero average carrier density implies the vanishing of the first moment of $P[n]$. The second moment of the carrier density distribution n_{rms} would then determine how the carriers in this inhomogeneous system screen any external potential, including the impurity potential that induced the fluctuations to begin with. This implies that the density fluctuations need to be calculated self-consistently.[11] We calculate the properties of this inhomogeneous system by assuming a global screening function that depends on the impurity profile only through an effective carrier density n_{eff} (where knowing n_{eff}, one can then calculate all moments of $P[n]$ including n_{rms}, e.g., for Dirac fermions, $n_{rms} \approx \sqrt{3}n_{eff}$).

This effective carrier density n_{eff} is nothing other than a measure of the typical carrier density inside the electron and hole puddles. After obtaining n_{eff}, we can then compute other properties of the Dirac point including its conductivity (that can be measured in a transport experiment) or its density-density correlation function [measured using a scanning tunneling microscope (STM)]. Calculating n_{eff} is therefore a central result of this work. We do this by requiring that the density induced by the second moment of the screened disorder potential is precisely the same as the density entering the global screening function. Applying this procedure to Eq. (1), and defining the dimensionless variable $y = |n_{eff}|/n_0$, we derive a system of equations that can be easily solved numerically:

$$\frac{y^2}{4} + y + sy^{3/2} = AC_0\left[\frac{B\sqrt{y}}{1 + s\sqrt{y}}\right],$$

$$C_0[x] = \partial_x\left[xe^x\int_x^\infty t^{-1}e^{-t}dt\right], \qquad (5)$$

$$A = \frac{1}{2}\frac{n_{imp}}{n_0}r_s^2, \quad B = 2r_s d\sqrt{4\pi n_0}.$$

Here, $s = \text{sgn}(n)$ denotes the electron ($s = +1$) and hole ($s = -1$) bands.

In Fig. 2, we show the effective carrier density obtained using this self-consistent method. Notice that the induced carrier density is different for electrons and for holes except in the limiting case $n_{imp} \ll n_0$ [where Eq. (1) gives a symmetric linear Hamiltonian]. This asymmetry implies that the charge neutrality point (where the number of electrons and holes are equal) does not necessarily coincide with the Dirac point (the crossing point between electron and hole bands). The numerical value of n_{eff}/n_{imp} is rather small (less than 0.03). This not only guarantees convergence of the theory, but also has important implications for experiments. Since n_0 and n_{imp} are comparable in current experiments, if n_{eff}/n_{imp} is not small, this would imply that the energy scale associated with the disorder-induced inhomogeneity would be comparable to the bulk band gap, and there would be no hope of observing any physics associated with the topological surface state. We note that in graphene, $n_{eff} \sim n_{imp}$, which is in contrast to 2D TI transport where we are finding $n_{eff} \ll n_{imp}$.
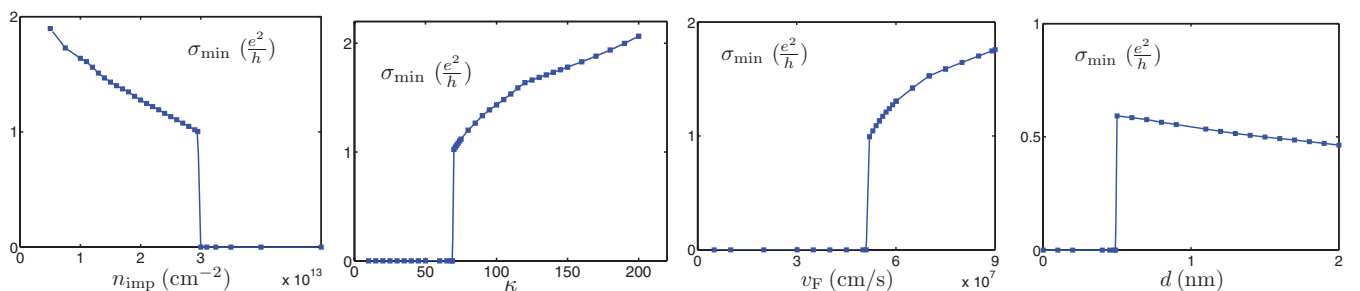

FIG. 3. (Color online) Minimum conductivity as a function of system parameters: (a) impurity concentration n_{imp}, (b) dielectric constant κ, (c) low-energy Fermi velocity v_F, and (d) impurity distance d. The values for the parameters not being varied are $n_{imp} = 5 \times 10^{13}$ cm^{-2}, $v_F = 4.5 \times 10^5$ m/s, $\kappa = 50$, $d = 0.1$ nm, and $m^* = 0.2$. A zero value for σ_{min} means that there was no local minimum in the conductivity in the vicinity of the Dirac point. Solid curves are a guide to the eye.

We now proceed to calculate the electronic transport in this inhomogeneous puddle-dominated carrier density landscape. The procedure is to first assume that the local conductivity can be calculated from the local carrier density using Eq. (4). Then, an effective-medium theory (EMT) can be used to calculate the global conductivity from this distribution of spatially fluctuating local conductivities,[12]

$$\int dn\, P[n]\, \frac{\sigma_B(n) - \sigma_{\text{EMT}}}{\sigma_B(n) + \sigma_{\text{EMT}}} = 0. \qquad (6)$$

The EMT assumes that the dominant contribution to the resistivity arises from scattering inside the electron and hole puddles, and not across the puddles. This is a reasonable assumption because the cross puddle backscattering is suppressed due to spin conservation. To simplify the theory, we assume that $P[n]$ is Gaussian, with variance $\sqrt{3}n^*$, where $n^* = \sqrt{n_{\text{eff}}^e n_{\text{eff}}^h}$ is the geometric mean of the electron and hole density fluctuations (see Fig. 2). Our results are shown in Fig. 1. An important result of performing the EMT average using an asymmetric conductivity is that the carrier density at which the conductivity is minimum is distinct from both the charge neutrality point and Dirac point, further complicating the analysis and interpretation of experiments. This is illustrated dramatically in the right-hand panel, where there is no conductivity minimum in the vicinity of the Dirac point. Rather, the conductivity increases monotonically from $\sigma = 0$ for $n = -n_0$ to $\sigma_B(|n|)$ for $n \gtrsim n_0$. Indeed, in the experiments of Ref. 5, the most disordered sample does not show a (clear) minimum conductivity, while other devices do have a local σ_{min} close to the Dirac point, similar to that shown in the left-hand panel of Fig. 1. The possible nonexistence of a conductivity minimum in 2D TI transport is a qualitative prediction of our theory.

We have explored how σ_{min} depends on a variety of physical parameters. In most cases, we find that the minimum conductivity *increases* with *increasing* κ, v_F, m^*, and sample purity (n_{imp}^{-1}). However, there is also a large range of parameter space where there is no conductivity minimum associated with the Dirac point (we denote this case as $\sigma_{\text{min}} = 0$). We illustrate our results in Fig. 3, where we show the dependence of σ_{min} on κ, v_F, and d. As seen in the figure, in each of these cases, the crossover between the presence or absence of a well-defined σ_{min} occurs quite sharply as a function of the parameter being varied.

The analytic results provided in Eq. (5) were done using the Thomas-Fermi (TF) approximation. We have also done calculations using the full static RPA. We find in general, as shown in Fig. 2, very good agreement between the TF and RPA results, with the RPA results being obtained completely numerically.

Finally, we turn our attention to the recent experiments which have directly observed the electron-hole puddles on
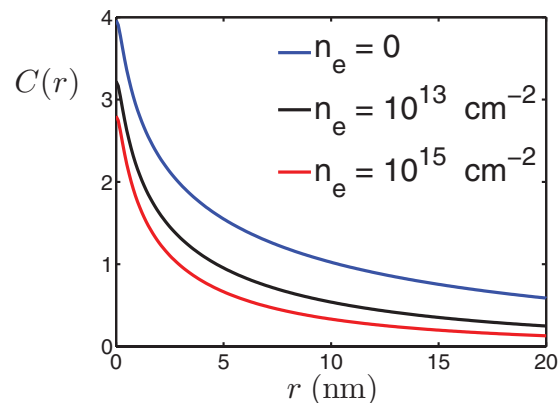

FIG. 4. (Color online) Theoretical calculations for potential correlation fluctuation $C(r) = \langle [V(r) - \bar{V}][V(0) - \bar{V}] \rangle$ for the electron band using the same parameters as in Fig. 1(b). The y axis is normalized by $(r_s \hbar v_F \sqrt{\pi n_{\text{imp}}})^2$. From top to bottom, the curves represent average electron doping of $n = 0$, 10^{13} cm^{-2}, and 10^{15} cm^{-2}.

the Bi$_2$Se$_3$ surface.[19] In addition to being a local probe, by tuning the bias voltage between the tip and the sample, these experiments can probe features in the density of states away from the Fermi energy. In current TI materials, this is especially important because unintentional doping pushes the Fermi energy far from the Dirac point. For each position in space, one can map out the energy of the Dirac point E_D. Shifts of E_D away from the average potential $\bar{V}$ give the screened disorder potential landscape whose autocorrelation function is $C(r) = \langle [V(r) - \bar{V}][V(0) - \bar{V}] \rangle$. Theoretical calculations for the correlation function are shown in Fig. 4 for different (average) electron carrier densities n using the Thomas-Fermi approximation and solving for the total density $n_{\text{tot}} = n + n_{\text{eff}}$ self-consistently. The effect of electron doping is to slightly reduce both the magnitude of the potential fluctuations $C(0)$ as well as the spatial correlation length of the puddles.

Our predictions for screening and carrier transport in disordered 2D TI surface states should be testable in future experiments. In particular, the band asymmetry gives rise to interesting and qualitatively novel effects, e.g., causing the Dirac point, the charge neutrality point, and the minimum conductivity to occur at different carrier densities. In addition, electron and hole 2D transport in TIs should manifest strong asymmetry, and in some situations with strong disorder (i.e., large n_{imp} or small d) there may not be any minimum conductivity plateau associated with the Dirac point. All of these features have important implications for using transport measurements to characterize these TI surface states, and eventually when utilizing them as an electronic material.

---

[1] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).

[2] J. E. Moore, Nature (London) **464**, 194 (2010).

[3] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).

[4] D. Culcer, Physica E **44**, 860 (2012).

[5] D. Kim, S. Cho, N. P. Butch, P. Syers, K. Kirshenbaum, S. Adam, J. Paglione, and M. S. Fuhrer, Nature Phys. **8**, 460 (2012).

[6] S. S. Hong, J. J. Cha, D. Kong, and Y. Cui, Nature Comm. **3**, 757 (2012).

[7] D. Kong, Y. Chen, J. J. Cha, Q. Zhang, J. G. Analytis, K. Lai, Z. Liu, S. S. Hong, K. J. Koski, S. K. Mo, Z. Hussain, I. R. Fisher, Z. X. Shen, and Y. Cui, Nature Nano. **6**, 705 (2011).

[8] H. Steinberg, J.-B. Laloe, V. Fatemi, J. S. Moodera, and P. Jarillo-Herrero, Phys. Rev. B **84**, 233101 (2011).

[9] S. Das Sarma, S. Adam, E. H. Hwang, and E. Rossi, Rev. Mod. Phys. **83**, 407 (2011).

[10] E. H. Hwang and S. Das Sarma, Phys. Rev. B **75**, 205418 (2007).

[11] S. Adam, E. H. Hwang, V. M. Galitski, and S. Das Sarma, Proc. Natl. Acad. Sci. USA **104**, 18392 (2007).

[12] E. Rossi, S. Adam, and S. Das Sarma, Phys. Rev. B **79**, 245423 (2009).

[13] Y.-W. Tan, Y. Zhang, K. Bolotin, Y. Zhao, S. Adam, E. H. Hwang, S. Das Sarma, H. L. Stormer, and P. Kim, Phys. Rev. Lett. **99**, 246803 (2007).

[14] J. H. Chen, C. Jang, S. Adam, M. S. Fuhrer, E. D. Williams, and M. Ishigami, Nature Phys. **4**, 377 (2008).

[15] S. Adam, S. Jung, N. N. Klimov, N. B. Zhitenev, J. A. Stroscio, and M. D. Stiles, Phys. Rev. B **84**, 235421 (2011).

[16] H. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, Nature Phys. **5**, 438 (2009).

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

[18] D. Culcer, E. H. Hwang, T. D. Stanescu, and S. Das Sarma, Phys. Rev. B **82**, 155457 (2010).

[19] H. Beidenkopf, P. Roushan, J. Seo, L. Gorman, I. Drozdov, Y. San Hor, R. J. Cava, and A. Yazdani, Nature Phys. **7**, 939 (2011).

[20] J. G. Analytis, R. D. McDonald, S. C. Riggs, J.-H. Chu, G. S. Boebinger, and I. R. Fisher, Nature Phys. **6**, 960 (2010).

[21] A. A. Taskin, Z. Ren, S. Sasaki, K. Segawa, and Y. Ando, Phys. Rev. Lett. **107**, 016801 (2011).

[22] N. P. Butch, K. Kirshenbaum, P. Syers, A. B. Sushkov, G. S. Jenkins, H. D. Drew, and J. Paglione, Phys. Rev. B **81**, 241301 (2010).

[23] A. M. Black-Schaffer and A. V. Balatsky, Phys. Rev. B **85**, 121103(R) (2012).

[24] The calculation here is done for zero applied magnetic field. For weak fields, one can plug the calculated σ_B into Eq. (4) of E. H. Hwang, S. Adam, and S. Das Sarma, Phys. Rev. B **76**, 195421 (2007) to obtain the leading-order response to a magnetic field.