2D Materials

PAPER

Thermoelectric transport in two-dimensional topological insulator state based on HgTe quantum well

To cite this article: G M Gusev *et al* 2019 *2D Mater.* **6** 014001

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

This content was downloaded from IP address 143.107.135.53 on 07/01/2019 at 15:19

IOP Publishing 2D Mater. 6 (2019) 014001 https://doi.org/10.1088/2053-1583/aaf702

2D Materials

RECEIVED
21 September 2018

REVISED
23 November 2018

ACCEPTED FOR PUBLICATION
7 December 2018

PUBLISHED
21 December 2018

PAPER

Thermoelectric transport in two-dimensional topological insulator state based on HgTe quantum well

G M Gusev[1], O E Raichev[2], E B Olshanetsky[3], A D Levin[1], Z D Kvon[3,4], N N Mikhailov[3] and S A Dvoretsky[3]

[1] Instituto de Física da Universidade de São Paulo, 135960-170, São Paulo, SP, Brazil
[2] Institute of Semiconductor Physics, NAS of Ukraine, Prospekt Nauki 41, 03028 Kyiv, Ukraine
[3] Institute of Semiconductor Physics, Novosibirsk 630090, Russia
[4] Novosibirsk State University, Novosibirsk, 630090, Russia

E-mail: gusev@if.usp.br

Keywords: topological insulator, thermopower, edge states, quantum transport, HgTe quantum well

Supplementary material for this article is available online

Abstract

The thermoelectric response of HgTe quantum wells in the state of two-dimensional topological insulator (2D TI) has been studied experimentally. Ambipolar thermopower, typical for an electron–hole system, has been observed across the charge neutrality point, where the carrier type changes from electrons to holes according to the resistance measurements. The hole-type thermopower is much stronger than the electron-type one. The thermopower linearly increases with temperature. We present a theoretical model which accounts for both the edge and bulk contributions to the electrical conductivity and thermoelectric effect in a 2D TI, including the effects of edge to bulk leakage. The model, contrary to previous theoretical studies, demonstrates that the 2D TI is not expected to show anomalies of thermopower near the band conductivity threshold, which is consistent with our experimental results. Based on the experimental data and theoretical analysis, we conclude that the observed thermopower is mostly of the bulk origin, while the resistance is determined by both the edge and bulk transport.

1. Introduction

The HgTe-based quantum well (QW) is a semiconductor system, where the two-dimensional (2D) conduction and valence subbands, divided by a narrow variable gap, can be created. The switch between the electron and the hole types of transport is achieved by the gate control. The gap energy is controlled by the well width d and goes to zero at $d \simeq 6.3$ nm, when the electron energy spectrum resembles a Dirac cone, like in graphene. Wider wells have an inverted energy band order, known to be in the state of 2D topological insulator (TI) [1] characterized by a pair of counterpropagating gapless edge modes. The edge states have a helical spin structure and are supposed to be robust to backscattering. The edge state transport in such HgTe QWs has been confirmed experimentally for ballistic transport in mesoscopic samples [2–4] and for diffusive transport in macroscopic samples [4, 5]. Similar results have been obtained in experiments on Si-doped InAs/GaSb quantum wells [6–9] which are also believed to be 2D TI. Application of novel experimental methods for the study of the transport properties of 2D TI is of particular interest.

The thermoelectric measurements can give complementary information about electron transport in metals and semiconductors and are used as a powerful tool for probing the sign of the charge carriers and the transport mechanisms. The diffusive thermopower is often described using the Mott relation, as the logarithmic derivative of the energy-dependent electrical conductivity. Apart from necessitating the validity of the Boltzmann equation and the degeneracy of the electron gas, the Mott relation has also other limitations discussed in the literature. In particular, a strong energy dependence of the relaxation time, when this time changes considerably within the kT interval around the Fermi level, causes a failure of the simple Mott relation. Indeed, the deviation from the Mott relation has been observed in thermopower experiments in graphene near the charge neutrality point (CNP) and at high temperatures [10, 11]. Application of the original Mott relation [12], which is more general because it is not based on the approximate Sommerfeld expansion, allows one to overcome these

© 2018 IOP Publishing Ltd

limitations and describe the thermopower in a wide range of temperature and carrier density [13]. However, in the case of strongly inelastic scattering by optical phonons, considerable deviations even from the Mott relation in its general form are expected, as recently demonstrated in high mobility graphene samples [14, 15].

Similar to graphene, the system based on HgTe quantum wells reveals ambipolar Hall effect accompanied by the resistance peak near the CNP. However, in contrast to the gapless graphene, in 2D TI the transition between the electron and hole types of conduction as the gate voltage is swept through the CNP, occurs when the Fermi level stays in the insulating gap and transport is determined by the edge states. The position of the Fermi level in the gap is stabilized by the bulk states which are present in the gap because of a random spatial inhomogeneity (disorder). These states are often considered as localized ones. However, electron transitions between the edge and the bulk states are possible and may influence transport properties in 2D topological insulators [4, 5, 16]. The edge to bulk mixing is believed to cause a strong enhancement of the thermopower in 2D TI. In particular, it is suggested [17, 18] that when the Fermi level approaches the bulk band edge, the scattering rate of electrons in the edge states increases rapidly and significantly, which is expected to cause an anomalous growth of the amplitude of Seebeck signal and a change of its sign. This offers a new opportunity to improve the thermoelectric parameter, such as the figure of merit zT, which is defined as $zT = GS^2T/(K_e + K_{ph})$, where G is the electron conductance, S is the Seebek coefficient, K_e and K_{ph} are the thermal conductances of electrons and phonons consequently. The interplay between the edge and the bulk conductances leads to a strong dependence of the parameter zT on the sample geometry and size. It has been predicted that the value of the figure of merit can be improved by more than $\sim$1 for a certain geometry at room temperature [17, 18]. Despite the interest to the thermoelectric properties in 2D topological insulators, the experimental studies have almost all been focused on the measurements of the electrical resistance.

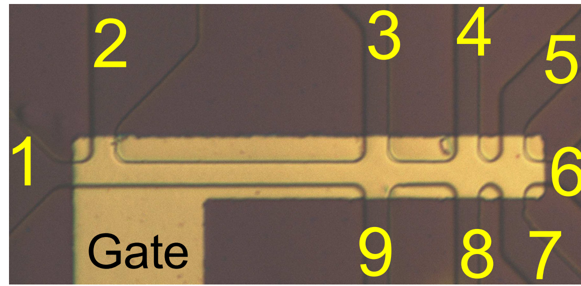
In the present paper we report an experimental study of the thermopower in band-inverted HgTe-based quantum wells. At the CNP where the resistance reaches its maximum, the thermopower changes its sign, showing the ambipolar behavior. The nonlocal resistance in our samples is comparable with the local one. This observation clearly proves the presence of the edge state transport, which dominates within the bulk gap. Importantly, we do not observe any of the anomalies of the Seebeck effect predicted in [17, 18], in particular, the sign of the effect changes like in a normal electron–hole system. This apparently suggests that the effect of the edge to bulk scattering on the transport is not as significant as it was expected. To verify this statement, we have carried out a calculation of conductivity and thermopower in the 2D TI, taking into account both the particle and energy balance in the coupled system of edge and bulk states. In brief, we demonstrated that the transport properties are determined not by the edge to bulk scattering rate alone, but by the spin current flowing between the edge and the bulk. If the spin relaxation in the bulk is slow, a bottleneck effect takes place, when the spin current is limited by the bulk conductivity rather than by the edge to bulk scattering rate. Therefore, in the region where the bulk conductivity is smaller than the conductance quantum e^2/h the scattering between the edge and the bulk is expected to be insignificant, while in the regions of larger bulk conductivity the edge-state contribution to transport is no longer important. Further, from a qualitative analysis of our experimental data supported by the theoretical considerations, we conclude that the observed thermopower is mostly of the bulk transport origin.

The paper is organized as follows. Section 2 contains the description of measurements and experimental results. Section 3 is devoted to the theoretical model and to discussion of the results based on this model. The concluding remarks are given in the last section.

2. Experiment

To probe the carrier transport, we measured thermoelectric voltage V and thermopower S together with the resistance. The quantum well structures $Cd_{0.65}Hg_{0.35}Te/HgTe/Cd_{0.65}Hg_{0.35}Te$ with [013] surface orientations and widths $d = 8$–8.3 nm were prepared by molecular beam epitaxy (figure 1, left panel). The sample is a long Hall bar consisting of three 3.2 μm wide consecutive segments of different length (3, 9, and 35 μm) and seven voltage probes, covered by the TiAu gate (see figure 1, top panel). The measurements were performed in a variable temperature insert cryostat in the temperature range 1.4–10 K using the standard four point scheme. A detailed description of the sample structure has been given in [5]. The electrically powered heater placed symmetrically near the contact 1 (see figure 2, top panel) creates temperature gradient in the system, while the other end is indium soldered to a small copper slab that serves as a thermal ground. The copper slab is, in turn, connected to the copper rod of the sample holder. One calibrated thermo sensor is attached at the end of the sample near the heater while the other is attached to the heat sink. The thermo sensors were used to measure the ΔT along the sample. The voltages induced by this gradient were measured by a lock-in detector at the frequency of $2f_0 = 0.8$–2 Hz across various voltage probes. The thermal conductance of the sample is overwhelmingly dominated by phonon transport in the GaAs substrate [19, 20]; diffusive heat transport by the two-dimensional gas is negligible in comparison. The thermal conductivity κ of a pure dielectric crystal is usually determined by the boundary scattering at low temperatures and depends

CdTe (40 nm)
$Hg_xCd_{1-x}Te$ (~30 nm)
HgTe QW (8.3 nm)
$Hg_xCd_{1-x}Te$ (~9nm)
In doping (~12 nm)
$Hg_xCd_{1-x}Te$ (~9nm)
CdTe
ZnTe
GaAs



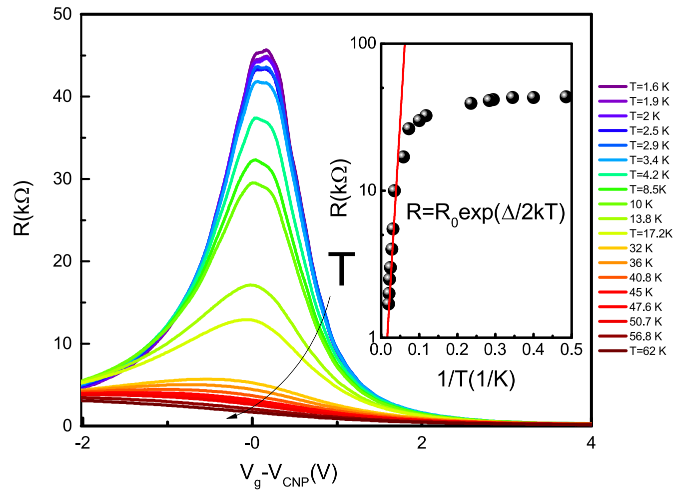


Figure 1. Sample geometry and resistance of shortest segment ($R_{I=1,6;V=4,5}$) as a function of the gate voltage for different temperatures (sample A). Left-schematic structure of the sample.

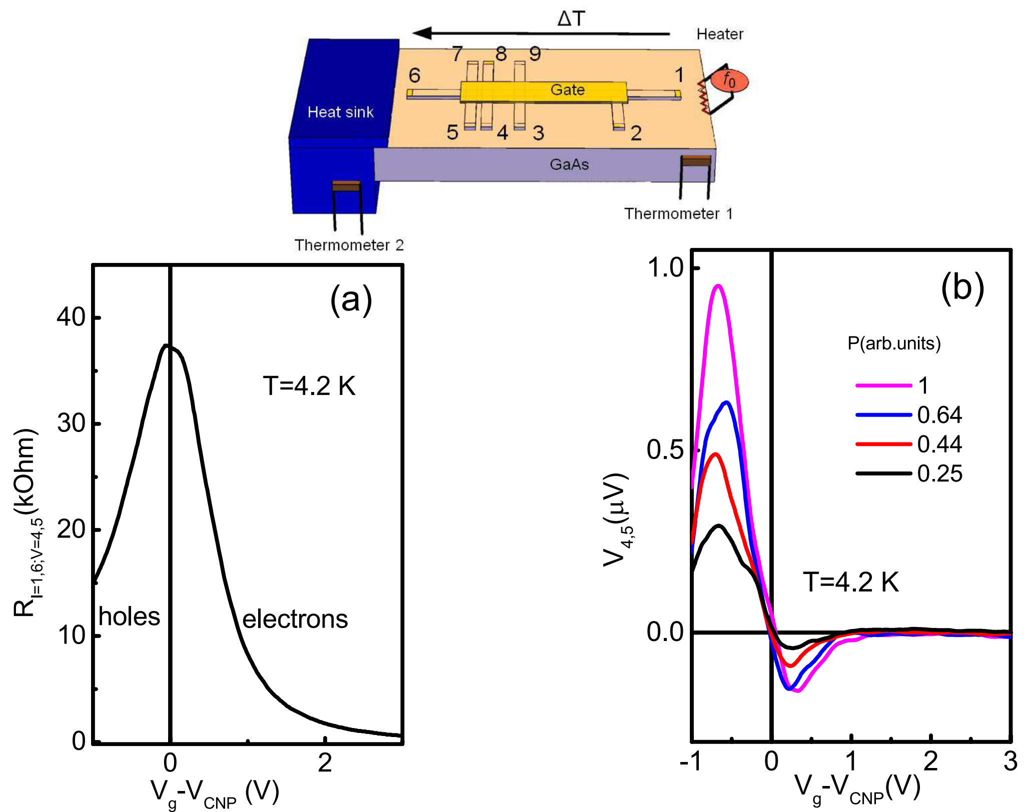


Figure 2. Resistance (a) and thermovoltage (b) for different heater powers (indicated) as a function of gate voltage, $T = 4.2$ K (sample A).

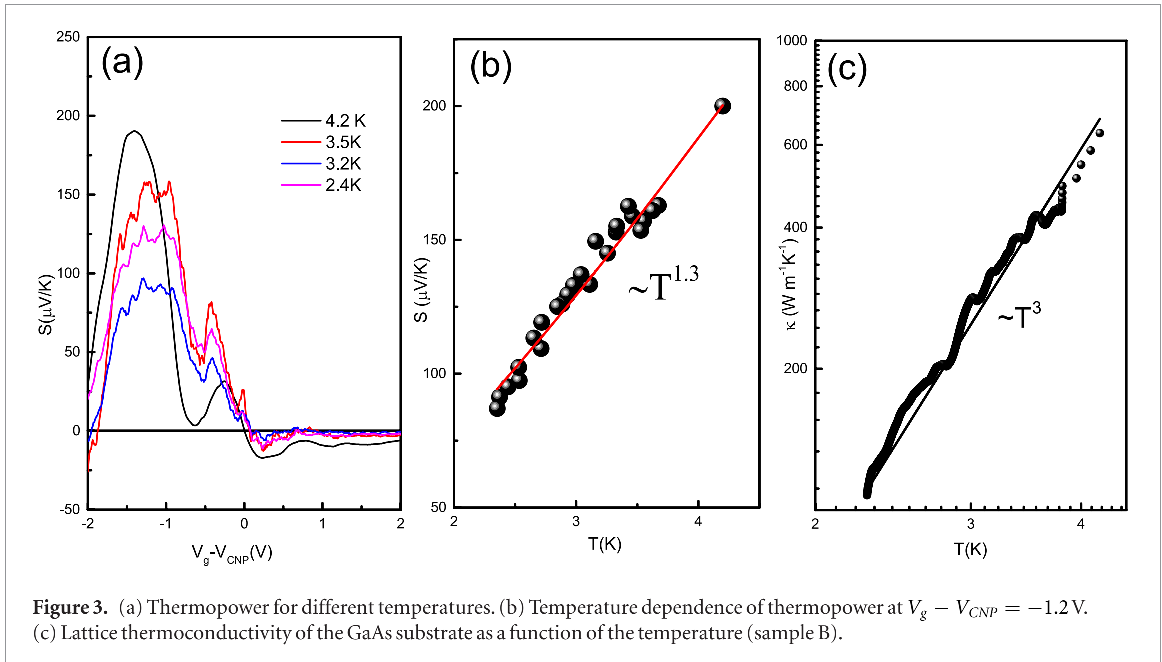


Figure 3. (a) Thermopower for different temperatures. (b) Temperature dependence of thermopower at $V_g - V_{CNP} = -1.2$ V. (c) Lattice thermoconductivity of the GaAs substrate as a function of the temperature (sample B).

on the temperature as $\kappa \sim T^3$. We performed the measurements of the thermal conductivity in our samples and obtained the value 300–200 W m^{-1} K^{-1} at $T = 4.2$ K for different substrates, which agrees with the previous measurements [19, 20]. We did not directly measure the temperature difference between the voltage probes, since the distance is very small. We estimate this difference between probes 3 and 2 as ~20 mK K for the heater power used in our experiment. For a given temperature difference between the sample extremities the temperature profile along the sample could be nonlinear, especially close to the ends of the substrates. The situation is somewhat similar to the electrical measurements, when the electric potential profile is inhomogeneous near the metallic contacts. However, we expect that in the center of the sample and along the short distances the profile is linear. We have also checked that the temperature difference varies linearly with heater power. Four different devices have been studied. Below we show the results obtained in two representative samples (A and B).

The variation of the resistance with the gate voltage and lattice (bath) temperature is shown in figure 1. The resistance of the shortest segment reveals a broad peak whose amplitude is larger than the value $h/2e^2$ expected in the ballistic case. We see that the resistance decreases sharply for temperatures above 15 K while saturating below 10 K. We find that the profile of the resistance temperature dependencies above $T > 15$ K fits very well the activation law $R\ exp(\Delta/2$ kT), where Δ is the activation gap. Insert in figure 1 shows the peak maximum resistance versus temperature. The thermally activated behavior of resistance above 15 K corresponds to a gap of 10 meV between the conduction and valence bands in the HgTe well. The mobility gap can be smaller than the energy gap due to disorder. Figure 2 shows the resistance and thermovoltage as a function of the gate voltage measured between probes 4 and 5 at $T = 4.2$ K. The thermovoltage increases nearly linearly with heater power, which proves that we measure the longitudinal (Seebeck) thermoelectric effect. The signal shows a behavior similar to other electron–hole systems such as graphene [10, 11]. It changes sign at the charge neutrality point (CNP) and decreases with the carrier density increasing. The voltage interval between the electron-like and hole-like regimes ($\Delta V_g \sim 1$ V) is almost two times smaller than the half-width of the resistance peak. The figure 3(a) displays the traces of the thermopower versus V_g for different temperatures. Figure 3(b) shows the temperature dependence of the thermopower measured across a longer bridge at a selected gate voltage $V_g - V_{CNP} = -1.2$ V (hole side) where the thermopower approaches its maximum. It is found that the signal grows almost linearly with temperature: $S \sim T^{1.3\pm0.1}$ (figure 3(b)) in the temperature interval $2.2 < T < 4.2$ K. This temperature interval was selected, because resistance becomes temperature dependent above $T > 10$ K (see figure 1), and metallic approximation for thermopower is no longer valid. It is worth noting that prior to the thermoelectric measurements the thermal conductance of the sample has been determined. The thermoconductance is dominated by the phonon transport in the substrate; the contribution from the diffusive heat transfer by the electrons is negligibly small. The thermal conductivity of the GaAs substrate is usually determined by the boundary scattering at low temperatures [19, 20]. The thermoconductivity is given by:

$$\kappa = \frac{2\pi^2}{15}\frac{k\Lambda}{v_{ph}^2}\left(\frac{kT}{h}\right)^3, \qquad (1)$$

where Λ is the phonon mean-free path and $v_{ph} = 3300$ m s^{-1} the appropriate mean acoustic phonon velocity. Figure 3(c) shows the thermal conductivity of the GaAs substrate as a function of the lattice temperature.

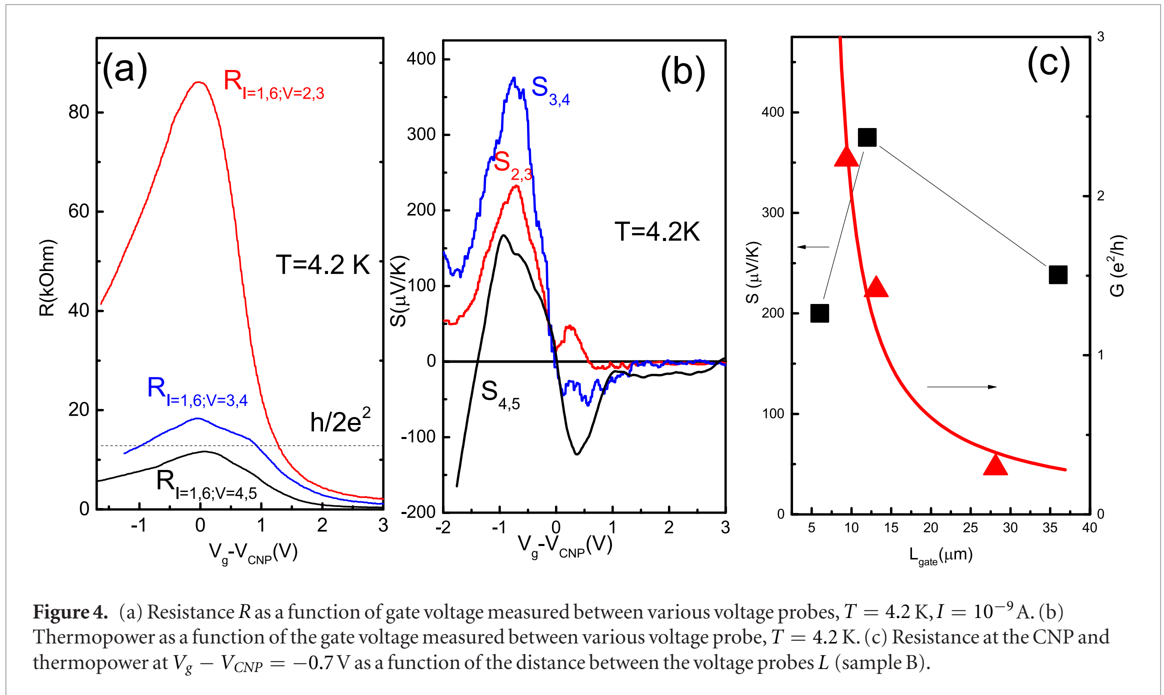


Figure 4. (a) Resistance R as a function of gate voltage measured between various voltage probes, $T = 4.2$ K, $I = 10^{-9}$ A. (b) Thermopower as a function of the gate voltage measured between various voltage probe, $T = 4.2$ K. (c) Resistance at the CNP and thermopower at $V_g - V_{CNP} = -0.7$ V as a function of the distance between the voltage probes L (sample B).

The thermal conductivity follows $\kappa \sim T^3$ law, in accordance with equation (1). The value of κ is ~600 W m^{-1} K^{-1} at $T = 4.2$ K, which agrees well with the previously measured thermoconductivity in pure GaAs substrates [19, 20]. Once the thermal conductance is found, the temperature gradient can be used to convert the measured thermoelectric voltages into the thermopower. Figure 4 shows resistance as a function of the gate voltage measured between different probes at $T = 4.2$ K. It is worth noting that the edge current flows along the gated sample edge whose length L_{gate} is longer than the distance between the probes L (bulk current path) and corresponds to 5–6 μm. For longer distances between the probes we see higher resistances. The large resistance can appear because of multiple transitions of electrons between counterpropagating helical edge states caused by either direct backscattering or electron transfer mediated by the bulk states in the puddles [21] emerging due to spatial potential fluctuations near the edge. The observation of a nonlocal resistance constitutes the main proof of the presence of the edge state transport in a 2D TI. A systematic study of the local and nonlocal transport in 2D TI has been preformed in the previous works in the ballistic [3, 4] and in the diffusive regimes [16]. The dependence of the resistance peak on the length L_{gate} is shown in figure 4(c). It is found to be very close to the $1/L_{gate}$ dependence. The thermopower signal has a nonmonotonic dependence on the distance L_{gate} (figure 4(c)) in contrast to the resistance. Below we consider the theory, which accounts for both the edge and bulk contributions to the electrical conductance and thermoelectric effect in 2D TI, including the effects of edge to bulk leakage.

Finally a few words need to be said about the thermoelectric efficiency in the 2D topological insulator regime. As mentioned in the introduction, in conventional semiconductors the factor zT is size independent because the geometrical factor is canceled between the conductance and the thermoconductance. In 2D topological insulator regime factor zT can be optimized by choosing appropriate geometries [18]. A large enhancement of the power factor is predicted for the gapless edge states near the charge neutrality point. Figure 5(a) shows the conductance measured near the CNP as a function of the gate voltage. It is expected that when the edge-state contribution to the transport is important, the nonlocal resistance should be observed [3, 4].

For comparison figure 5(a) presents an example of the nonlocal resistance when the current flows between contacts 4–8 and the voltage is measured between contacts 5–7, i.e. $R_{I=4,8;V=5,7}$. The nonlocal resistance peaks are narrower and lower as compared to the local resistance peaks measured in the same device. Simple estimation from Kirchoff formula (see for details [16]), gives $R_{I=4,8;V=5,7} = 0.6\frac{h}{e^2}$ for the mean free path $l = 10$ μm, which in 2 times large than the experimental value $R_{I=4,8;V=5,7} = 0.3\frac{h}{e^2}$. It is not surprising, since when the edge current flows over a long distance (in this particular case $L_{4,8} \gg l$) there is a high probability for the coupling with the bulk states, and the total current experiences considerable leakage into the bulk. An advanced theory, considered in [16], is required for a more detailed analysis. However, we may conclude here that the edge state transport dominates in the voltage interval -1 V $< V_g <$ 1 V. Figure 5(b) shows the coefficient GS^2 as a function of the gate voltage near the CNP in the edge state transport regime at $T = 4.2$ K. It is clear that an enhancement of GS^2 is observed compared to other than TI cases on the hole-side of the resistance peak. Unfortunately this enhancement is observed at low temperature and is unlikely to survive

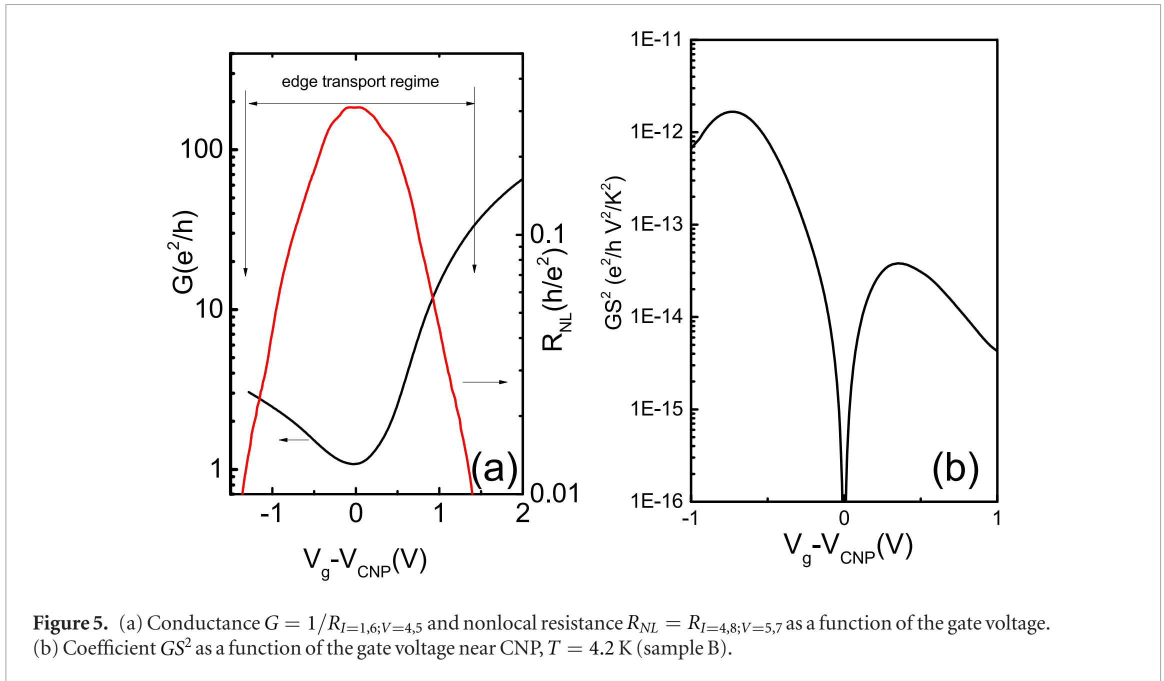


Figure 5. (a) Conductance $G = 1/R_{I=1,6;V=4,5}$ and nonlocal resistance $R_{NL} = R_{I=4,8;V=5,7}$ as a function of the gate voltage. (b) Coefficient GS^2 as a function of the gate voltage near CNP, $T = 4.2$ K (sample B).

at high T (see figure 1). However, quite recently the TI state has been observed at $T \sim 100$ K [22], and one may hope that the thermopower characteristics presented here could be valid at higher temperatures in these new materials.

3. Theory

Since the thermovoltage follows almost linear temperature dependence and monotonically decreases with increasing carrier density far away from the CNP, the thermopower is likely determined by the diffusive mechanism. The phonon drag thermopower would have different temperature dependence. In comparison to GaAs quantum wells, where the phonon drag mechanism is essential, in HgTe quantum wells this mechanism should be much less important because of relative smallness of deformation-potential and piezoelectric constants in HgTe, and because of relative smallness of the density of states of 2D carriers. In the edge state transport, the possibility of phonon drag is negligible because of topological protection of the edge states. Therefore, we consider the diffusive mechanism in the following.

The transport in the edge channel k (since the channel is helical, $k = 1, 2$ denote both the direction of motion and spin orientation) along the axis Ox is described by the Boltzmann equations for the energy distribution functions of electron, $f_{k\varepsilon}(x)$:

$$s_k \frac{\partial f_{k\varepsilon}}{\partial x} = \gamma_\varepsilon (f_{k'\varepsilon} - f_{k\varepsilon}) + g_\varepsilon (F_{k\varepsilon} - f_{k\varepsilon}) + \frac{J^{ee}_{k\varepsilon}}{v} + \frac{J^{ph}_{k\varepsilon}}{v}, \quad (2)$$

where $k' \neq k$, v is the edge state velocity, $F_{k\varepsilon}(x, y)$ is the isotropic part of the electron distribution function in the bulk, $\gamma_\varepsilon = \nu^{bs}_\varepsilon / v$ and $g_\varepsilon = \nu^{eb}_\varepsilon / v$ are the inverse mean free path lengths for elastic backscattering and edge to bulk scattering (ν^{bs} and ν^{eb} are the corresponding scattering rates). Next, J^{ee} and J^{ph} are the collision integrals for inelastic processes, electron–electron and electron–phonon scattering. Assuming that the electrons in the state 1 move from the left to the right (i.e. have positive velocity), we put $s_1 = 1$ and $s_2 = -1$. The transitions between counterpropagating edge states (either elastic or inelastic [23]) are rare because of topological protection, while the scattering between the edge and the bulk is weak because of low probability of finding the bulk-state puddles [21] in the close vicinity to the edge (it may become strong, however, above the threshold of band conductivity). On the other hand, the inelastic electron–electron scattering within a single edge channel is free from these restrictions and, therefore, is much stronger. Moreover, such kind of scattering has a very large phase space, especially if v is energy-independent so that momentum and energy conservation rules are satisfied simultaneously for any two electrons participating in the collisions [24]. As the electron–electron scattering controls electron distribution, we search the distribution function in the Fermi-like form, $f_{k\varepsilon}(x) = \{\exp[(\varepsilon - \varphi_k(x))/T_{ke}(x)] + 1\}^{-1}$ [24], characterized by the coordinate-dependent elecrochemical potential φ_k and temperature T_{ke}. Assuming that in the bulk the different spin states are weakly coupled, we use the similar form for $F_{k\varepsilon}(x, y)$. Indeed, in the original Bernevig–Hughes–Zhang Hamiltonian [25] describing HgTe quantum wells the spin states are uncoupled, the coupling appears when the spin–orbit terms are introduced [2]. The coupling in symmetric wells is described by the bulk inversion asymmetry parameter Δ [2], and the probability ratio of spin-flip to spin-conserving transitions is estimated as a ratio of the energy Δ to the gap energy. This ratio is small (about 0.05) for our quantum wells.

Integrating equation (2) over ε, then multiplying equation (2) by ε and integrating again, we obtain four balance equations expressing conservation of particles and energy (for details see supplementary information (stacks.iop.org/TDM/6/014001/mmedia)). They are solved together with the balance equations for the bulk states and boundary conditions expressing conservation of currents and energy fluxes at the edge. We consider the case when particle transfer between edge and bulk occurs only in the narrow regions near the contacts, while in the most part of the edge a dynamical equilibrium is reached, when only spin currents between edge and bulk are flowing. This corresponds to the condition $g \gg 1/L$ (g is energy averaged quantity of g_ε, defined below) and leads to a homogeneous distributions of temperatures and potentials at the edge: the spin-averaged potentials and temperatures linearly depend on x, while the quantities describing spin polarization, $\delta\varphi = \varphi_1 - \varphi_2$ and $\delta T_e = T_{1e} - T_{2e}$, are constants. Only this condition is relevant, because under the opposite condition $g \leqslant 1/L$ the particle transfer between edge and bulk cannot have a sizeable influence on the transport. The current carried by a single edge is given as $I_e = (e/h)\int d\varepsilon(f_{1\varepsilon} - f_{2\varepsilon}) = (e/h)\delta\varphi$, and the thermal current (energy flux minus $\mu I_e/e$) along the edge is $W_e = (1/h)\int d\varepsilon(\varepsilon-\mu)(f_{1\varepsilon} - f_{2\varepsilon}) = (\pi^2/3h)T\delta T_e$, where μ and T are the equilibrium chemical potential and temperature. The total current $I = I_{bulk} + 2I_e$ and the total thermal current $W = W_{bulk} + 2W_e$ are connected to the voltage ΔV and temperature difference ΔT between the contacts by a linear relation

$$\begin{pmatrix} I \\ (e/T)W \end{pmatrix} = \frac{2e^2}{h}\hat{\mathcal{G}}\begin{pmatrix} \Delta V \\ \Delta T/e \end{pmatrix}, \tag{3}$$

where we introduced a thermoelectric response matrix

$$\hat{\mathcal{G}} = \frac{w\tilde{\sigma}}{L}\hat{M} + \hat{c}\left[\hat{c} + L\hat{\gamma} + \frac{L}{2}\left(\frac{w}{2\tilde{\sigma}}\hat{M}^{-1} + \hat{g}^{-1}\right)^{-1}\right]^{-1}\hat{c}. \tag{4}$$

The first and the second terms of this matrix describe the bulk and the edge contributions, respectively, and the term in the round braces describes the influence of the edge to bulk leakage on the edge transport. Here and below, $\tilde{\sigma} = \sigma/(e^2/h)$, σ is the bulk conductivity per spin, L is the distance between the contacts, w is the sample width, and the matrices are defined as

$$\hat{\gamma} = \begin{pmatrix} \gamma & \gamma_I \\ \gamma_I & \gamma_{II} \end{pmatrix}, \hat{g} = \begin{pmatrix} g & g_I \\ g_I & g_{II} \end{pmatrix}, \hat{c} = \begin{pmatrix} 1 & 0 \\ 0 & \pi^2/3 \end{pmatrix}, \tag{5}$$

$$\hat{M} = \frac{1}{\sigma}\begin{pmatrix} \sigma & \sigma_I \\ \sigma_I & \sigma_{II} \end{pmatrix} = \begin{pmatrix} 1 & eS_b \\ eS_b & e^2(\kappa_b/\sigma T + S_b^2) \end{pmatrix}. \tag{6}$$

The matrix $\hat{M}$ describes thermoelectric response in the bulk (for details see supplementary information) and is expressed through the bulk thermopower $S = S_b$ and electronic thermal conductivity κ_b. The quantities X, X_I, and X_{II} are introduced as the averages $X \equiv \langle X_\varepsilon \rangle = \int d\varepsilon(-\partial f_\varepsilon^{(0)}/\partial\varepsilon)X_\varepsilon$, $X_I \equiv \langle X_\varepsilon(\varepsilon-\mu)/T \rangle$, and $X_{II} \equiv \langle X_\varepsilon(\varepsilon-\mu)^2/T^2 \rangle$, where $f_\varepsilon^{(0)} = \{\exp[(\varepsilon-\mu)/T] + 1\}^{-1}$ is the equilibrium distribution. Equation (3) is written under an additional assumption that spin relaxation length in the bulk exceeds the sample halfwidth $w/2$. We also neglected energy transfer between the edge states and the lattice, which is justified, according to our estimates, in the samples of submillimeter length. The total thermopower $S_{tot} = -(\Delta V/\Delta T)_{I=0}$ is determined by the ratio of non-diagonal to diagonal elements of the matrix (3), $S_{tot} = e^{-1}\mathcal{G}_{12}/\mathcal{G}_{11}$. The total conductance is $G_{tot} = I/\Delta V = (2e^2/h)\mathcal{G}_{11}$.

In the general case, equation (3) tells us that under conditions $\tilde{\sigma} < 1$, where the edge conductivity can be significant, the contribution due to $\hat{g}$ is not important (unless $gw/2 < \tilde{\sigma} < 1$ which necessarily implies $w \ll L$ because $g \gg 1/L$ is assumed). The whole contribution of the term describing edge to bulk leakage in $\hat{\mathcal{G}}$ cannot exceed $(L/w)\tilde{\sigma}\hat{M}$ and can be neglected in the case of $\gamma L \gg 1$. This is a manifestation of the bottleneck effect described in the Introduction. On the other hand, when $\tilde{\sigma} \gg 1$, there is no need to take the edge transport into account.

In the case when Sommerfeld expansion for each of the energy-dependent parameters is valid, i.e. $\hat{X} \simeq \hat{c}X + (\pi^2/3)TX'\hat{\sigma}_x$, where $X' = dX_\mu/d\mu$ and $\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, the Mott relation $S = (\pi^2/3e)T(G'/G)$ is valid as well. Then the second term in equation (3) is simplified and leads to the edge state contribution to the conductance and thermopower as follows:

$$G_e = \frac{e^2}{h}\mathcal{F}^{-1},\ \mathcal{F} = 1 + \gamma L + \frac{(\frac{L}{w})\tilde{\sigma}gL}{(\frac{2L}{w})\tilde{\sigma} + gL}, \tag{7}$$

$$S_e = \frac{\pi^2 T}{3|e|}\frac{\mathcal{F}'}{\mathcal{F}}, \tag{8}$$

while the total conductance and thermopower are

$$G_{tot} = 2(G_e + G_b),\ S_{tot} = \frac{S_eG_e + S_bG_b}{G_e + G_b}, \tag{9}$$

where $G_b = \sigma w/L = (e^2/h)(w/L)\tilde{\sigma}$ is the bulk conductance per spin. Though the theory leading to equations (7) and (8) initially assumes $gL \gg 1$, these equations remain formally valid in the opposite limit $gL \ll 1$, when the edge state contribution to the conductance is not influenced by the bulk-edge currents and $G_e = \frac{e^2}{h}(1+\gamma L)^{-1}$. This means that equations (7) and (8) can be used as a reasonable approximation for arbitrary gL.

The main results of the our theoretical model is given by equations (7) and (8). One can see that the function $\mathcal{F}$ contains two terms: the first depends on the scattering between edges and the second describing the edge to bulk leakage. Two cases can be considered:

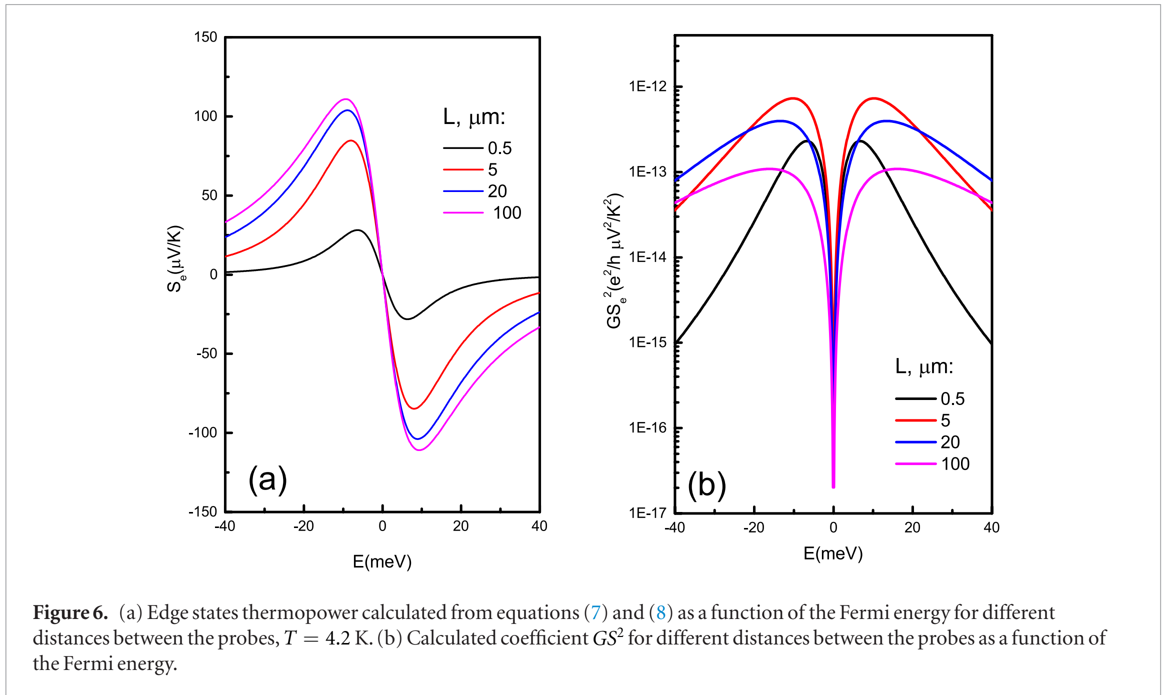


Figure 6. (a) Edge states thermopower calculated from equations (7) and (8) as a function of the Fermi energy for different distances between the probes, $T = 4.2$ K. (b) Calculated coefficient GS^2 for different distances between the probes as a function of the Fermi energy.

(a) Fermi level is inside of the gap.

In this case if the bulk transport is suppressed by localization the conductance and thermopower are governed by the backscattering between the edge states.

(b) The Fermi level enters conductance (valence) band.

The edge state conductance and thermopower are determined by equations (7) and (8). If $(L/w)\tilde{\sigma} \ll gL$, the presence of the edge to bulk scattering does not strongly affect either the resistance or the thermopower, because:

$$\mathcal{F} = 1 + \gamma L + \frac{L}{w}\tilde{\sigma}. \qquad (10)$$

This also means that even when energy dependence of g is strong, the thermopower sign alteration considered in model [18] does not occur. For the long narrow sample and for high mobility bulk carriers (high conductivity), if $(L/w)\tilde{\sigma} \gg gL$ the edge state contribution to function $\mathcal{F}$ should be proportional to gL, and the conductance and thermopower should be given by:

$$G = \frac{e^2}{h}(1 + \gamma L + g)^{-1} \qquad (11)$$

$$S_e = \left(\frac{\pi^2 T}{3e}\right)\frac{\gamma' + g'}{1 + \gamma L + gL}. \qquad (12)$$

Note that for the limiting case $\gamma' \ll g' \ll 0$ the thermopower is large and has a positive anomalous sign for electrons (for conventional 2D and 3D system Seebek coefficient always negative for electrons). This mechanism has been predicted in model [18]. Thus, for observations of the anomalous thermopower long samples with suppressed edge to edge scattering is required.

4. Discussion and comparison with experiment

In the ballistic case, the edge state contribution to the thermopower is absent. However, in the experiment we observe thermopower signal in a quasi-ballistic case (figures 2 and 4). Generally, it is expected that γ is a smooth function of chemical potential, monotonically decreasing away from the CNP, because the CNP roughly corresponds to the crossing point in the edge-state spectrum, where the transferred momentum is zero, and an elastic scattering rate usually decreases with increasing transferred momentum. This property of the scattering should cause a decrease of the resistance with gate voltage $|V_g - V_{CNP}|$ and a negative slope of the thermopower near the CNP, as it is observed in experiment. If the energy is counted from the Dirac point, the transferred momentum is $q = 2\varepsilon/v_e$. Assuming, for example, that the backscattering rate ν_ε is proportional to a Lorentzian function $\gamma_0/[1 + (\varepsilon/\varepsilon_0)^2]$, one can find $\gamma'/\gamma = -(2\mu/\varepsilon_0^2)/[1 + (\varepsilon/\varepsilon_0)^2]$. The thermopower is a linear function of the Fermi energy near the CNP, negative in the electron part and positive in the hole part. If we assume that the transport near the CNP is dominated by the edge states, then the slope of the thermopower near the CNP is entirely determined by the energy dependence of the backscattering rate.

In this region the assumed approximation $g = 0$ is relevant.

Figure 6(a) shows the thermopower calculated from equations (7) and (8) taking into account the Lorentzian energy dependence of the backscattering rate for different edge channel lengths L with the following parameters: $\varepsilon_0 = 10$ meV, $\gamma_0 = 1/\mu$m, $\tilde{\sigma} = 0.1$. One can see a nonlinear length dependence. A nonmonotonic length dependence has also been observed in our samples (see figure 4(c)). Figure 6(b) shows the coefficient GS^2 as a function of the energy near the CNP in the edge state transport regime at $T = 4.2$ K for different edge channel lengths. The theory (equations (7) and (8)) confirms the predictions of the model [18] concerning the dependence on geometry. In practice, however, geometry-related optimization would necessitate the possibility to minimize the thermoconductance of a realistic 2D TI. This is not feasible in our case, since the thermoconductance in our structures is mostly determined by the GaAs substrate. One can see that on approaching the CNP GS^2 increases, goes through a maximum, and then reaches its minimum at the CNP, which has also been observed in our experiment (figure 5(b)). Note, however, that the theory is simplified and does not consider the asymmetry between electron and hole side near the Dirac point. The model reproduces the key feature of the thermopower, for example, the ambipolar behaviour of the signal, the linear temperature dependence and the nonmonotonic dependence on the length.

As we mentioned above, for narrow and long sample, when if $(L/w)\tilde{\sigma} \gg gL$, the contribution of the γ' and g' are equally important. Note , however, that observation of the nonlocal resistance (figure 5(a)) clearly demonstrates that the edge state transport inside of the mobility gap (-1.5 V $< Vg < 1.5$V). Let us assume that $gL \sim \gamma L \sim 1$. In this case the mixing between edge states and the bulk becomes important for probes 3-2 ($L/w = 10$) if $\sigma \gg 0.1e^2/h$. From our model [16] we estimated approximately equal value of the bulk conductivity $\sim 0.08e^2/h$). Note, however, that alternatively, this bulk conductivity may lead to bulk mechanism of the thermopower, considered below.

We can suggest that the bulk transport is important in our samples even in the vicinity of the CNP, despite the existence of a considerable nonlocal resistance signifying the presence of the edge transport. One would assume that our samples are characterized by a significant disorder level which results in the persistence of the bulk electron transport even inside the gap, so that the bulk component of the conductance exists in the whole gate voltage range. Near the center of the gap, the bulk transport may be of a hopping variety. A signature of such transport would be an exponential temperature dependence of the kind $R \propto \exp[(T_c/T)^{1/3}]$ corresponding to the Mott law for 2D carriers. However, we do not see such a dependence because the bulk conductivity is short-circuited by the edge states. On the other hand, the temperature behavior of the bulk thermopower is not expected to change considerably at the transition from the band transport to the hopping transport. For the band transport $S_b \propto T$. For the hopping transport, the thermopower was calculated in [26]. Applied to 2D electrons the result of [26] can be written as

$$S_b = -\frac{\lambda}{|e|}\left[\frac{\pi - 2}{\pi}(T_c^2 T)^{1/3} + \frac{2\pi}{3}T\right]\left(\frac{d\ln\rho_\mu}{d\mu}\right) \quad (13)$$

where λ is a numerical constant of the order of 1, ρ_ε is the density of states, T_c is a characteristic temperature proportional to $1/(\rho_\mu a_0^2)$, and a_0 is a localization length of the electron wavefunction. Since the Mott law is only valid for T_c considerably larger than T, the first term in equation (10) is important together with the second one. However, in the regime close to the onset of the band transport the localization length becomes large and T_c is of the order of T or smaller than T, so only the second term in equation (10) remains important and S_b is linear in T.

The presence of the logarithmic derivative of the density of states is quite a general feature, because the conductivity in the hopping transport regime is proportional to the square of the density of states. Any physically reasonable model of the density of states gives a dependence of S qualitatively similar to the experimental one. We approximate the density of states by the function:

$$\rho_e = \rho_0 + \frac{\varepsilon}{\pi A^2}, \quad \varepsilon > 0 \quad (14)$$

$$\rho_h = \rho_0 + \beta\frac{\varepsilon}{\pi A^2}, \quad \varepsilon < 0 \quad (15)$$

where the energy ε is counted from the center of the gap ε_g and ρ_0 is the constant background density of states describing localized states. The linear function $\varepsilon/\pi A^2$, where $A = 0.36$ eV nm for a well of width $d = 8$ nm, describes the band density of states at $\varepsilon > \varepsilon_g/2$ in the Dirac model with the electron spectrum $\varepsilon = \sqrt{(\varepsilon_g/2)^2 + (Ak)^2}$. To avoid discontinuities at $\varepsilon = \varepsilon_g/2$, the function $\varepsilon/\pi A^2$ is extended to the region $0 < \varepsilon < \varepsilon_g/2$. The equation (12) gives a reasonably good description of the density of states in the electron part of the spectrum. In the hole part, where the dispersion relation is more complicated and may include several subbands, no simple model exists. However, it can be roughly approximated by the same form, with an extra coefficient β ($\beta > 1$) describing the electron–hole asymmetry, which leads to equation (13).

Any physically reasonable model of the density of states results in a dependence of S qualitatively similar to the experimental one. For example, applying a Lorentz spectral function to describe the broadening of electron states, one obtains

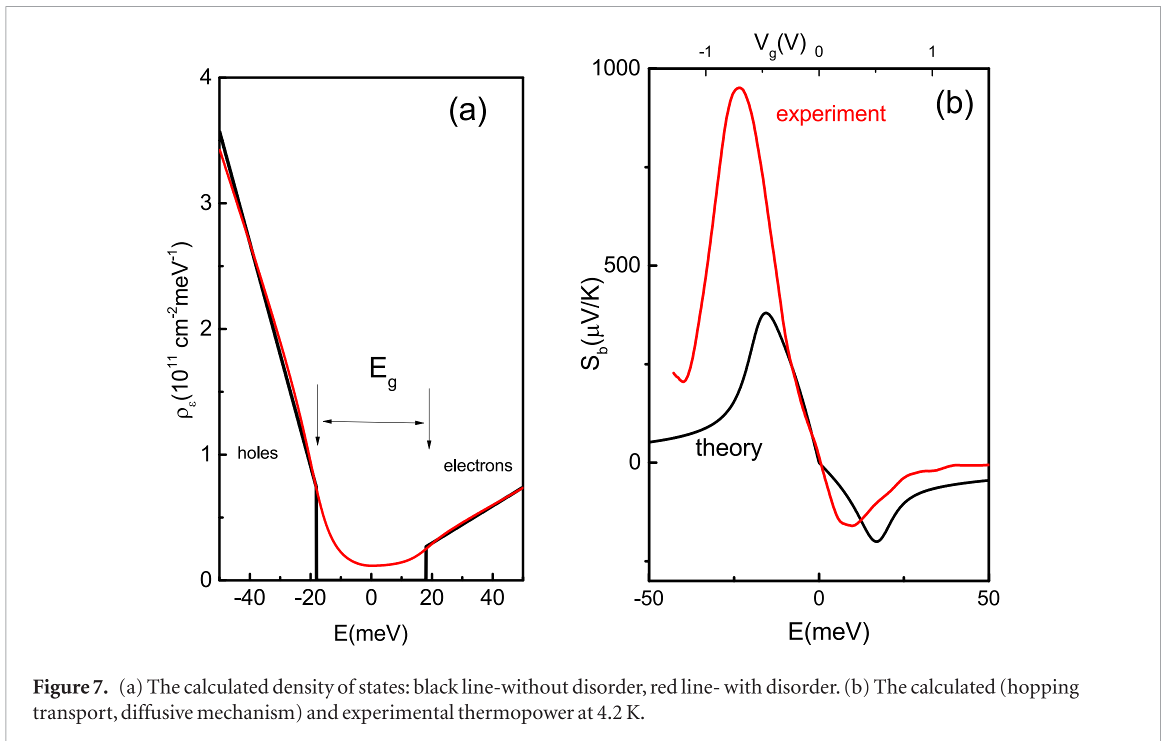


Figure 7. (a) The calculated density of states: black line-without disorder, red line- with disorder. (b) The calculated (hopping transport, diffusive mechanism) and experimental thermopower at 4.2 K.

$$\rho_e = \frac{1}{\pi A^2}\left[\varepsilon \arctan\frac{\varepsilon - \Delta/2}{\delta} + \varepsilon \arctan\frac{\varepsilon + \Delta/2}{\delta} + \frac{1}{2}\delta C_\varepsilon\right], \quad (16)$$

where

$$C_\varepsilon = \ln\left(\frac{(\varepsilon - E_c)^2 + \delta^2}{(\varepsilon - \Delta/)^2 + \delta^2}\right) + \ln\left(\frac{(\varepsilon + E_c)^2 + \delta^2}{(\varepsilon + \Delta/2)^2 + \delta^2}\right), \quad (17)$$

where Δ is the broadening energy, E_c is a large cutoff energy. In the hole region ($\varepsilon < 0$) the extra coefficient $\beta \approx 6$ is added, and the density of band states is larger. An example of the density of states calculated according to this expression is shown in figure 7(a) with disorder parameters $E_c = 150$ meV and $\delta = 6$ meV. Applying equation (11) with a scaling constant $\lambda \approx \pi$, one may plot the thermopower in the hopping transport regime as shown in figure 7(b). This calculation is in a reasonable agreement with experiment in the region close to the CNP. The numerical constant is closer to the metallic case. Strictly speaking, neither the hopping nor metallic transport are expected to work well in this intermediate regime. Still, it is important to get the idea of how the thermopower can depend on the electron density around the CNP. Note that the better agreement of the experimental results with equation (11) requires exact knowledge of the bulk conductivity behavior in the gap and hole side regions.

The temperature dependence of thermopower does not allow to separate the bulk transport from the edge transport, because $S_e \propto T$ as well. Nevertheless, there are several reasons to argue that the observed thermopower is mostly of the bulk origin. First, the asymmetry of the thermopower signal is much larger than the asymmetry of the resistance peak. The large asymmetry is associated with the bulk transport and is the consequence of electron–hole asymmetry in HgTe quantum wells (the asymmetry increases with the width d in the 2D TI regime $d > 6.3$ nm). Second, the absolute value of S_e, according to equation (10), is determined by the energy derivative of γ, which is not expected to be large (γ' is small if the backscattering rate is weakly sensitive to the momentum transfer). In contrast, the absolute value of S_b is not small even in the region of hopping transport, where S_b is determined by the energy derivative of the density of states according to equations (14) and (15). In the transition region between the hopping and band transport, S_b is expected to be large because both the density of states and the bulk conductivity σ strongly depend on the energy in this region. Thus, we expect $S_e \ll S_b$. The total thermopower in these conditions is reduced to $S \simeq S_b/(1 + G_e/G_b)$.

To conclude this chapter we should note that until now only few 2D topological insulators has been discovered: HgTe, InAs/GaSb based quantum wells, and recently—WTe_2 monolayer. Unfortunately in all these systems $e^2/2h$ in the conductance quantization has been observed in micron size devices, which indicates the presence of the backscattering between the counter propagating edge states. Thus, for observation of anomalous sign of the Seebek coefficient, predicted in the paper [18] it is important to develop a technology that reduces the impurity concentration in the bulk.

5. Summary and conclusions

In summary, we have studied the thermoelectric power together with the resistance behavior in HgTe quantum wells. The dependence of the thermopower on the gate-controlled carrier density, temperature, and device length has been investigated. The thermopower shows a behavior expected for electron–hole systems,

changing sign at the CNP, where the resistance reaches the maximum, and decreasing with carrier density increasing. The hole thermopower is much stronger than the electron one. The temperature dependence of the thermopower is close to linear.

Near the CNP, when the Fermi level lies in the gap, the resistance is comparable to or larger than $h/2e^2$. The bulk of the sample is likely to be localized under these conditions, and the resistance is determined mostly by the edge transport. Away from the CNP, the bulk diffusive transport takes place. In contrast, the thermopower appears to be mostly of the bulk origin, regardless to the position of the Fermi level. The transition from the localized states in the gap to the band conductance does not show the anomalies such as strong enhancement and sign variation of the thermopower recently suggested in the theoretical works [17, 18]. Our theory of a linear thermoelectric response in 2D TI explains the absence of these anomalies and supports the conclusion about the bulk origin of the observed thermopower.

Acknowledgment

The financial support of this work by the Russian Science Foundation (Grant No. 16-12-10041, MBE growth of HgTe QWs, fabrication of the field effect transistors and carrying out of the experiment and data analysis) and FAPESP (Brazil), CNPq (Brazil) is acknowledged.

ORCID iDs

G M Gusev https://orcid.org/0000-0003-3646-6024

References

[1] Hasan M Z and Kane C L 2010 *Rev. Mod. Phys.* **82** 2045
Qi X-L and Zhang S-C 2011 *Rev. Mod. Phys.* **83** 1057
[2] König M, Wiedmann S, Brune C, Roth A, Buhmann H, Molenkamp L W, Qi X-L and Zhang S-C 2007 *Science* **318** 766
[3] Roth A, Brüne C, Buhmann H, Molenkamp L W, Maciejko J, Qi X-L and Zhang S-C 2009 *Science* **325** 294
[4] Olshanetsky E B, Kvon Z D, Gusev G M, Levin A D, Raichev O E, Mikhailov N N and Dvoretsky S A 2015 *Phys. Rev. Lett.* **114** 126802
[5] Gusev G M, Kvon Z D, Shegai O A, Mikhailov N N, Dvoretsky S A and Portal J C 2011 *Phys. Rev.* B **84** 121202
[6] Knez I, Du R-R and Sullivan G 2011 *Phys. Rev. Lett.* **107** 136603
Knez I, Rettner C T, Yang S-H, Parkin S S P, Du L, Du R-R and Sullivan G 2014 *Phys. Rev. Lett.* **112** 026602
[7] Du L, Knez I, Sullivan G and Du R-R 2015 *Phys. Rev. Lett.* **114** 096802
[8] Nichele F, Pal A N, Pietsch P, Ihn T, Ensslin K, Charpentier C and Wegscheider W 2014 *Phys. Rev. Lett.* **112** 036802
[9] Suzuki K, Harada Y, Onomitsu K and Muraki K 2015 *Phys. Rev.* B **91** 245309
[10] Zuev Y M, Chang W and Kim P 2009 *Phys. Rev. Lett.* **102** 096807
[11] Wei P, Bao W, Pu Y, Lau C N and Shi J 2009 *Phys. Rev. Lett.* **102** 166808
[12] Cutler M and Mott N F 1969 *Phys. Rev.* **181** 1336
[13] Hwang E H, Rossi E and Das Sarma S 2009 *Phys. Rev.* B **80** 235415
[14] Ghahari F, Xie H-Y, Taniguchi T, Watanabe K, Foster M S and Kim P 2016 *Phys. Rev. Lett.* **116** 136802
[15] Xie H-Y and Foster M S 2016 *Phys. Rev.* B **93** 195103
[16] Rahim A, Levin A D, Gusev G M, Kvon Z D, Olshanetsky E B, Mikhailov N N and Dvoretsky S A 2015 *2D Mater.* **2** 044015
[17] Takahashi R and Murakami S 2010 *Phys. Rev.* B **81** 161302
Takahashi R and Murakami S 2012 *Semicond. Sci. Technol.* **27** 124005
[18] Xu Y, Gan Z and Zhang S-C 2014 *Phys. Rev. Lett.* **112** 226801
[19] Fletcher R, Maan J C, Ploog K and Weimann G 1986 *Phys. Rev.* B **33** 7182
[20] Chickering W E, Eisenstein J P, Pfeiffer L N and West K W 2010 *Phys. Rev.* B **81** 245319
[21] Vayrynen J I, Goldstein M and Glazman L I 2013 *Phys. Rev. Lett.* **110** 216402
[22] Wu S, Fatemi V, Gibson Q D, Watanabe K, Taniguchi T, Cava R J and Jarillo-Herrero P 2018 *Science* **359** 76–9
[23] Schmidt T L, Rachel S, von Oppen F and Glazman L I 2012 *Phys. Rev. Lett.* **108** 156402
[24] Kuroda M and Leburton J P 2008 *Phys. Rev. Lett.* **101** 256805
[25] Bernevig B A, Hughes T L and Zhang S C 2006 *Science* **314** 1757
[26] Parfenov O E and Shklyaruk F A 2007 *Semiconductors* **41** 1021