Kondo Effect in the Helical Edge Liquid of the Quantum Spin Hall State

Joseph Maciejko,1 Chaoxing Liu,1,2 Yuval Oreg,3 Xiao-Liang Qi,1 Congjun Wu,4 and Shou-Cheng Zhang1
1Department of Physics, McCallough Building, Stanford University, Stanford, California 94305-4045, USA
2Center for Advanced Study, Tsinghua University, Beijing, 100084, China
3Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel
4Department of Physics, University of California, San Diego, California 92093, USA

(Received 13 January 2009; published 26 June 2009)

Following the recent observation of the quantum spin Hall (QSH) effect in HgTe quantum wells, an important issue is to understand the effect of impurities on transport in the QSH regime. Using linear response and renormalization group methods, we calculate the edge conductance of a QSH insulator as a function of temperature in the presence of a magnetic impurity. At high temperatures, Kondo and/or two-particle backscattering give rise to a logarithmic temperature dependence. At low temperatures, for weak Coulomb interactions in the edge liquid, the conductance is restored to unitarity with unusual power laws characteristic of a “local helical liquid,” while for strong interactions, transport proceeds by weak tunneling through the impurity where only half an electron charge is transferred in each tunneling event.

FIG. 1 (color online). Temperature dependence of the conductance $G$ in a Hall bar measurement is approximately quantized to $G_0 = 2e^2/h$, independent of temperature, for samples of about a micron length [2,7]. However, larger samples exhibit $G < G_0$ and $G$ decreases with decreasing temperature [7]. Deviations from the expected quantized value have been attributed to local doped regions due to potential inhomogeneities within the sample arising from impurities or roughness of the well-barrier interface [7]. Although pure potential scattering cannot backscatter the edge states, the role of these potential inhomogeneities is to trap bulk electrons which may then interact with the edge electrons. These localized regions act as dephasing centers for the edge channels due to interaction effects and may cause backscattering.

In this Letter, we study the temperature dependence of the edge conductance $G$ of a QSH insulator. We consider the case where a local doped region in the vicinity of the edge traps an odd number of electrons and acts as a magnetic impurity coupled to the HL. Our main results are (Fig. 1): (1) At high temperatures, $G$ is logarithmic, $-\Delta G \equiv -(G - G_0) = \eta + \gamma \ln(D/T)$ where $\eta$, $\gamma$ are interaction-dependent parameters and $D$ is an energy scale on the order of the bulk gap. (2) For weak Coulomb interactions $K > 1/4$ where $K$ is the Luttinger parameter of the HL, $G$ is restored to unitarity at $T = 0$ due to the formation of a Kondo singlet. This is in stark contrast with the Kondo problem in a usual spinful 1D liquid where $G$ vanishes at $T = 0$ for all $K_\rho < 1$ where $K_\rho$ is the Luttinger parameter in the charge sector [8]. At low but finite $T$, $G$ decreases as an unusual power law $\Delta G \propto -T^{2(4K-1)}$ due to correlated two-particle backscattering (2PB). The edge liquid being helical, the decrease in $G$ is a direct measure of the spin-flip rate [9]. (3) For strong Coulomb interactions $K < 1/4$, 2PB processes are relevant and the system is insulating at $T = 0$. At low but finite $T$, $G$ is restored by...
tunneling of excitations with fractional charge $e/2$, and we obtain $G(T) \propto T^{(1/4K-1)}$.

**Model.**—We model the impurity by a $S = \frac{1}{2}$ local spin coupled by exchange to the 1D HL with Coulomb interactions. The HL having the same number of degrees of freedom as a spinless fermion, a single nonchiral boson $\phi$ is sufficient to bosonize the HL [5]. The system is described by the Hamiltonian $H = H_0 + H_K + H_Z$, where $H_0$ is the usual Tomonaga-Luttinger Hamiltonian $H_0 = \frac{\pi}{8} \times \int dx (K_{1}|^2 + \frac{\chi}{2}(\partial_x \phi)^2)$, with $K$ the Luttinger parameter and $\nu$ the edge state velocity. The Kondo Hamiltonian $H_K$ has the form

$$H_K = \frac{J\nu}{2\pi\xi} \langle S_+ : e^{-\nu/S}\phi(0) : + \text{H.c.} \rangle - \frac{J\nu}{\sqrt{\pi}} S_+ \Pi(0) \ .$$

where $S_+ = S_x \pm iS_y$ and $S_+$ are the spin operators for the impurity localized at $x = 0$. $a$ is the lattice constant of the underlying 2D lattice and corresponds to the size of the impurity (we assume that the impurity occupies a single lattice site). $\xi$, the penetration length of the helical edge states into the bulk, acts as a short-distance cutoff for the 1D continuum theory in the same way that the magnetic length, the penetration length of the chiral edge states, acts as a short-distance cutoff for the chiral Luttinger liquid theory of the quantum Hall edge [10]. In addition to Kondo scattering, 2PB is allowed by TRS [5,6]. In HgTe QW, the wave vector where the edge dispersion enters the bulk is usually smaller than $\pi/2a$ such that the uniform 2PB (unklapp) term requiring $4k_F = 2\pi/a$ can be ignored. The impurity can however provide a $4k_F$ momentum transfer, and we must generally also consider a local impurity-induced 2PB term [11].

**Weak coupling regime.**—We first consider the weak coupling regime where $J_||, J_z, A_2$ are small parameters. We first calculate $G$ to second order in the bare couplings $J_0$ and $A_2$ using the Kubo formula. This result is then extended to include all leading logarithmic terms in the perturbation expansion by means of a weak coupling renormalization group (RG) analysis of the scale-dependent couplings $J_||(T)$ and $A_2(T)$ where the scale is set by the temperature $T$.

The $J_z$ term can be removed from the Hamiltonian by a unitary transformation [12] of the form $U = e^{i\lambda S_z \phi(0)}$ with $\lambda = -J_0 a/\nu K\sqrt{\pi}$, which changes the scaling dimension of the vertex operator: $e^{i\sqrt{\nu/S}\phi(0)}$. The transformed Hamiltonian is $\tilde{H} = UHU^{-1} = H_0 + H_2 + \frac{J\nu}{2\pi\xi} \langle S_+ : e^{-i\sqrt{\nu/S}\phi(0)} : + \text{H.c.} \rangle$, where $\chi = 1 - \nu J_z/2K$ with $\nu = a/\nu v$ the density of states of the HL. The scaling dimension of $e^{i\sqrt{\nu/S}\phi(0)}$ is $\tilde{K} = K\chi^2$.

Using $\tilde{H}$, we calculate $\Delta G(T) = \Delta G_K(T) + \Delta G_2(T)$ to second order in $J_0$ and $A_2$ using the Kubo formula, where $\Delta G_K$ is the correction due to Kondo scattering [13], $\Delta G_2$ is the correction due to 2PB, and $\Delta G_3$ is the correction due to Kondo scattering [13].

$$\Delta G_K = \frac{\Delta G_2}{\nu K} = \frac{\nu J_0}{\nu K} \ .$$

where $\nu J_0 = \nu J_0(T/D)^{1-K}$ to $O(\nu)$ and $D = \nu v/\nu \xi$ is a high-energy cutoff on the order of the bulk gap. $\Delta G_2$ is the correction due to 2PB,

$$\Delta G_2 = \frac{\Delta G_4}{\nu K} = \frac{\nu J_0}{\nu K} \ .$$

where $\nu J_0 = \nu J_0(T/D)^{1-K}$.

These results can be complemented by a RG analysis. The RG equation for $A_2$ follows by dimensional analysis $dA_2/d\ell = (1 - 4K)A_2$ so that $A_2$ is relevant for $K < 1/4$ and irrelevant for $K > 1/4$. The renormalized coupling is $A_2(T) = A_2(T/D)^{1-K}$, and second order renormalized perturbation theory $\Delta G_2(T) \propto -A_2(T)^2$ simply reproduces the Kubo formula result (3). Perturbation theory fails for $T \lesssim T_2$ where $T_2^* = (\nu J_0)^{1/(1-K)}$ is a scale for the crossover from weak to strong 2PB with $\lambda_2^2$ the bare 2PB amplitude. The 1-loop RG equations [5,15] for the Kondo couplings $J_0, J_z$ read

$$\frac{dJ_0}{d\ell} = (1 - K)J_0 - \nu J_0 J_z, \quad \frac{dJ_z}{d\ell} = \nu J_0 J_z \ ,$$

The family of RG trajectories is indexed by a single scaling invariant $c = (\nu J_0^2 - (\nu J_z)^2)$ which is fixed by the couplings at energy scale $D$. In contrast to the spinful case [8], the absence of spin-flip interactions dominate over Kondo physics, and we obtain $T^* = D^{1/(1-K)}$. In the limit $1 - K \gg \nu J_0^2, \nu J_z^2$, Coulomb interactions dominate over Kondo physics, and we obtain $T^* \approx D^{1/(1-K)}$, a power-law form similar to the result for spinful Luttinger liquids [17]. From the exponent, we see that $T^*$ corresponds to the scale of the mass gap opened in the isotropic case $\alpha = 0$, one recovers the usual form $T^* = D^{1/(1-K)}$. In the limit $1 - K \gg \nu J_0^2, \nu J_z^2$, Coulomb interactions dominate over Kondo physics, and we obtain $T^* \approx D^{1/(1-K)}$, a power-law form similar to the result for spinful Luttinger liquids [17]. From the exponent, we see that $T^*$ corresponds to the scale of the mass gap opened
in a spinless Luttinger liquid by a nonmagnetic impurity of strength $\nu J_{\parallel}^0$ and the corresponding crossover is that of weak to strong single-particle backscattering.

In the high-temperature limit $\max\{|T^4_2, T^4_3\} \ll T \leq D$, both the Kondo and 2PB processes contribute logarithmically to the suppression of $G$, in the limit of small $T$. The topological nature of the QSH edge state as a “holochromatic liquid” living on the boundary of a 2D system [5] results in a drastic change of the low-energy effective theory in the vicinity of the strong coupling fixed point as compared to that of a usual 1D quantum wire. As suggested by the perturbative RG analysis, the nature of the $T = 0$ fixed point depends on whether $K$ is greater or lesser than $1/4$.

For $K > 1/4$, 2PB is irrelevant and $\Delta G_2$ flows to zero. On the other hand, for antiferromagnetic $J_z$, the Kondo strong coupling fixed point $J_{\parallel}, J_z \to +\infty$ is reached at $T = 0$, with formation of a local Kramers singlet and complete screening of the impurity spin by the HL. As a result, the formation of the Kondo singlet effectively removes the impurity site from the underlying 2D lattice [Fig. 2(b)]. In a strictly 1D spinful liquid, this has the effect of cutting the system into two disconnected semi-infinite 1D liquids [Fig. 2(a)] and transport is blocked at $T = 0$ for all $K^p < 1$ [8]. In contrast, due to its topological nature, the QSH edge state simply follows the new shape of the edge, and we expect the unitarity limit $G = G_0$ to be restored at $T = 0$. For finite $T \ll T_k$, the effective low-energy Hamiltonian contains the leading irrelevant operators in the vicinity of the fixed point. In the case of spinful conduction electrons, the lowest-dimensional operator causing a reduction of $G$ is single-particle backscattering. However, the helical property of the QSH edge states forbids such a term, and it is natural to conjecture that the leading irrelevant operator must be the 2PB operator with scaling dimension $4K$.

We thus expect a correction to $G$ at low temperatures $T \ll T_k$ for $K > 1/4$ of the form $\Delta G \propto -(T/T_k)^{2(4K-1)}$.

In particular, in the noninteracting case $K = 1$, we predict a $T^6$ dependence in marked contrast to both the usual Fermi liquid [18] and spinful 1D liquid [8] behaviors. This dependence characteristic of a “local helical Fermi liquid” can be understood from a simple phase space argument. The Pauli principle requires the 2PB operator to be defined through point-splitting [5] with the short-distance cutoff $\xi$, which translates into a derivative coupling $\psi^\dagger_k(0)\psi^\dagger_k(\xi)\psi_k(0)\psi_k(\xi) \to \xi^2\psi^\dagger_k\partial_\xi\psi^\dagger_k\partial_\xi\psi_k$ in the limit of small $\xi$. In the absence of derivatives, the 4-fermion term contributes $T^2$ to the inverse lifetime $\tau_k^{-1}$. The derivatives correspond to the first powers of momenta close to the Fermi points in the scattering rate $I_{k,k'\rightarrow p,p'} \propto (k-k')^2(p-p')^2$, which translates into an additional factor of $T^4$. Furthermore, since for $T \ll T_k$, the suppression of $G$ is entirely due to 2PB, we expect the effective charge

$$e^* = S/(2\langle|I_B|\rangle)$$

obtained from a measurement of the shot noise $S$ in the backscattering current $\langle I_B \rangle$ to be $e^* = 2e$ [11,19].

For $K < 1/4$, the $\lambda_2 \to \infty$ fixed point is reached at $T = 0$ and the system is insulating. The field $\phi(x = 0, \tau)$ is pinned at the minima of the cosine potential $H_1$ located at $\pm(2n + 1)\sqrt{\pi}/4$ for $\lambda_2 > 0$ and $\pm 2n\sqrt{\pi}/4$ for $\lambda_2 < 0$, with $n \in \mathbb{Z}$. $G$ is restored at finite $T$ by instanton processes corresponding to tunneling between nearby minima separated by $\Delta \phi = \sqrt{\pi}/2$. From the relation $\partial_x \phi = e\phi/\sqrt{\pi}$ where $\phi$ is the electric current, the charge pumped by a single instanton [Fig. 2(c)] is $\Delta Q_{\text{inst}} = \sqrt{\pi} / \Delta \phi = \frac{\pi}{2}$. This fractionalized tunneling current can be understood as the Goldstone-Wilczek current [20–22] for 1D Dirac fermions with a mass term $\delta L = g\Psi(m_1 + i\gamma^0 m_2)\Psi$ where $g \sim \lambda_2$, and the mass order parameters $m_1 = \cos 2\sqrt{\pi}/\phi$, $m_2 = \sin 2\sqrt{\pi}/\phi$ change sign during an instanton process. The order is Ising-like because the 2PB term explicitly breaks the spin $U(1)$ symmetry of the helical liquid $H_0 = \frac{\pi}{2} \int dx (K\sigma^2 + \frac{1}{K} \rho^2)$ down to $\mathbb{Z}_2$, where $\sigma_\uparrow = \rho_+ - \rho_-$ and $\rho = \rho_+ + \rho_-$ are the spin and charge densities, and $\rho_\uparrow$ are the chiral densities for the two members of the Kramers pair.

Fractionalization of the tunneling current is confirmed by a saddle-point evaluation of the low-energy effective action $S_{\text{eff}}$ for large $\lambda_2$ in the dilute instanton gas approximation, which yields a Coulomb gas model that can be mapped exactly to the boundary sine-Gordon theory

$$S_{\text{eff}}[\theta] = \frac{K}{B} \sum_{\langle i,j\rangle} \langle \omega_n \rangle |\theta(i\omega_n)|^2 + t \int_0^\beta d\tau \cos \sqrt{\pi}/\theta(\tau),$$

where $t$ is the instanton fugacity. The RG equation for $t$ follows as $\frac{dt}{dT} = (1 - \frac{1}{K})t$ and $G(T) \approx t(T)$ is a power law $G \propto (T/T^*_c)^{2(1/4K-1)}$ for $T \ll T^*_c$, $K < 1/4$. In contrast to the strong coupling regime in a usual Luttinger liquid
where $t$ corresponds to a single-particle hopping amplitude \cite{23}, the unusual scaling dimension of the tunneling operator in the present case corresponds to half-charge tunneling. In particular, we calculate the shot noise in the strong coupling regime using the Keldysh approach \cite{24} and find $S = 2e^2|\langle J \rangle|$ where $\langle J \rangle$ is the tunneling current and $e^* = e/2$.

Experimental realization.—We find that the experimental results of Ref. \cite{7} are consistent with our expressions for the weak coupling regime with a weak Luttinger parameter $K \approx 1$, but the small number of available data points does not allow for a reliable determination of the model parameters. The temperature dependence of the conductance being exponentially sensitive to $K$, our predictions can be best verified in QW with stronger interaction effects. Because of reduced screening of the Coulomb interaction \cite{25}, we expect to see a steeper decrease of conductance with decreasing temperature in HgTe samples with only a backgate. Because of lower Fermi velocities $v_F$, we expect even stronger interaction effects to occur in InAs/GaSb/AlSb type-II QW which are predicted to exhibit the QSH effect \cite{26}. For QW widths $w_{\text{InAs}} = w_{\text{GaSb}} = 10 \text{ nm}$ and considering only screening from the front gate closest to the QW layer, from a $k \cdot p$ calculation of material parameters, we obtain $K \approx 0.2 < 1/4$, making the insulating phase observable at low temperatures. Although the backgate will cause additional screening, $v_F$ can be further decreased by adding a thin AlSb barrier layer between the InAs and GaSb QW layers. $v_F$ is controlled by the overlap between electron and hole subband wave functions \cite{3} which are localized in different layers for type-II QW \cite{26}, and an additional barrier layer will decrease this overlap. A lower $v_F$ also translates into higher $T^*_K$ since $vJ \propto 1/v_F^2$, where one factor of $v_F$ comes from the matrix element of the localized impurity potential between edge states, and the other factor comes from the density of states $\nu$. Since $T^*_K$ depends on $vJ$ exponentially, we expect an experimentally accessible $T^*_K$ in type-II QW.

We are grateful to E.-A. Kim, T. Hughes, M. König, E. Berg, T. Ong, A. Furusaki, H. Yao, and S. Raghu for valuable discussions. J.M. is supported by NSERC of Canada and FQRNT, and the Stanford Graduate Program. C.X.L. is supported by the CSC, NSF under Grant Nos. 10774086, 10574076, and Basic Research Development of China under Grant Number 2006CB921500. Y.O. is supported by BSF, DIP, and GIF grants, the Mel Schwartz Research Fund, an NSF CAREER Grant No. DMR-0349354, and the SITP. C.W. is supported by the NSF under Grant No. DMR-0804775, the Sloan Research Foundation, and the Army Research Office under Grant No. W911NF0810291. S. C. Z. is supported by the U.S. Department of Energy, Office of Basic Energy Sciences under Contract No. DE-AC03-76SF00515.

[13] This result holds for $\hbar v/L \ll T < D$ and for Fermi liquid leads \cite{14} where $L$ is the length of the QSH region.
[16] The Kondo model derived from the Anderson model for a single level coupled to the HL is isotropic due to TRS with $J_k = J_k^* = (|j|^2 + |\nu|^2)(\frac{1}{x^2 + k^2} + \frac{1}{x^2 + l^2})$ where $\epsilon_F$ is the Fermi energy, $\epsilon_d$ is the impurity level with on-site Coulomb repulsion $U$, and $u, t$ are the spin-flip and non-spin-flip hopping amplitudes, respectively. Coulomb interactions ($K \neq 1$) may however induce an effective anisotropy ($\alpha \neq 0$) even if the original Kondo model is isotropic.
[19] In the high temperature regime $T \gg T^*_2, T^*_K$, both the Kondo ($\epsilon^* = \epsilon$) and 2PB ($\epsilon^* = 2\epsilon$) contributions to the effective carrier charge are present, such that we expect a nonuniversal value for the Fano factor.
[25] $K$ can be estimated \cite{23} by $K = [1 + \alpha \ln(d/\ell)]^{-1/2}$ where $\alpha = \frac{2}{\pi} \frac{e^2/\epsilon}{\hbar^2}$ and $\epsilon$ is the bulk dielectric constant. The distance $d$ from the QW layer to a nearby metallic gate acts as a screening length for the Coulomb potential, and $\ell$ is a microscopic length scale $\ell = \max[\xi, w]$ which acts as a short-distance cutoff for the Coulomb potential, where $w$ is the thickness of the QW layer.