Helical Edge Resistance Introduced by Charge Puddles

Jukka I. Väyrynen, Moshe Goldstein, and Leonid I. Glazman

Phys. Rev. Lett. **110**, 216402 — Published 21 May 2013

DOI: 10.1103/PhysRevLett.110.216402

Helical edge resistance introduced by charge puddles

Jukka I. Väyrynen, Moshe Goldstein, and Leonid I. Glazman
Department of Physics, Yale University, New Haven, CT 06520, USA
(Dated: March 12, 2013)

We study the influence of electron puddles created by doping of a 2D topological insulator on its helical edge conductance. A single puddle is modeled by a quantum dot tunnel-coupled to the helical edge. It may lead to significant inelastic backscattering within the edge because of the long electron dwelling time in the dot. We find the resulting correction to the perfect edge conductance. Generalizing to multiple puddles, we assess the dependence of the helical edge resistance on temperature and doping level, and compare it with recent experimental data.

PACS numbers: 71.10.Pm,73.63.Kv

The realization that a boundary separating a topologically-nontrivial insulator from a conventional one should carry delocalized electron states [1, 2] has led to the prediction of such states in concrete materials and their experimental observation [3–5]. One of the stunning theoretical predictions is that in 2D the zero-temperature electron transport along an edge is reflectionless, as long as time-reversal symmetry is not broken, which should lead to the quantization of the edge conductance [2]. Experiments with HgTe quantum wells of the appropriate thickness confirmed the existence of highly-conducting channels in a nominally insulating state of a heterostructure [6–8]. The Fermi energy E_F in a heterostructure was tuned by a gate to reside in the gap between the valence and conduction bands. The values of the conductance G measured under these conditions were indeed close to the predicted quantized value $G_0 = e^2/h$ per edge, but only for small $\sim 1\times 1\,\mu\text{m}^2$ samples. Deviations $\Delta G \equiv G_0 - G$ towards lower conductance values were clearly seen in larger samples [6–11].

In short samples, ΔG fluctuated with gate voltage. The temperature dependence of G has not been systematically measured yet, but the existing data indicate it to be rather weak. These observations should be contrasted with the theoretical predictions of a strong T-dependence of electron inelastic backscattering rate, with a characteristic scale set by the band gap E_G. Depending on the model, ΔG scales as $\propto (T/E_G)^6$ or $(T/E_G)^4$, unless time-reversal symmetry is broken [2, 12–14]. Spontaneous symmetry breaking is improbable for weak electron-electron interaction (noting the high dielectric constant, $\kappa \approx 13$ [5]) and in the absence of a high density of magnetic impurities. Luttinger liquid effects [2, 12, 14, 15] are also suppressed in short samples.

The existing theory considers inelastic electron backscattering by either uniform interactions along an edge [2, 15, 16], or at isolated points [2, 12, 13, 17]. Helical edges formed in a semiconductor heterostructure are likely to deviate considerably from either limit. The structures are doped [6–8, 10, 11]; the presence of charged donors and acceptors results in a non-uniform potential landscape for electrons. These inhomogeneities are not point-like because of the long-range of the Coulomb potential. Moreover, the topologically non-trivial insulators are in fact narrow-gap semiconductors with a typical gap of only $E_G \simeq 10\,\text{meV}$ for HgTe quantum wells [6–8, 10, 11, 18]. To place E_F inside the band gap, an appropriate gate voltage is applied, so that the gate charge balances out the uncompensated donor (n_d) or acceptor (n_a) charge density. The joint effect of the gate and ionized dopant atoms may lead to the accidental formation of electron and hole puddles in the quantum well, cf. Fig. 1, similar to the known phenomenon in compensated bulk semiconductors [19]. In the existing measurements doping varied from $n_d \sim 3.5\cdot 10^{11}\,\text{cm}^{-2}$ to $n_a \sim 5\cdot 10^{10}\,\text{cm}^{-2}$ [6–8, 10, 11], and the results indeed seem to indicate that a lower doping level improves the quality of the edge conductance quantization. Furthermore, the uncovered strong sensitivity of the edge conductance to the potential of a scanning probe [10] may imply the presence of puddles, i.e., spontaneously formed quantum dots, in the vicinity of the edge.

In this Letter we elucidate the role of tunneling between an edge and a quantum dot on the edge conductance. Elastic processes involving electron dwelling in

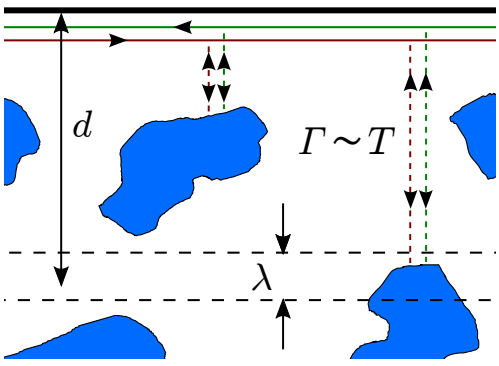

FIG. 1. (Color online) Electrons moving along a helical edge tunnel in and out of puddles created by the inhomogeneous charge distribution in the heterostructure. In the puddles electrons may undergo inelastic backscattering. The main contribution comes from puddles whose distance d from the edge is within a strip where the resulting level width $\Gamma \sim T$, cf. Eqs. (7)–(8). The strip width is the tunneling length $\lambda = v/E_G$. Summation over the puddles yields the average resistivity, Eq. (12).

the dot do not lead to any backscattering. However, dwelling enhances the inelastic backscattering by increasing the time electrons interact with each other. At $T < \delta$, the dwelling-time effect makes the conductance correction strongly dependent on the position of the Fermi level E_F with respect to the dot energy levels, and on the tunneling widths Γ of these levels ($\delta \ll E_G$ is the mean level spacing in the dot). At a given temperature T, the tallest peaks $\Delta G^{\rm peak} \propto (T/\delta)^2$ in $\Delta G(E_F)$ are produced by levels with elastic widths $\Gamma \sim T$, see Eq. (8) and Fig. 1. Such peaks in $\Delta G(E_F)$ are of widths $\sim T$, and the "peak-to-valley" ratio is $\sim (\delta/T)^6$.

Dots, or puddles of charge carriers in a quantum well, are formed by fluctuations in the donor density [21, 22]. We establish a crossover value n_0 of n_d below which puddles are rare. At $n_d \ll n_0$ the density of puddles, n_p, is exponentially small in n_0/n_d. In short samples of length $L \lesssim n_p^{-1/2}$ only a few puddles are in the vicinity of the edge, resulting in mesoscopic fluctuations of G with the gate voltage. This model agrees with the results of scanning-gate experiments [10] and could explain the variations of G with back gate voltage in earlier experiments [6–8], if the condition $n_d \lesssim n_0$ would hold there. For longer samples, $L \gg n_p^{-1/2}$, many puddles couple effectively to the edge. That leads to edge resistance, $R \propto n_p L (T/\delta)^3$, which varies smoothly with the gate voltage and possibly greatly exceeds the quantized value h/e^2. At the same time, the "bulk" hopping conductivity, which is proportional to factors exponentially small in $n_p^{-1/2}$ and T/δ, may still remain negligible. In this case, current would flow along the edges, despite edge resistance being high compared to h/e^2, as observed in Ref. [11]. The model would also explain the earlier measurements [6–9] on larger samples, if the condition $n_d \lesssim n_0$ would be satisfied [our crude estimate of n_0, Eq. (10), turns out to be too low for that].

We start by considering a helical edge coupled to a single quantum dot via a point contact. In the absence of interactions, the corresponding Hamiltonian takes form:

$$\hat{H}_0 = -iv_F \sum_\gamma \gamma \int dx \psi_\gamma^\dagger(x) \partial_x \psi_\gamma(x) + \sum_{n\gamma} \varepsilon_n c_{n\gamma}^\dagger c_{n\gamma} + \sum_{n,\gamma} t_n c_{n\gamma}^\dagger \psi_\gamma(0) + \text{H.c.} \quad (1)$$

Here v_F is the helical edge velocity, $\gamma = \pm 1 \equiv R, L$ labels the right- and left-movers, respectively, and ε_n are the discrete energy levels in the dot, measured from E_F. The dot is coupled to the edge at $x = 0$ by tunneling amplitudes t_n, leading to elastic level widths $\Gamma_n = |t_n|^2/(2v_F)$. The Kramers degeneracy of each level gave us the freedom to pick the eigenfunctions $|n\gamma\rangle$ such that the right- and left-movers are coupled to two orthogonal components of each doublet. There is thus no backscattering in the free-electron problem. Interaction in the dot,

$$\hat{U} = \frac{1}{2} \sum_{n_i,\gamma_i} U_{n_1\gamma_1 n_2\gamma_2; n_3\gamma_3 n_4\gamma_4} c_{n_1\gamma_1}^\dagger c_{n_2\gamma_2}^\dagger c_{n_4\gamma_4} c_{n_3\gamma_3}, \quad (2)$$

may lead to inelastic backscattering (hereinafter, we assume $\hat{U}$ respects time-reversal symmetry).

Inelastic backscattering reduces the steady-state current $I = I_0 - \Delta I$ from its ideal value $I_0 = G_0 V$ by

$$\Delta I = e \sum_{\gamma_i} \Delta N_{\gamma_1\gamma_2;\gamma_3\gamma_4} \int dE_1 dE_2 dE_3 dE_4 \times S_{\gamma_1\gamma_2;\gamma_3\gamma_4}(E_1, E_2; E_3, E_4) \delta(E_1 + E_2 - E_3 - E_4) \times \left[\tilde{f}_{\gamma_1}(E_1) \tilde{f}_{\gamma_2}(E_2) (1 - \tilde{f}_{\gamma_3}(E_3))(1 - \tilde{f}_{\gamma_4}(E_4)) - \tilde{f}_{\gamma_3}(E_3) \tilde{f}_{\gamma_4}(E_4) (1 - \tilde{f}_{\gamma_1}(E_1))(1 - \tilde{f}_{\gamma_2}(E_2)) \right]. \quad (3)$$

Here V is the source-drain voltage, $\Delta N_{\gamma_1\gamma_2;\gamma_3\gamma_4} = (\gamma_1 + \gamma_2 - \gamma_3 - \gamma_4)/2$ counts the net number of right-movers scattered into left-movers, $\tilde{f}_{\gamma_i}(E) = 1/[e^{(E+\gamma_i eV/2)/T} + 1]$ is the Fermi function shifted by $\pm eV/2$, and $S_{\gamma_1\gamma_2;\gamma_3\gamma_4}(E_1, E_2; E_3, E_4)$ is the cross section for the two-electron scattering process $|E_1\gamma_1, E_2\gamma_2\rangle \to |E_3\gamma_3, E_4\gamma_4\rangle$ between exact right- and left-propagating eigenstates of the Hamiltonian (1). In general, S allows backscattering of one ($RR \to LR$) or two ($RR \to LL$) electrons. There are two respective contributions, ΔG_1 and ΔG_2, to the conductance $G = G_0 - \Delta G_1 - \Delta G_2$.

In the Born approximation the cross section is

$$S_{\gamma_1\gamma_2;\gamma_3\gamma_4}(E_1, E_2; E_3, E_4) = \frac{2}{\pi^3} \sum_{m_i,n_i} \left[\prod_{i=1}^4 \text{Im} \mathcal{G}_{n_i m_i}^R(E_i) \right] \times U_{m_1\gamma_1 m_2\gamma_2; m_3\gamma_3 m_4\gamma_4}^* U_{n_1\gamma_1 n_2\gamma_2; n_3\gamma_3 n_4\gamma_4}. \quad (4)$$

Here $\mathcal{G}_{n_1 n_2}^R(E)$ is the noninteracting retarded Green function of an electron in the dot. All interaction matrix elements must be small compared to the mean level width Γ to allow the perturbative treatment at arbitrary position of the Fermi level with respect to the dot levels. This condition is more easily satisfied for the off-diagonal matrix elements [20] appearing explicitly in Eq. (4), than for the diagonal ones $U_{n_1\gamma_1 n_2\gamma_2; n_1\gamma_1 n_2\gamma_2} \sim E_C$. The introduced charging energy E_C is small, $E_C \ll \delta$, if the spacer between the quantum well and gate is thinner than the Debye radius for electrons in the well. In the opposite case of $E_C \gtrsim \delta$, Coulomb blockade may develop. We first treat the entire interaction perturbatively, and later point out how Coulomb blockade modifies the results. We will also see that backscattering is dominated by puddles with $\Gamma \sim T$; thus Kondo correlations [17] setting in only at the temperature $T_K \ll \Gamma$ can be ignored.

Using properties of the interaction matrix elements in Eq. (4), it is straightforward to check that in the low-temperature limit $\Delta G_1 \propto T^4$ and $\Delta G_2 \propto T^6$, in agreement with Refs. [2, 12–14]. For a generic form of strong spin-orbit interaction in the dot, all interaction matrix elements in Eq. (4) are of the same order [20]. In this case, $\Delta G_2/\Delta G_1 \ll 1$ if $T \ll \delta$. The proportionality coefficient of the temperature dependence $\Delta G_1 \propto T^4$ is a function of the dot parameters; in the case of weak tunneling it peaks every time a level crosses the Fermi energy.

Weak tunneling corresponds to $\Gamma_n \ll |\varepsilon_n - \varepsilon_{n\pm1}|$. Then the leading-order approximations for the diagonal

and off-diagonal ($n_1 \neq n_2$) matrix elements of $\hat{\mathcal{G}}^R(E)$ read $\mathcal{G}^R_{nn}(E) = (E - \varepsilon_n + i\Gamma_n)^{-1}$ and $\mathcal{G}^R_{n_1n_2}(E) = -i\sqrt{\Gamma_{n_1}\Gamma_{n_2}}[(E-\varepsilon_{n_1}+i\Gamma_{n_1})(E-\varepsilon_{n_2}+i\Gamma_{n_2})]^{-1}$, respectively. Using this simplification in Eq. (4) we find

$$\frac{\Delta G_1^{\rm peak}}{G_0} = \frac{2^7\pi}{15}\left(\frac{T}{\Gamma_1}\right)^4 \left|\sum_{n\neq 1}\sqrt{\frac{\Gamma_n}{\Gamma_1}}\cdot\frac{U_{1L1R;1RnR}}{\varepsilon_n}\right|^2 \quad (5)$$

for the peak in ΔG_1 corresponding to the level ε_1 crossing the Fermi level ($\varepsilon_1 = 0$). The peak height and its width, $|\varepsilon_1 - E_F| \sim \Gamma_1$, display mesoscopic fluctuations; Eq. (5) is applicable at $T \ll \Gamma_1$. The peak value $\Delta G_1^{\rm peak}$ grows with temperature till T reaches a value $T \sim \Gamma_1$. At higher temperatures, some of the incoming electrons with energies $|E| \lesssim T$ which contribute to ΔG_1 are off resonance. This leads to a decreasing T-dependence of $\Delta G_1^{\rm peak}$ at $T \gtrsim \Gamma_1$,

$$\frac{\Delta G_1^{\rm peak}}{G_0} = \frac{\Gamma_1}{T}\left|\sum_{n\neq 1}\sqrt{\frac{\Gamma_n}{\Gamma_1}}\cdot\frac{U_{1L1R;1RnR}}{\varepsilon_n}\right|^2, \quad (6)$$

and a peak width $|\varepsilon_1 - E_F| \sim T$.

In a weakly-disordered dot the Thouless energy $E_T = g\delta \gg \delta$ ($g \gg 1$ is the dimensionless conductance within the dot). The disorder-averaged matrix elements $\langle U^2\rangle$ of interaction present in Eqs. (5) and (6) can be evaluated using the standard diagrammatic techniques [20]. Further simplification is possible for the screened Coulomb interaction, which is dominated by its universal zero-momentum component, leading to $\langle U^2\rangle \sim \delta^2/g^2$. Using this estimate in Eqs. (5)–(6) and dropping numerical factors, we arrive at the interpolation

$$\frac{\langle\Delta G_1^{\rm peak}\rangle}{G_0} \sim \frac{1}{g^2}\frac{T^4}{\Gamma^4}\theta(\Gamma - T) + \frac{1}{g^2}\frac{\Gamma}{T}\theta(T-\Gamma) \quad (7)$$

for the typical peak conductance as a function of T at small charging energy, $E_C \ll \max\{T, \Gamma\}$.

The backscattering processes leading to Eqs. (5) and (6) involve a sequence of virtual states. Those with energy deficit $|\varepsilon_n| \neq 0$ are represented by the denominators in the sums over $n \neq 1$. One of the virtual states, however, has two electrons on level $n = 1$ and brings a large factor $\sim 1/\Gamma^2$ to Eqs. (5) and (6). It is replaced by $1/E_C^2$ in the presence of charging energy $E_C \gg \Gamma$. For the same reason, the cross section Eq. (4) loses sensitivity to the energy E_i of one of the two electrons involved. The corresponding integration range in Eq. (3) is then restricted by T rather than Γ at any T/Γ. In the important (see below) case $E_C \sim \delta$, the two modifications change Eq. (7) by a factor $\sim (\Gamma/\delta)^2 \cdot \max\{1, T/\Gamma\}$, leading to:

$$\frac{\langle\Delta G_1^{\rm peak}\rangle}{G_0} \sim \frac{1}{g^2}\frac{T^4}{\Gamma^2\delta^2}\theta(\Gamma - T) + \frac{1}{g^2}\frac{\Gamma^2}{\delta^2}\theta(T-\Gamma). \quad (8)$$

Backscattering in the "valley" (Fermi level in between two subsequent dot levels) regime does not involve any low-energy virtual state and is not affected qualitatively by $E_C \sim \delta$. The corresponding estimate, $\langle\Delta G_1^{\rm valley}\rangle/G_0 \sim T^4\Gamma^4/g^2\delta^8$ is smaller than the peak value Eq. (8) by a factor $\sim (\Gamma^2/\delta^6)\cdot\max\{\Gamma^4, T^4\}$.

The main contribution to the backscattering correction averaged over the position of the Fermi level comes from the peak values, Eq. (8), as $\langle\Delta G_1^{\rm valley}\rangle/G_0$ is parametrically smaller. Accounting for the peak widths, $|\varepsilon_i - E_F| \sim \max\{\Gamma, T\}$, we find

$$\frac{\langle\Delta G_1^{\rm av}\rangle}{G_0} \sim \frac{1}{g^2}\frac{T^4}{\Gamma\delta^3}\theta(\Gamma - T) + \frac{1}{g^2}\frac{\Gamma^2 T}{\delta^3}\theta(T-\Gamma). \quad (9)$$

At higher temperatures the above mechanism gives way to thermally-activated backscattering processes. Those originate from the diagonal elements $\mathcal{G}^R_{nn}(E)$ in Eq. (4) only. Since this regime is probably not relevant for the interpretation of existing experiments (see below) we only sketch the results, deferring a detailed discussion [23]. There are two types of activated contributions to ΔG. The first one involves transitions within a pair of levels, $\{n_3, n_4\} = \{n_1, n_2\}$. The other one involves more levels, $\{n_3, n_4\} \neq \{n_1, n_2\}$, and gains importance at higher temperatures ($T \gg \delta$) due to the larger phase space available for transitions. At $T \lesssim \delta$, backscattering is dominated by the two levels closest to E_F, and $\Delta G \sim (\delta^2/g^2\Gamma T)\cdot\exp(-\varepsilon/T)$ with $\varepsilon \sim \delta$. Comparison with Eq. (9) shows that activated backscattering exceeds $\langle\Delta G_1^{\rm peak}\rangle$ at $T \gtrsim \delta/\ln(\delta/\Gamma)$. The distinction between peaks and valleys is lost at these temperatures, although ΔG does experience strong mesoscopic fluctuations at $T \sim \delta$ due to the randomness of the activation energy ε.

Now we turn to the typical experimental case [6–9, 18] of a doped, gate-controlled heterostructure. Donors of density n_d are confined to a plane between the quantum well and the gate, and separated from the two by distances ℓ_d and $\ell_g - \ell_d$, respectively. In the absence of carriers, the random potential [21] of donors $V(\mathbf{r})$ has variance $\langle V^2\rangle = V_0^2\ln\{\ell_g^2/[(2\ell_g - \ell_d)\ell_d]\}$ with $V_0 = \sqrt{2\pi n_d}e^2/\kappa$ (κ is the dielectric constant). At the point of full depletion (the gate charge density is $-en_d$) the probability of creation of electron and hole puddles depends on the ratio $E_G/(2\sqrt{\langle V^2\rangle})$. The relation $\sqrt{\langle V^2\rangle} = E_G/2$ defines a characteristic donor density,

$$n_0 = \frac{E_G^2\kappa^2}{8\pi e^4\ln\{\ell_g^2/[(2\ell_g - \ell_d)\ell_d]\}}. \quad (10)$$

At $n_d \ll n_0$, the carrier puddles are small and rare. In the opposite limit ($n_d \gg n_0$), puddles are large and separated by p-n junctions of typical [24] thickness $E_G/|\nabla V(\mathbf{r})|$. The junctions are *thin* compared to the electron penetration depth $\lambda \sim v/E_G$ (here v is the electron velocity at $E_F \gg E_G/2$) and therefore transparent, $\Gamma \gtrsim \delta$; the corresponding average bulk conductivity $\sigma_{\rm bulk} \gtrsim e^2/h$. Analysis [25] suggests a transition from the topological insulator to conductor state at $\sigma_{\rm bulk} \approx (1.4 - 2.5)e^2/h$. This makes the limit $n_d \gg n_0$ unfavorable for the helical edge conductance quantization at any temperature.

In the opposite limit, $n_d \ll n_0$, the puddles are sparse if the electron chemical potential is in the band gap, away by $\gtrsim V_0$ from the band edges [21]. At $\alpha \ll 1$ [$\alpha = e^2/(\kappa\hbar v)$ is the interaction parameter], we describe a puddle in the Thomas-Fermi approximation. The puddle size w is found by matching its number of electrons N and electrochemical potential $N/(2m^*w^2)$ to the typical impurity charge fluctuation $N \sim w n_d^{1/2}$ and potential fluctuation V_0, respectively [$m^* = E_G/2v^2$ is the effective electron mass]. This leads to $w \sim a_B$ and

$$E_C \sim \delta \sim \alpha^2 E_G\,;\; g \sim (n_d/n_0)^{1/4}/(2\pi)^{1/4}\alpha\,. \qquad (11)$$

Here $a_B = 2\hbar v/\alpha E_G$ is the effective Bohr radius; we used the estimates $\delta \sim V_0/N$, $g \sim \sqrt{N}$ [20], and $E_C \sim e^2/(\kappa w)$ [assuming $w \lesssim \ell_g$]. We also assume $n_d a_B^2 \gg 1$, justifying the Thomas-Fermi approximation [21]. Puddles are located at rare strong fluctuations of the potential, and are thus far away from each other. Their density n_p is estimated as the ratio of the total carrier density to the number $N \sim a_B n_d^{1/2}$ of electrons in a puddle. To find the former quantity we note that the distance to the gate ℓ_g serves as screening length for the potential fluctuations; hence we can divide the sample into roughly independent regions of size ℓ_g. A region becomes populated by carriers only if the local potential experiences an exponentially rare fluctuation exceeding $E_G/2$. The carrier number is such that they compensate for the fluctuation [21]. This leads to $n_p \sim 1/(\ell_g a_B)(n_d/n_0)^{1/2} e^{-n_0/n_d}$.

The hopping conductivity facilitated by the puddles is proportional to a product of two small parameters: the tunneling probability, exponential in $-(\lambda^2 n_p)^{-1/2}$; and the thermal activation probability, exponential in $-\delta/T$. The latter one remains small at $T \ll \delta$, even when approaching the crossover region $n_d \lesssim n_0$. Under the same conditions, the rate of backscattering into the helical edge scales as a relatively low power of T/δ, cf. Eqs. (8)–(9). For samples of length $L \lesssim n_p^{-1/2}$ only a few puddles occur in the vicinity of the edge. That would make ΔG sensitive to a local probe potential, consistent with recent scanning gate measurements [10], and may also provide an explanation for the mesoscopic fluctuations of $\Delta G(E_F)$ in earlier measurements [6–8, 10]. Eq. (8) predicts that the largest peaks in $\Delta G(E_F)$ at a given T scale as $(T/\delta)^2$ and are produced by levels with $\Gamma \sim T$. For $L \gg n_p^{-1/2}$ contributions of many puddles add up incoherently, as scattering off each puddle is inelastic. The exponential dependence of the level widths $\Gamma \propto e^{-2d/\lambda}$ on the distance d between a puddle and the edge leads to a broad distribution of Γ. Hence, by Eq. (9), backscattering will be dominated by puddles whose distance from the edge is such that $\Gamma \sim T$. Summing over puddles (see Fig. 1), we find that a long edge displays resistance $R = \rho_{\rm edge} L$ with self-averaging resistivity

$$\rho_{\rm edge} \sim \frac{1}{G_0}\frac{1}{g^2} n_p \lambda \left(\frac{T}{\delta}\right)^3. \qquad (12)$$

While $\rho_{\rm edge} \propto (T/\delta)^3$, the quantum well hopping conductivity $\sigma_{\rm bulk}$ is exponentially small in δ/T. Therefore, leakage into the bulk at $n_d \lesssim n_0$ is insignificant for samples shorter than the exponentially-large "leakage length" $L^* = 1/(\sigma_{\rm bulk}\rho_{\rm edge})$, which may explain recent scanning SQUID results [11].

Our findings thus match with observations provided that $n_d \lesssim n_0$ and $T \ll \delta$. To estimate δ for a HgTe/CdTe heterostructure, we use Eq. (11) with $E_G = 10$meV and $\alpha \approx 0.32$ (found with $\kappa \approx 13$ and $v = 5.5\cdot 10^7$cm/sec [5]). We arrive at $\delta \approx 1$meV, comfortably above $k_B T$ in most experiments. For the crossover density, Eq. (10), we find $n_0 \approx 3\cdot 10^{10}$cm^{-2}. The doping levels reported in Ref. [10] and in [6–8] are, respectively, moderate ($n_a/n_0 \sim 1$) and high ($n_d/n_0 \sim 10$) with respect to this value. On the other hand, from the total resistance of long samples in Ref. [6] we deduce that $\sigma_{\rm bulk} \lesssim 0.45 G_0$, consistent with an insulating bulk [25]. It may mean that our crude estimate of n_0 is off by a factor of 10. The characteristic length $n_p^{-1/2}$, separating mesoscopic samples from the "self-averaging" ones, provides another check. The pre-exponential factor in it, ~ 100nm, is only ~ 10 times shorter than 1μm-long "mesoscopic" samples in Refs. [6–8]. That too may indicate that the samples doping was close to the true crossover value n_0.

To conclude, disorder in a doped heterostructure may lead to appreciable backscattering within a helical edge, while hopping conductivity in the quantum well remains negligible. It may explain some of the experiments [6–8, 10, 11]. The samples doping level n_d was too close to the crossover value Eq. (10) to allow a reliable analysis of the edge resistance dependence on n_d. On the other hand, robust qualitative features of the resistance T-dependence, in both the mesoscopic and self-averaging regimes, Eqs. (8)–(9) and (12), make its detailed measurement very desirable.

We thank D. Goldhaber-Gordon, K. Moler, and K. Nowack for stimulating discussions, and C. Varma for his request to write for the Journal Club for Condensed Matter Physics, which partially motivated this study. This work was supported by NSF DMR Grant No. 1206612, the Simons Foundation, and the Bikura (FIRST) program of the Israel Academy of Science.

[1] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).
[2] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).
[3] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, Science **314**, 1757 (2006).
[4] C. L. Kane and M. Z. Hasan, Rev. Mod. Phys. **82**, 3045 (2010).

[5] X.-L Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).
[6] M. König *et al.*, Science **318**, 766 (2007).
[7] M. König *et al.*, J. Phys. Soc. Jpn. **77**, 031007 (2008).
[8] A. Roth *et al.*, Science **325**, 294 (2009).
[9] G. M. Gusev *et al.*, Phys. Rev. B **84**, 121302(R) (2011).
[10] M. König *et al.*, arXiv:1211.3917.
[11] K. C. Nowack *et al.*, arXiv:1212.2203.
[12] C. Xu and J. E. Moore, Phys. Rev. B **73**, 045322 (2006).
[13] T. L. Schmidt, S. Rachel, F. von Oppen, and L. I. Glazman, Phys. Rev. Lett. **108**, 156402 (2012).
[14] N. Lezmy, Y. Oreg, and M. Berkooz, Phys. Rev. B **85**, 235304 (2012).
[15] Anders Ström, Henrik Johannesson, and G. I. Japaridze, Phys. Rev. Lett. **104**, 256804 (2010).
[16] J. C. Budich, F. Dolcini, P. Recher, and B. Trauzettel, Phys. Rev. Lett. **108**, 086602 (2012).
[17] J. Maciejko *et al.*, Phys. Rev. Lett. **102**, 256803 (2009); Y. Tanaka, A. Furusaki, and K. A. Matveev, Phys. Rev. Lett. **106**, 236402 (2011).
[18] E. G. Novik *et al.*, Phys. Rev. B **72**, 035321 (2005).
[19] A. L. Efros, and B. I. Shklovskii, *Electronic Properties of Doped Semiconductors* (Springer-Verlag, New York, 1984).
[20] I. L. Aleiner, P. W. Brouwer, and L. I. Glazman, Phys. Rep. **358**, 309 (2002).
[21] V. A. Gergel' and R. A. Suris, Zh. Eksp. Teor. Fiz. **75**, 191 (1978) [Sov. Phys. JETP **48**, 95 (1978)].
[22] A. L. Efros, F. G. Pikus, and V. G. Burnett, Phys. Rev. B **47**, 2233 (1993).
[23] J. I. Väyrynen, M. Goldstein, and L. I. Glazman, in preparation.
[24] B. Skinner and B. I. Shklovskii, arXiv:1212.6653; we find typical $|\nabla V(\mathbf{r})|$ from the correlation function $\langle V(\mathbf{r})V(\mathbf{0})\rangle$ at the scale $|\mathbf{r}| \sim \lambda$, different from the limit considered there.
[25] L. Fu and C.L. Kane, Phys. Rev. Lett. **109**, 246605 (2012), and references cited therein.