Electrically tunable surface-to-bulk coherent coupling in topological insulator thin films

H. Steinberg,[1] J.-B. Laloë,[2] V. Fatemi,[1] J. S. Moodera,[1,2] and P. Jarillo-Herrero[1]
[1]*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[2]*Francis Bitter Magnet Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
(Received 14 July 2011; revised manuscript received 1 October 2011; published 5 December 2011)

We study coherent electronic transport in charge-density-tunable microdevices patterned from thin films of the topological insulator (TI) Bi_2Se_3. The devices exhibit pronounced electric field effect, including ambipolar modulation of the resistance with an on-and-off ratio of 500%. We show that the weak antilocalization correction to conductance is sensitive to the number of coherently coupled channels, which in a TI includes the top and bottom surfaces and the bulk carriers. These are separated into coherently independent channels by the application of gate voltage and at elevated temperatures. Our results are consistent with a model where channel separation is determined by a competition between the phase coherence time and the surface-to-bulk scattering time.

DOI: 10.1103/PhysRevB.84.233101 PACS number(s): 73.20.Fz, 71.70.Ej, 72.15.Rn, 73.25.+i

Topological insulators (TIs) are gapped bulk insulators with gapless Dirac surface states which have emerged as a new paradigm in the study of topological phases of matter.[1] TI-based electronic devices are attractive as platforms for spintronic applications,[2] and for the detection of emergent properties such as Majorana excitations,[3] electron-hole condensates[4] and the topological magnetoelectric effect.[5] Most theoretical proposals envision an experimental geometry consisting of a planar TI device, where electrical current is carried by the surface states.

Despite considerable recent evidence of TI surface states in angle-resolved photoemission spectroscopy (ARPES)[6,7] and scanning tunneling microscopy (STM),[8] their observation and manipulation in transport experiments remain difficult: TI devices require a density-tunable surface state which is decoupled from the residual bulk carriers. Realistic devices, however, conduct through parallel channels consisting of the top and bottom surfaces, and of bulk carriers which, due to unintentional doping, can account for a significant part of the conductance, limit the surface density tunability, and create an uncertainty in the surface-to-bulk coupling. It is therefore desirable to minimize the bulk contribution and to simultaneously investigate the various ways in which it is involved in electronic transport. Suppression of the bulk channel can be obtained by minimizing carrier density, as demonstrated in millimeter-size single crystals,[9] and by fabrication of nanoscale devices, such as nanoribbons[10] and flakes exfoliated from single crystals.[11–13] The latter studies have also demonstrated that the surface carrier density can be tuned by electrostatic gating and have all detected an ambipolar modulation of conductance.

An alternative route for fabrication of thin TI devices employs the growth of thin films in ultrahigh vacuum chambers. Thin-film growth offers fine control over geometry and composition and a straightforward approach for the growth of heterostructures. High-quality Bi_2Se_3 thin films were studied by *in situ* STM[14] and ARPES.[15] Transport studies[16,17] carried out on the same films are still dominated by bulk transport since the material is highly doped. Interestingly, most thin-film studies report a pronounced weak antilocalization (WAL) feature, which is an indication of phase coherent transport.

WAL and weak localization emerge from the correction of coherent time-reversed closed paths to electronic transport.[18] They are sensitive to the competition between the phase coherence time τ_ϕ and other time scales, and have been extensively employed as a probe for coherent dynamics in solid-state systems.[19] In TIs both the bulk and surface states may contribute to coherent transport, and in both cases they should exhibit WAL: In the bulk, the strong spin-orbit (SO) coupling leads to random rotations of the spin orientation,[20] which on average result in destructive interference for backscattering paths, hence leading to enhanced conductance. On the chiral TI surface state, momentum is coupled to the spin degree of freedom, so time-reversed paths around a closed trajectory acquire a relative phase of π.[21]

Given the various parallel conduction channels, so far it was not clear why most of the recent TI transport studies report WAL corrections corresponding to a single coherent channel,[16,17] and it was suggested that one of the surfaces has a significantly higher coherence length than the other.[17] A few studies have found the number of channels to be tunable by a gate voltage,[13,17,22] but the microscopic mechanism underlying this tunability was not thoroughly investigated. Here we study the WAL effect in charge-density-tunable Bi_2Se_3 devices. We find that the number of independent coherent channels is tunable by the applied electric field and by the temperature. Our results indicate that bulk carriers play a crucial role in TI coherent transport, and that channel coupling is controlled by a competition between the phase coherence time and the surface-to-bulk coupling time.

We fabricate low-density Bi_2Se_3 devices by growing large-area thin films on a Si(111) substrate and subsequent lithographic processing.[23] A device image and schematics are shown in Figs. 1(a) and 1(b). X-ray diffraction [Fig. 1(c)] reveals sharp $(0,0,3n)$ peaks, confirming that the films are c-axis oriented along the growth direction. The Bi_2Se_3 is typically capped *in situ* by a sputtered layer of 4-nm Al_2O_3. An additional 16 nm of HfO_2 gate dielectric is deposited after device patterning.

Figure 2 shows the electric field effect behavior of three representative devices (devices A1, A2, A3), all patterned from the same 20-nm-thick capped Bi_2Se_3 wafer, and measured at $T = 4$ K. Every resistance (conductance) shown is resistance (conductance) per square. For $V_{TG} > 0$, R is modulated weakly by the gate, but begins rising sharply at approximately $V_{TG} = 0$, and peaks at 5–6 kΩ at approximately $V_{TG} = -5$ V.

 ©2011 American Physical Society

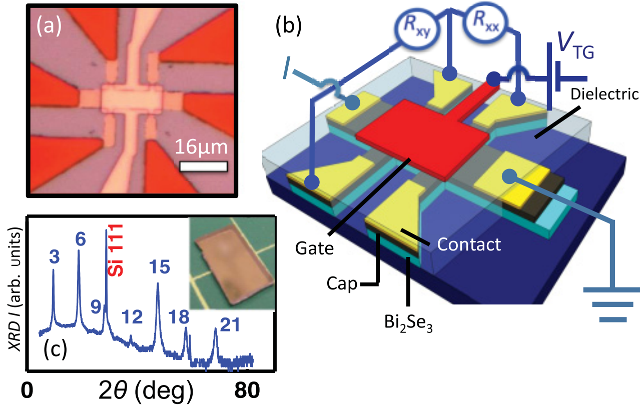

FIG. 1. (Color online) (a) Optical image of a gated Bi_2Se_3 device, showing a Bi_2Se_3 Hall bar, evaporated contact pads, and top gate. The gate electrode covers the entire area of the Hall bar. (b) Schematic of the device and measurement circuitry, showing the Bi_2Se_3 layer, cap layer, contact pads, dielectric, and top gate. (c) X-ray diffraction intensity of a Bi_2Se_3 thin film, showing the (0,0,3n) family of diffraction lines indicating growth along the c axis. Some lines associated with the Si substrate are also visible. Inset: Image of Bi_2Se_3 film grown on a Si wafer.

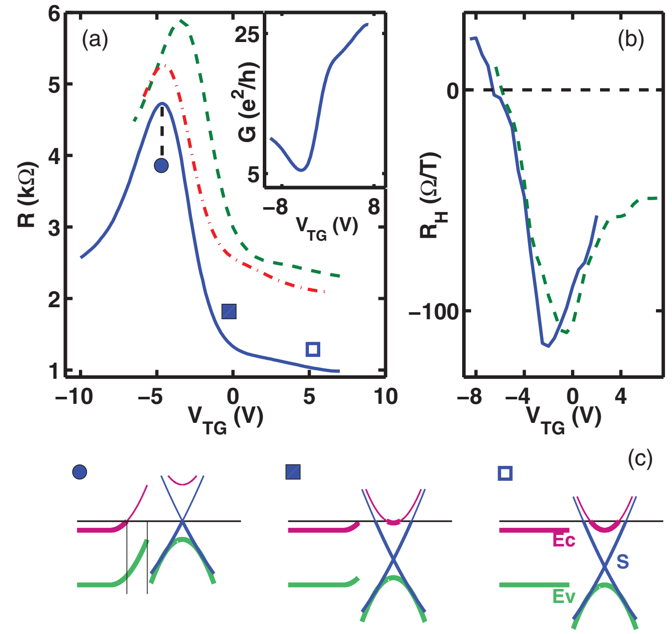

FIG. 2. (Color online) (a) Square resistance R vs top-gate voltage $V_{\rm TG}$ measured on devices A1 (continuous line), A2 (dashed), and A3 (dashed-dotted) at 4 K. Inset: Square conductance G for device A1. (b) Low-field Hall coefficient R_H in two-dimensional (2D) units. R_H is observed to cross over to positive values at $V_{\rm TG} \sim -7$ V. (c) Schematics of the band structure and spatial variation of bulk bands near the surface at different gate voltages. In each subpanel, the left-hand side shows the band energy vs vertical position, and the right-hand side is the energy-momentum dispersion at the surface. The bulk conduction band is marked by "Ec," the valence band by "Ev," and surface by "S." Right-hand panel (hollow square): $V_{\rm TG} > 0$. The bulk states near the surface are populated. Middle panel (solid square): Bulk states near the surface are at the depletion point (bottom of conduction band). Left-hand panel (circle): $V_{\rm TG} < 0$. The top surface state is gated to the Dirac point. The bulk bands bend to form a depletion layer near the surface, marked by vertical lines. This depletion region increases with increasing negative $V_{\rm TG}$. The markers in (a) correspond to the band-structure schemes shown in the bottom panels.

The inset shows the square conductance G of device A1 in units of e^2/h. For this device G varies between 5 and $25e^2/h$, exhibiting a pronounced ambipolar modulation characteristic of Dirac dispersions, and indicating that the conductance is dominated by the top surface. This is supported by the Hall coefficient R_H, which for a single channel is given by $R_H = -1/ne$, where n is the density of charge carriers and e is the electron charge, but which is more complex for multichannel systems.[11] R_H is strongly modulated by the gate and crosses over to positive values [Fig. 2(b)] at $V_{\rm TG} \sim -7$ V, indicating that the dominant Hall current-carrying population has changed from electrons to holes, as expected for Dirac systems with low doping. This is in stark contrast to previously reported studies on Bi_2Se_3 thin films, which are very strongly electron doped, and do not exhibit such a strong carrier-type modulation.[16,17] Gating of the surface states shifts the surface energy bands vertically and hence is accompanied by bending of the bulk bands near the surface, as shown in Fig. 2(c). We associate the sharp change in the slope of $R(V_{\rm TG})$ near $V_{\rm TG} = 0$ with the depletion of bulk carriers immediately near the surface. For $V_{\rm TG} < 0$, the main effect of the gate is to change the TI surface-state charge density, which results in a more rapid variation of conductance. However, any change of the surface charge density has to be accompanied by a change in the width of the bulk depletion region. This depletion region is crucial for coherently separating the surface state from the bulk bands.

We now turn to the electric field effect on coherent transport, studied by magnetoconductance. These data were taken on device B, which is patterned from a different wafer than the one reported above but has similar transport characteristics, with a well-defined resistance maximum [Fig. 3(b)]. Figure 3(a) shows R versus perpendicular magnetic field B at $T = 0.3$ K and where WAL appears as a sharp suppression of resistance at low magnetic field. We repeat the same measurement for different gate voltages, and find that the WAL feature also evolves with $V_{\rm TG}$, as shown in Fig. 3(c), where the change in conductance $\Delta G(B) = G(B) - G(0)$ is plotted at $V_{\rm TG} = +6$ V and $V_{\rm TG} = -8$ V. Both traces exhibit a correction of $\Delta G \sim e^2/h$, although the former is sharper in the magnetic field. The WAL correction depends on the phase-coherence time τ_ϕ, or length l_ϕ, which are related through $l_\phi = \sqrt{D\tau_\phi}$ (D being the diffusion constant). The data agree well with[24]

$$\Delta G_{\rm WAL}(B) = -\alpha \frac{e^2}{2\pi^2\hbar}\left[\log\frac{B_o}{B} - \psi\left(\frac{1}{2} + \frac{B_o}{B}\right)\right], \quad (1)$$

where $B_o = \hbar/(4el_\phi^2)$, ψ is the digamma function, and α is a prefactor which should be equal to $-1/2$ for a single coherent channel. Equation (1) is valid for massless Dirac fermions (i.e., the surface states)[21] and is also valid in the bulk where the spin-orbit scattering time $\tau_{\rm SO}$ is significantly shorter than τ_ϕ.[23] The data (blue/dark gray) agree well with the fit to Eq. (1) (yellow/light gray) over the entire measured range in both cases, indicating that WAL constitutes the entire correction

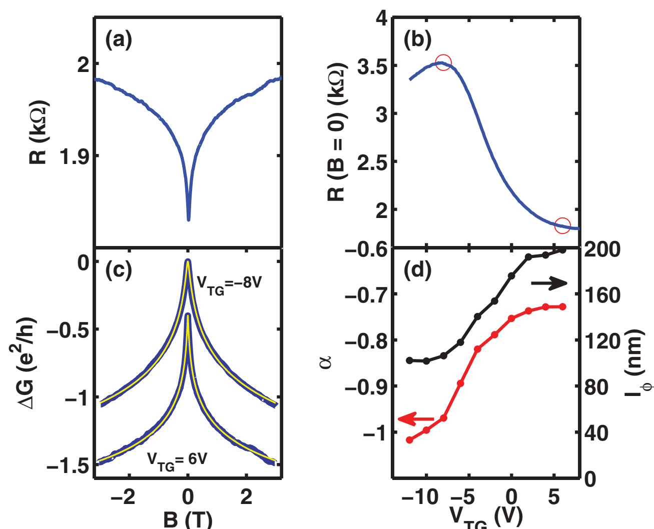

FIG. 3. (Color online) (a) Magnetoconductance of device B taken at $V_{TG} = +6$ V, showing a pronounced WAL suppression of resistance. (b) $R(B = 0)$ vs V_{TG}. R peaks at $V_{TG} = -8$ V. (c) ΔG vs B plotted for $V_{TG} = +6$ and -8 V [marked by circles in (b)]. The two plots are offset for clarity. The data are plotted in blue/dark gray, and fits to WAL theory [Eq. (1)] are plotted in yellow/light gray. (d) Coherence length l_ϕ and prefactor α extracted from Eq. (1) for each V_{TG}. Both l_ϕ and α are modulated by the applied gate voltage.

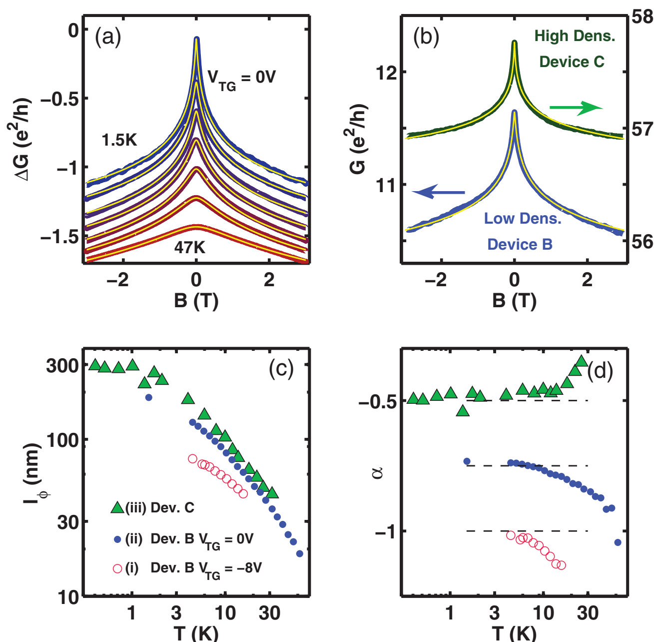

FIG. 4. (Color online) (a) Magnetoconductance ΔG measured on low-density device B at $V_{TG} = 0$ V for different temperatures (from 1.5 to 47 K). The curves are vertically shifted for clarity. Superimposed on each scan, in yellow/light gray, is the fit to Eq. (1). (b) WAL traces and fits to Eq. (1) taken on low-density device B and high-density device C. Device B: $l_\phi = 175$ nm, $\alpha = -0.75$; device C: $l_\phi = 300$ nm, $\alpha = -0.5$. (c) l_ϕ vs T for the three regimes discussed in the text. Open circle: Device B at the resistance peak (i). Full circle: Device B at $V_{TG} = 0$ V (ii). Triangle: Device C (iii). (d) α vs T for all three regimes.

to magnetoconductance. The fit contains two free parameters, the coherence length l_ϕ and the prefactor α, both of which change with the applied gate voltage [Fig. 3(d)]. Here we focus on the gate tunability of α, which changes from -0.7 to ~ -1, reflecting a change in the effective number of coherent channels.

From the WAL point of view, the surface and bulk states can be regarded as independent phase coherent channels as long as the carriers in one channel lose coherence before being scattered into the other. In this case, the majority of closed loops responsible for WAL will involve states from a single channel, and each channel will exhibit its own WAL correction. This condition can be formulated as $\tau_{SB} > \tau_\phi$, where τ_{SB} is the effective surface-to-bulk scattering time. In the opposite limit $\tau_{SB} < \tau_\phi$, charge carriers scatter between the bulk and surface states while maintaining phase coherence, effectively becoming a single phase-coherent channel. We therefore interpret the gate dependence of α as direct modulation of τ_{SB} via the formation of a depletion layer between the top surface and bulk carriers [see Fig. 2(c)], spatially separating them and suppressing the scattering probability between them.

It is important to discuss the validity of Eq. (1) when the coupling between channels is tunable. In the fully decoupled regime, each channel i has correction ΔG_i which follows Eq. (1) (with $\alpha = -0.5$) and depends on $l_{\phi,i}$, yielding for the total correction $\Delta G_{tot} = \Delta G_1 + \Delta G_2$. Decomposing Eq. (1) into its logarithmic and digamma components,[23] we note that the latter approaches a constant value for $B > B_o \approx$ 10–20 mT. Since our data extend to a few T, where the change in ΔG is dominated by the logarithmic component, Eq. (1) is a very good approximation to ΔG_{tot}, with $\alpha = -1$ and an effective coherence length $l_\phi^{eff} = \sqrt{l_{\phi 1} l_{\phi 2}}$, as shown in Ref. 23. As a consistency check, we can find the dataset where Eq. (1) yields $\alpha = -1$, and fit the data to ΔG_{tot}.[23] We find the fit to agree very well with the data both for $B > B_o$ and $B < B_o$, with coherence lengths $l_{\phi 1,2} = 135$ and 77 nm. When $-0.5 > \alpha > -1.0$ the physical interpretation of the parameters extracted from Eq. (1) is not trivial and no theoretical model exists to predict the magnetoconductance in the crossover coupling region. The evident success of the fit indicates that the logarithmic correction is robust, and α can be used phenomenologically as a measure for channel separation.

We have so far encountered two effective regimes differing in the degree of channel separation: (i) $V_{TG} = -8$ V, where $\alpha \sim -1 = -(1/2 + 1/2)$, indicating a decoupling of the top surface from the rest of the system; (ii) $V_{TG} = +6$ V, where $\alpha \sim -0.7$, indicating that the top surface is only partially decoupled. In highly doped samples [device C, Fig. 4(b)] we find a third regime, where $\alpha = -0.5$, $R_H = -3.7\Omega/T$, and $G = 57e^2/h$. This result is consistent with other WAL studies[16,17] and suggests that the surface and bulk channels in highly doped TIs are fully coupled. This is consistent with our model, since for $V_{TG} \geqslant 0$ the surface and bulk states coexist in space and $1/\tau_{SB}$ should depend on the momentum difference between the surface and bulk bands, which become closer as the density increases.[7,25]

The effective number of phase coherent channels reflected in the magnitude of α depends on the ratio τ_{SB} to τ_ϕ, which

was controlled above by modulating τ_{SB} via the electric field effect. An independent control over this ratio can be obtained by varying the temperature, which directly controls τ_ϕ through dephasing. Figure 4 shows the change in the WAL correction with temperature in the different mixing regimes discussed above. Figure 4(a) shows $\Delta G(B)$ in regime (ii) from $T = 1.5$ to 47 K, together with the corresponding fits to Eq. (1). Similar data and fits are obtained for regimes (i) and (iii). The corresponding temperature dependence of α and l_ϕ is shown in Figs. 4(c) and 4(d). α decreases with temperature for regimes (ii) (from -0.75 to ~ -1) and (i) (-1 to -1.15), suggesting increased channel separation with temperature. This may seem surprising, since naively one expects larger channel mixing at higher temperatures. However, the behavior of $\alpha(T)$ is consistent with our model, and can be understood once the temperature dependence of l_ϕ is examined. In Fig. 4(c) we see that l_ϕ is strongly temperature dependent, and since $\tau_\phi \sim l_\phi^2$, this indicates that τ_ϕ decreases rapidly as T is increased, which should result in a decrease in the τ_ϕ/τ_{SB} ratio, and consequently an increase in the channel separation. The validity of this model requires that τ_{SB} changes slower than τ_ϕ. This is expected in view of recent mobility data,[11] suggesting that the impurity scattering rate, which should govern τ_{SB}, is indeed nearly temperature independent below 40 K. The results in regime (i), where α becomes smaller than -1, suggest that, in addition to the top-bulk separation identified above, also the bottom surface could be decoupling from the bulk. At very high densities (device C), this behavior is not observed, likely because τ_{SB} is much shorter than τ_ϕ in the temperature range explored. The corresponding increase in α at high T could be consistent with coherent transport being dominated by the bulk, where WAL is suppressed when $\tau_\phi \approx \tau_{SO}$.

Recently we became aware of work by Chen *et al.*,[22] which report a similar modulation of α, but associate it with changes in the coherence lengths of electron and hole channels. Such an interpretation cannot explain our data, since our samples are thicker and the bottom surface density is not tunable, nor can it explain our observed temperature dependence nor the fact that $\alpha = -0.7$ at positive gate voltages.

We are grateful to L. Fu, P. A. Lee, K. Michaeli, L. I. Glazman, and A. Yacoby for useful discussions. H.S. acknowledges financial support from the Israeli Ministry of Science. J.-B.L. and J.S.M. thank support from NSF DMR Grant No. 0504158 and ONR Grant No. N00014-09-1-0177. P.J-H. acknowledges support from a Packard Fellowship. This work was performed in part at the NSF-funded Harvard Center for Nanoscale Systems and MIT Center for Materials Science and Engineering.

[1] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010); J. E. Moore, Nature (London) **464**, 194 (2010); X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).

[2] O. V. Yazyev, J. E. Moore, and S. G. Louie, Phys. Rev. Lett. **105**, 266806 (2010).

[3] L. Fu and C. L. Kane, Phys. Rev. Lett. **100**, 096407 (2008).

[4] B. Seradjeh, J. E. Moore, and M. Franz, Phys. Rev. Lett. **103**, 066402 (2009).

[5] X.-L. Qi, T. L. Hughes, S. Raghu, and S.-C. Zhang, Phys. Rev. Lett. **102**, 187001 (2009).

[6] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature (London) **452**, 970 (2008).

[7] D. Hsieh, Y. Xia, D. Qian, L. Wray, J. H. Dil, F. Meier, J. Osterwalder, L. Patthey, J. G. Checkelsky, and N. P. Ong, Nature (London) **460**, 1101 (2009); Y. Chen, J. Analytis, J. Chu, Z. Liu, S. Mo, X. Qi, H. Zhang, D. Lu, X. Dai, Z. Fang, S. C. Zhang, I. R. Fisher, Z. Hussain, and Z. X. Shen, Science **325**, 178 (2009).

[8] P. Roushan, J. Seo, C. V. Parker, Y. S. Hor, D. Hsieh, D. Qian, A. Richardella, M. Z. Hasan, R. J. Cava, and A. Yazdani, Nature (London) **460**, 1106 (2009); Z. Alpichshev, J. G. Analytis, J.-H. Chu, I. R. Fisher, Y. L. Chen, Z. X. Shen, A. Fang, and A. Kapitulnik, Phys. Rev. Lett. **104**, 016401 (2010).

[9] J. G. Analytis, R. D. McDonald, S. C. Riggs, J.-H. Chu, G. S. Boebinger, and I. R. Fisher, Nat. Phys. **6**, 960 (2010), D.-X. Qu, Y. S. Hor, J. Xiong, R. J. Cava, and N. P. Ong, Science **329**, 821 (2010).

[10] H. Peng, K. Lai, D. Kong, S. Meister, Y. Chen, X.-L. Qi, S.-C. Zhang, Z.-X. Shen, and Y. Cui, Nat. Mater. **9**, 225 (2010).

[11] H. Steinberg, D. R. Gardner, Y. S. Lee, and P. Jarillo-Herrero, Nano Lett. **10**, 5032 (2010).

[12] B. Sacépé, J. B. Oostinga, J. Li, A. Ubaldini, N. J. G. Couto, E. Giannini, and A. F. Morpurgo, e-print arXiv:1101.2352.

[13] J. G. Checkelsky, Y. S. Hor, R. J. Cava, and N. P. Ong, Phys. Rev. Lett. **106**, 196801 (2011).

[14] P. Cheng, C. Song, T. Zhang, Y. Zhang, Y. Wang, J. F. Jia, J. Wang, Y. Wang, B. F. Zhu, X. Chen, X. Ma, K. He, L. Wang, X. Dai, Z. Fang, X. Xie, X. L. Qi, C. X. Liu, S. C. Zhang, and Q. K. Xue, Phys. Rev. Lett. **105**, 076801 (2010).

[15] Y. Zhang, K. He, C.-Z. Chang, C. L. Song, L. L. Wang, X. Chen, J. F. Jia, Z. Fang, X. Dai, W. Y. Shan, S.-Q. Shen, Q. Niu, Z. L. Qi, S. C. Zhang, X. Ma, and Q. K. Xue, Nat. Phys. **6**, 584 (2010).

[16] J. Wang, A. M. DaSilva, C.-Z. Chang, K. He, J. K. Jain, N. Samarth, X.-C. Ma, Q.-K. Xue, and M. H. W. Chan, Phys. Rev. B **83**, 245438 (2011); M. Liu, C.-Z. Chang, Z. Zhang, Y. Zhang, W. Ruan, K. He, L.-l. Wang, X. Chen, J.-F. Jia, S.-C. Zhang, Q.-K. Xue, X.-C. Ma, and Y. Wang, *ibid.* **83**, 165440 (2011); H.-T. He, G. Wang, T. Zhang, I.-K. Sou, G. K. L. Wong, and J.-N. Wang, Phys. Rev. Lett. **106**, 166805 (2011).

[17] J. Chen, H. J. Qin, F. Yang, J. Liu, T. Guan, F. M. Qu, G. H. Zhang, J. R. Shi, X. C. Xie, C. L. Yang, K. H. Wu, Y. Q. Li, and L. Lu, Phys. Rev. Lett. **105**, 176602 (2010).

[18] G. Bergmann, Phys. Rep. **107**, 1 (1984).

[19] C. W. J. Beenakker and H. van Houten, *Quantum Transport in Semiconductor Nanostructures* (Academic, San Diego, 1991).

[20] G. Bergmann, Solid State Commun. **42**, 815 (1982).

[21] E. McCann, K. Kechedzhi, V. I. Fal'ko, H. Suzuura, T. Ando, and B. L. Altshuler, Phys. Rev. Lett. **97**, 146805 (2006).

[22] J. Chen, X. Y. He, K. H. Wu, Z. Q. Ji, L. Lu, J. R. Shi, J. H. Smet, and Y. Q. Li, Phys. Rev. B **83**, 241304 (2011).

[23] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.84.233101 for: (i) Details on material growth and device fabrication; (ii) Discussion of the applicability of the strong spin-orbit limit formula; (iii) WAL in the multi-channel limit.

[24] S. Hikami, A. I. Larkin, and Y. Nagaoka, Prog. Theor. Phys. **63**, 707 (1980).

[25] S. R. Park, W. S. Jung, C. Kim, D. J. Song, C. Kim, S. Kimura, K. D. Lee, and N. Hur, Phys. Rev. B **81**, 041405 (2010).