Current-induced breakdown of the quantum anomalous Hall effect

Gertjan Lippertz [1,2,*] Andrea Bliesener,[1,*] Anjana Uday [1,*] Lino M. C. Pereira [2]
A. A. Taskin [1,†] and Yoichi Ando [1,‡]

[1]*Physics Institute II, University of Cologne, Zülpicher Straße 77, 50937 Köln, Germany*
[2]*KU Leuven, Quantum Solid State Physics, Celestijnenlaan 200D, 3001 Leuven, Belgium*

The quantum anomalous Hall effect (QAHE) realizes dissipationless longitudinal resistivity and quantized Hall resistance without the need of an external magnetic field. However, when reducing the device dimensions or increasing the current density, an abrupt breakdown of the dissipationless state occurs with a relatively small critical current, limiting the applications of the QAHE. We investigate the mechanism of this breakdown by studying multiterminal devices and identified that the electric field created between opposing chiral edge states lies at the origin. We propose that electric-field-driven percolation of two-dimensional charge puddles in the gapped surface states of compensated topological-insulator films is the most likely cause of the breakdown.

I. INTRODUCTION

The quantum anomalous Hall effect (QAHE) [1] has been achieved in thin films of the topological insulator (TI) material $(Bi_xSb_{1-x})_2Te_3$ doped with Cr or V [1–3], where spontaneous magnetization M perpendicular to the surface opens an exchange gap in the two-dimensional (2D) surface states [Fig. 1(a)]. The hallmark of the QAHE is the dissipationless longitudinal transport accompanied by quantized Hall resistance of h/e^2. Since the chemical potential of the 2D surface states must be tuned into the exchange gap to realize the QAHE, the system is called a quantum anomalous Hall insulator (QAHI) in which chiral edge states are the only transport channels. The QAHI is a promising platform [4] for such phenomena as topological magnetoelectric effects [5] and chiral topological superconductivity [6].

However, the actual nature of the QAHI is still not well understood: For example, the nonlocal transport in the QAHE has shown deviations from an ideal behavior [7,8], which was explained by employing the coexistence of additional, quasihelical edge states [7–9]. Also, it has been observed that the dissipationless state breaks down with a much smaller critical current than in the integer quantum Hall effect (QHE), and the origin of this premature breakdown is under significant debate [10–14]. The breakdown current was reported to scale linearly with the sample width [11], which is problematic for applications of the QAHE in mesoscopic devices such as those to braid Majorana zero modes generated in the proximity-induced superconducting state of QAHI [15–17].

In this work, we investigate the breakdown of the QAHE, to be understood as the loss of dissipationless edge transport, in both local and nonlocal measurement geometries

to gain insights into its mechanism. Our detailed nonlocal transport data allow us not only to dismiss the contribution of additional edge states, but also to identify the transverse electric field E_{yx} as the most probable driving force for the breakdown, which is corroborated by the persistence of dissipationless edge transport in the nonlocal region above the breakdown current. To understand the very low breakdown current, we propose that percolation of charge puddles through the 2D bulk causes the abrupt onset of dissipation at the critical value of E_{yx}. We demonstrate that the effect of the current-induced breakdown can be minimized by spatially separating the source and drain contacts from the voltage probes, ideally in a nonlocal measurement configuration. We also show that the breakdown can result in a large spurious contribution in the three-terminal configuration with a narrow contact, which should be avoided in designing the experiments.

II. EXPERIMENTAL

The experiments were carried out on thin films of V-doped $(Bi_xSb_{1-x})_2Te_3$, grown by molecular beam epitaxy (MBE) on InP (111)A. The InP substrate was kept at 190°C while V, Bi, Sb, and Te were coevaporated to produce uniform films of ≈ 8 nm thickness. To tune the chemical potential into the magnetic exchange gap, a Bi:Sb beam-equivalent-pressure ratio of 1:4 was used. The films were protected from degradation in air by depositing a 3-nm Al_2O_3 capping layer using atomic layer deposition immediately after taking the samples out of the MBE chamber. The films were patterned into multiterminal Hall-bar devices using photolithography and chemical wet etching. Devices A, B, and C are long 26-terminal Hall-bar devices, while devices D to I are regular 6-terminal Hall-bars (see the Supplemental Material [18] for details). The metal contacts were fabricated by sputtering 5 nm Pt and 45 nm Au. All the reported films showed a clean QAHE without the need of gating.

*These authors contributed equally to this work.

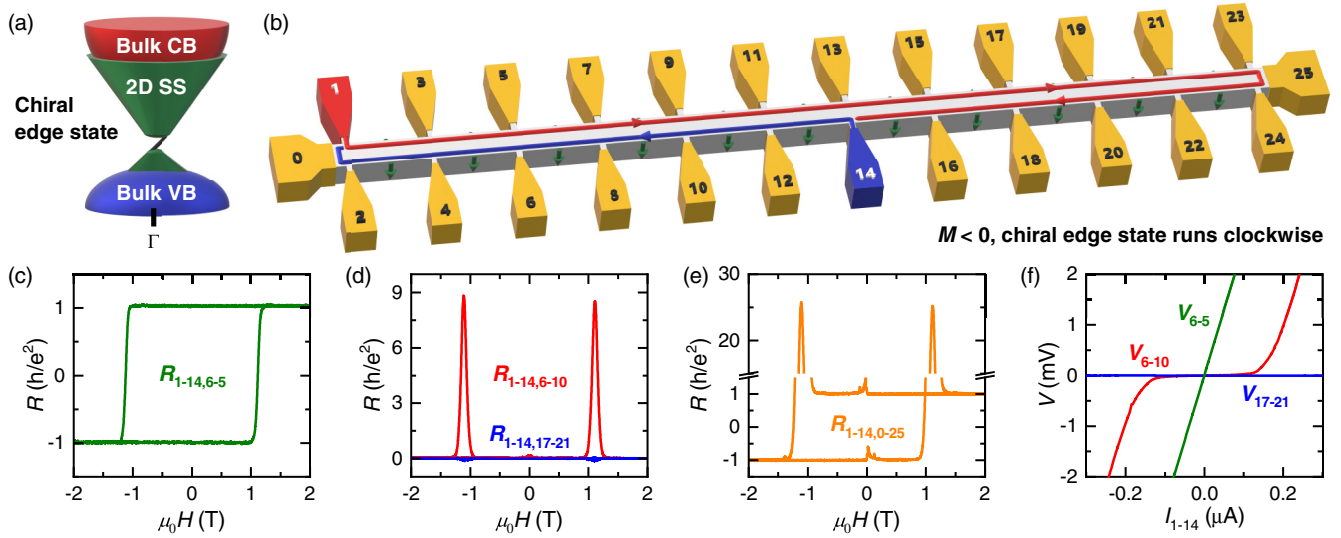†taskin@ph2.uni-koeln.de

‡ando@ph2.uni-koeln.de

FIG. 1. Local and nonlocal transport measured below the breakdown current in device A. (a) Illustration of the energy spectrum: Bulk conduction band (CB) and valence band (VB) are well separated and the 2D surface states (SS) have an exchange gap at the Γ point, but the chiral edge state connects the upper and lower branches. (b) Schematic picture of the 26-terminal Hall-bar device with the width of 100 μm. The red (blue) line shows the high (low) potential portion of the chiral edge state for a downward, out-of-plane magnetization ($M < 0$) when current flows from contact 1 to 14; arrow heads signify the direction of the current flow. Note that the chirality of the edge state inverts for $M > 0$. [(c)–(e)] Magnetic-field dependence of the resistance measured between contacts 6-5 (c), 6-10 and 17-21 (d), and 0-25 (e), showing the QAHE in the local and nonlocal regions at 15 mK with a DC current of 10 nA. (f) Four-terminal current-voltage characteristics in the local and nonlocal regions, measured at 15 mK in +2 T. The breakdown of the QAHE occurs in the local region (voltage between contacts 6-10) at $\approx 0.16\,\mu$A.

III. CURRENT-INDUCED BREAKDOWN

A. Dissipationless transport and breakdown

To clearly disentangle dissipative bulk current paths from the edge transport, we investigate 26-terminal Hall-bar devices, as shown in Fig. 1(b). In the measurement shown in Figs. 1(c)–1(e), the current flowed from contact 1 to 14; namely, a voltage is applied to contact 1 and contact 14 is grounded. As a result, the portion of the edge colored red (blue) in Fig. 1(b) is at the source (drain) potential due to the QAHE. From the point of view of local-nonlocal transport, contacts 0 to 14 belong to the local transport region while contacts 15 to 25 are in the nonlocal region. The four-terminal resistance $R_{1\text{-}14,6\text{-}10}$ and $R_{1\text{-}14,6\text{-}5}$ correspond to the longitudinal and transverse resistance, respectively; here, the first index denotes the current probes (1-14), while the second index denotes the voltage probes (6-10 or 6-5). The sample shows a clean QAHE with $R_{1\text{-}14,6\text{-}5}$ equal to the von Klitzing constant h/e^2 and a vanishing $R_{1\text{-}14,6\text{-}10}$.

The nonlocal resistance $R_{1\text{-}14,17\text{-}21}$ shows a near-perfect zero resistance throughout the magnetic field sweep. This indicates either a near-perfect nonlocal (and dissipationless) edge transport or the absence of current flow in the nonlocal region. If the latter is the case, the potential of contact 25 would always be equal to the drain potential of contact 14, and hence $R_{1\text{-}14,0\text{-}25}$ would be zero for $M < 0$. However, the observed $R_{1\text{-}14,0\text{-}25}$ is quantized to $-h/e^2$ for $M < 0$, meaning that the potential of contact 25 is equal to the source potential of contact 1. This proves that the nonlocal edge transport is realized.

Having demonstrated a clean QAHE and nonlocal transport, we now address the breakdown of the QAHE with increasing probe current. Figure 1(f) shows the voltages appearing at three different contact pairs as functions of the DC probe current $I_{1\text{-}14}$. The longitudinal voltage $V_{6\text{-}10}$ shows a broad plateau at 0 V up to $\approx 0.16\,\mu$A, above which a sharp increase signifies the breakdown of the dissipationless state. Note that this breakdown current value is among the highest reported so far [10–13]. The Hall voltage $V_{6\text{-}5}$ follows the expected linear behavior of the QAHE, $V_{6\text{-}5} \approx (h/e^2)I$, with only a little deviation at high current. Obviously, the breakdown in $V_{6\text{-}5}$ is much less pronounced than in $V_{6\text{-}10}$. In fact, a simple Landauer-Büttiker formalism [19] discussed in Sec. III C and in the Supplemental Material [18] leads to the conclusion that a small leakage current crossing the width of the Hall-bar directly affects the longitudinal resistance, while leaving the Hall resistance unaffected. This is consistent with the earliest studies of the QAHE, where the transverse resistance was close to h/e^2 while having a sizable longitudinal resistance of several kΩ [2].

It is important to notice that while the breakdown is clearly observed in the local transport region, dissipationless edge transport is maintained in the nonlocal region: As partly shown in Fig. 1(f), $V_{17\text{-}21}$ remains zero beyond $\approx 0.16\,\mu$A up to the maximum current employed (0.4 μA), to within the accuracy of our DC measurement ($\approx 1\,\mu$V). This points to the transverse electric field as the driving force of the breakdown of the dissipationless state, as was also suggested in Refs. [10,11], because the transverse electric field is absent in the nonlocal region where the edge potential is constant.
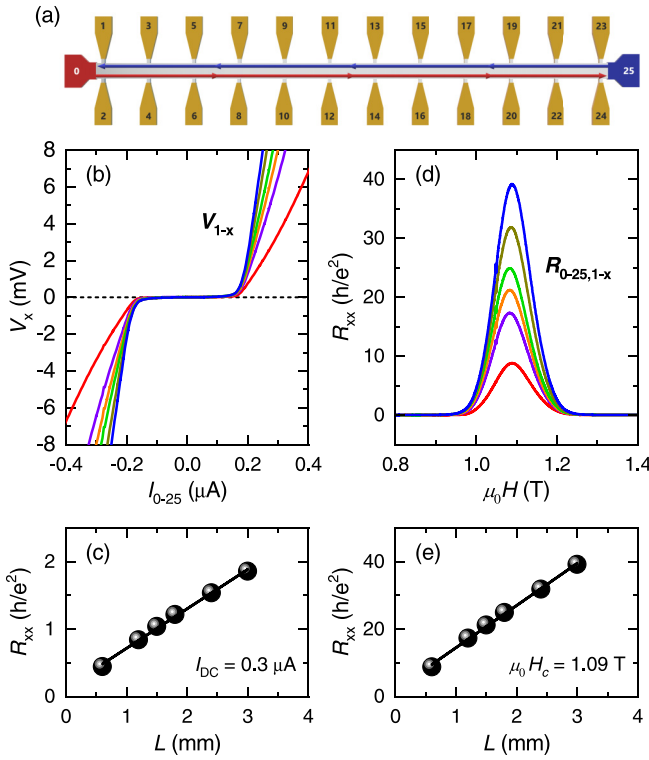
FIG. 2. Length dependence of the current-induced breakdown of the QAHE, measured on the 26-terminal Hall-bar device B at 40 mK. (a) Schematic of the potential distribution in the chiral edge state for $M > 0$; the red (blue) color signifies the source (drain) potential. (b) Plots of the longitudinal voltage V_{1-x} vs I_{0-25} with $x = \{5, 9, 11, 13, 17, 21\}$, measured at +2 T. The breakdown of the QAHE occurs at ≈ 0.16 μA. (c) The value of V_{1-x}/I_{0-25} at 0.3 μA as a function of the voltage contact spacing L. (d) Longitudinal resistance $R_{0-25,1-x}$ near the coercive field $\mu_0 H_c = 1.09$ T, measured with $I_{DC} = 30$ nA. (e) The peak values of $R_{0-25,1-x}$ at the coercive field H_c as a function of L.

The presence of edge current in the nonlocal region above the breakdown is shown in Ref. [18] with additional data (Fig. S8).

Wang *et al.* predicted an exponential length dependence of the edge potential in case quasihelical edge states are coexisting with the chiral edge state [9]. As seen in Fig. 1(f), we observed no sign of dissipation in the nonlocal region (see Ref. [18] for additional data). Moreover, when the I-V characteristics are recorded for a varying voltage contact spacing L, a clear linear dependence of the longitudinal resistance R_{xx} on L is observed above the breakdown [see Figs. 2(a)–2(c)]. The peak values of R_{xx} at the coercive field H_c are shown for comparison in Figs. 2(d) and 2(e), which also presents a linear dependence on L. The absence of a nonlocal position-dependent edge potential, together with the linear relation of R_{xx} on L, point to 2D diffusive transport as the origin of dissipation and speak against the presence of additional dissipative edge states as was proposed in Refs. [7–9].

B. Spurious contributions to the three-terminal resistance

Since the separation between the high- and low-potential branches of the chiral edge state directly determines the
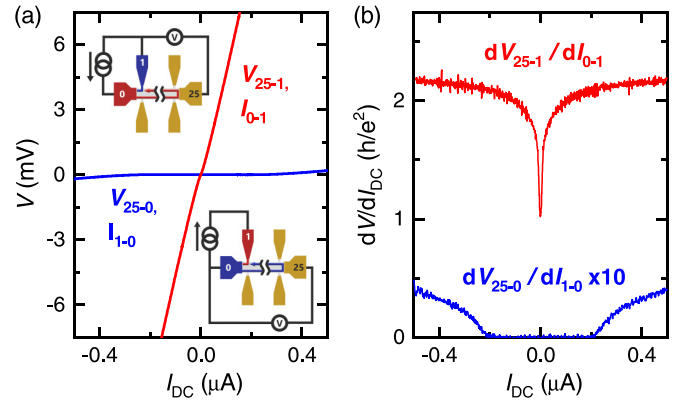

FIG. 3. (a) The three-terminal current-voltage characteristic of device B for contact 0 (100 μm, blue) and contact 1 (20 μm, red) with contact 25 as reference, measured at 40 mK in +2 T ($M > 0$). Insets show the measurement configurations. (b) Corresponding differential resistance for contacts 0 and 1. The curve for dV_{25-0}/dI_{1-0} is magnified by a factor of 10 for clarity.

strength of the transverse electric field, one would expect the breakdown to play a strong role near the source and drain contacts where the two branches come together. In our long Hall-bar device shown in Fig. 1(b), contacts 1 to 24 are made via a 20-μm-wide section of the magnetic TI film (see Fig. S1(b) in Ref. [18] for schematics), while contacts 0 and 25 are made along the full width (100 μm) of the device.

To elucidate the adverse effect of a narrow contact on the breakdown, we employed three-terminal measurements involving contacts 0, 1, and 25. Figure 3 shows the three-terminal I-V characteristics and corresponding differential resistance for two configurations with $M > 0$. The voltage V_{25-0} was measured with contact 0 as the drain (I_{1-0}). Since the edge current flows counterclockwise for $M > 0$, contact 25 was at the drain potential in this measurement. Indeed, the differential resistance dV_{25-0}/dI_{1-0} is approximately zero with an upturn at ≈ 0.16 μA due to the breakdown of the QAHE. Hence, there is no additional resistance associated with contact 0.

On the other hand, V_{25-1} was measured with contact 1 as the drain (I_{0-1}). Now contact 25 is at the source potential, and hence dV_{25-1}/dI_{0-1} should be equal to h/e^2. However, dV_{25-1}/dI_{0-1} immediately deviates from h/e^2 with a finite I_{DC} [see the red curve in Fig. 3(b)]. As shown in Ref. [18] and in Sec. III C using the Landauer-Büttiker formalism [19], this can be explained by the large electric field appearing in the narrow contact arm, causing an immediate breakdown in the contact arm and enhancing the three-terminal resistance [see Eqs. (5) and (6) below]. This demonstrates the necessity to avoid a three-terminal configuration with a narrow contact to minimize the breakdown effect. For example, in a recent study using an $\approx$200-nm-wide Nb electrode on top of a QAHI film, the breakdown of the QAHE was the dominant contribution to the measured conductance [20], making it difficult to detect the Andreev reflection at the Nb/QAHI interface.

C. Breakdown in the Landauer-Büttiker formalism

As discussed in detail in Ref. [18], the effect of breakdown on the measured three- and four-terminal resistances can
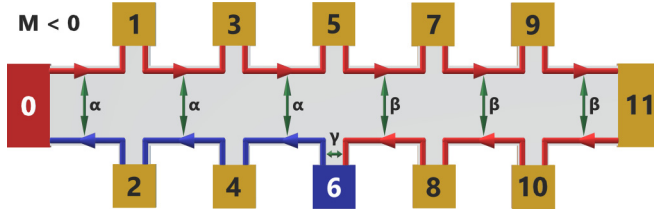
FIG. 4. Schematics of the edge transport in a 12-terminal Hall-bar device for $M < 0$, with the leakage between opposing counter-propagating edges parametrized by the scattering probabilities α, β, and γ. The contact spacing between neighboring contacts are chosen to be equal, i.e., $L_{0\text{-}1} = L_{1\text{-}3} = L_{3\text{-}5} = L_{5\text{-}7} = L_{7\text{-}9} = L_{9\text{-}11}$.

be phenomenologically understood by using the Landauer-Büttiker formalism [19]. Here, we briefly discuss the extension of the Landauer-Büttiker treatment to the nonlocal regime by considering the situation shown in Fig. 4, which schematically models a 12-terminal Hall-bar device for $M < 0$ with the current flowing from contact 0 to 6, similar to the setup used in Fig. 1. The scattering probabilities α, β, and γ describe the fraction of the current leaking between the opposite edge states in the local region, nonlocal region, and current contact 6, respectively. This simplistic model by no means describes the physics of the current-induced breakdown accurately, but it can provide a qualitative expressions for the observed resistances above the breakdown.

We use the Landauer-Büttiker formula [19]

$$I_i = \frac{e^2}{h} \sum_j (T_{ji} V_i - T_{ij} V_j), \quad (1)$$

where V_i is the voltage on the ith contact, I_i is the current flowing through the ith contact into the sample, and T_{ji} is the transmission probability from the ith to the jth contact. The full set of transmission coefficients for the case of Fig. 4 is given in Ref. [18]. Note that the spacing between neighboring contacts are chosen to be equal for simplicity, since α, β, and γ are length dependent. Using $V_6 = 0$, $I_0 = -I_6 = I$, and $I_i = 0$ for all other contacts, Eq. (1) can be solved for I and V_i, which allows us to calculate the longitudinal (R_{xx}), transverse (R_{yx}), and nonlocal (R^{NL}) resistances as follows:

$$R_{xx} = R_{0\text{-}6,1\text{-}3} = R_{0\text{-}6,2\text{-}4} = R_{0\text{-}6,3\text{-}5} = \frac{\alpha}{1-\alpha}\frac{h}{e^2}, \quad (2)$$

$$R_{yx} = R_{0\text{-}6,2\text{-}1} = R_{0\text{-}6,4\text{-}3} = -\frac{h}{e^2}, \quad (3)$$

$$R^{\text{NL}} = R_{0\text{-}6,5\text{-}7} = R_{0\text{-}6,5\text{-}8} = \cdots = 0. \quad (4)$$

Note that R^{NL} is the resistance between any pair of contacts in the nonlocal region (contacts 5, 7, 8, 9, 10, 11) for the current flowing from contact 0 to 6. The results for R_{xx} and R_{yx} are the same as those for the simple four-terminal geometry discussed in Ref. [18]. The result $R^{\text{NL}} = 0$ means that additional resistive channels in the middle of the nonlocal region (represented by β) do not carry any nonequilibrium current, as they are short-circuited by the dissipationless chiral edge channel.

If the resistive channels in the 2D bulk are allowed to carry some leak current to the drain contact, or additional, dissipative quasihelical edge states are included in the model,
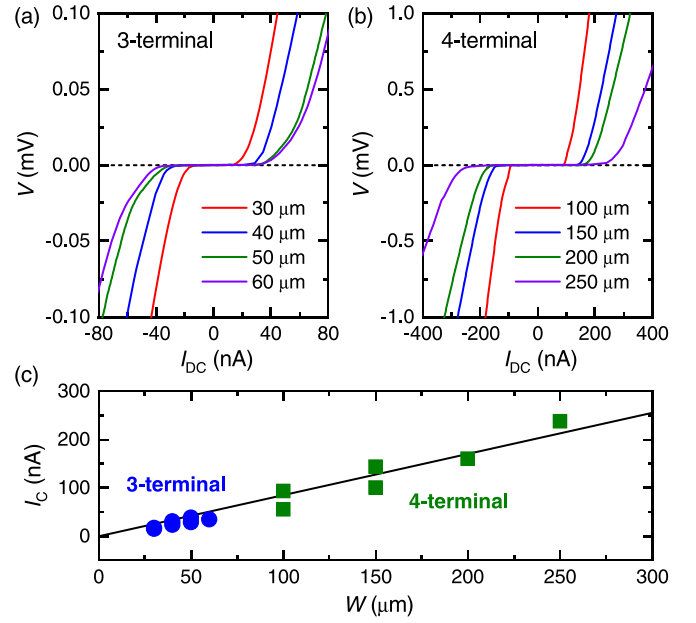

FIG. 5. Current-voltage characteristics showing the current-induced breakdown of the QAHE at 25 mK in 0 T (after training at +2 T), measured for various widths of the sample in the three-terminal (a) and four-terminal (b) geometries. The specified widths in panels (a) and (b) correspond to the contact and Hall-bar widths, respectively, of devices F2, G2, H, and I (cf. Table I in Ref. [18]). The devices were fabricated on the same wafer. (c) Corresponding width dependence of the critical current I_c. The black solid line is a linear fit through both data sets, yielding a slope of ≈ 0.85 nA/μm.

a position-dependent expression for the edge potential in the nonlocal region is obtained (see Ref. [18] for details). However, experimentally we found $R^{\text{NL}} = 0$, which indicates that the edge current is not lost in the nonlocal region through such leakage paths and the dissipationless edge transport is maintained in the nonlocal region.

In passing, the three-terminal resistances obtained from the solution of the Landauer-Büttiker formula for the case of Fig. 4 are

$$R_{0\text{-}6,0\text{-}1} = \frac{\alpha}{1-\alpha}\frac{h}{e^2}, \quad R_{0\text{-}6,0\text{-}2} = \frac{1}{1-\alpha}\frac{h}{e^2}, \quad (5)$$

$$R_{0\text{-}6,4\text{-}6} = \frac{\alpha - 2\alpha\gamma + \gamma}{(1-\alpha)(1-\gamma)}\frac{h}{e^2}, \quad R_{0\text{-}6,8\text{-}6} = \frac{1}{1-\gamma}\frac{h}{e^2}. \quad (6)$$

The differential resistances shown in Fig. 3(b) correspond to the case of $\gamma \gg \alpha$.

D. Width dependence of breakdown

To investigate the scaling of the critical current I_c on the edge state separation W, we fabricated four types of six-terminal Hall-bar devices F, G, H, and I on the same V-doped (Bi$_x$Sb$_{1-x}$)$_2$Te$_3$-InP wafer (see Table I in Ref. [18] for details) with their Hall-bar (contact) widths equal to 100 (30), 150 (40), 200 (50), and 250 μm (60 μm), respectively. Figures 5(a) and 5(b) shows representative three- and four-terminal *I-V* characteristics for devices F2, G2, H, and I. One can see an approximately linear dependence of I_c on W in Fig. 5(c), which corroborates the assumption that the
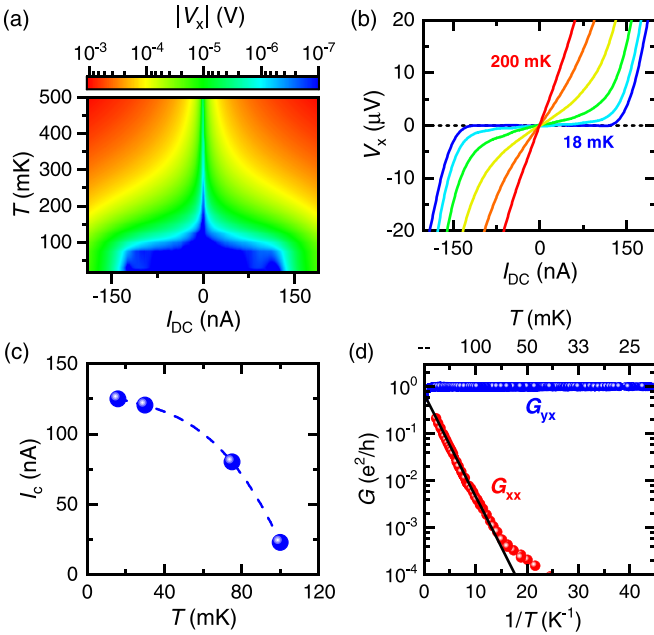
FIG. 6. Temperature dependence of the QAHE in 0 T after training at +2 T. (a) 2D color mapping of the longitudinal voltage $|V_x|$ as functions of T and I_{DC} for the 100-μm-wide device E. (b) V_x vs I_{DC} at 18, 100, 125, 150, 175, and 200 mK, showing the evolution of the current-induced breakdown curve. (c) Plot of I_c vs T; the dashed line is a guide to the eyes. (d) Arrhenius plot of the longitudinal conductance G_{xx} and transverse conductance G_{yx} of device D, measured with an AC rms current of 10 nA. Black solid line is a fit to the linear behavior $G_{xx} = G_0 e^{-T_0/T}$, yielding $T_0 \approx 0.5$ K.

transverse electric field governs the breakdown process. Namely, breakdown occurs at a particular value of the nominal critical current density

$$\frac{I_c}{W} \sim \frac{e^2}{h} E_{yx}, \tag{7}$$

independent of the device dimensions, as one can infer from Figs. 2(b) and 5(c).

E. Temperature dependence of breakdown

Now we turn to the temperature dependence of the breakdown effect. The six-terminal devices D and E were measured in the regular four-terminal Hall-bar geometry. Figure 6(a) shows a 2D mapping of the longitudinal voltage V_x as functions of temperature T and probe current I_{DC}; the plots of V_x vs I_{DC} at selected temperatures are shown in Fig. 6(b). A well-extended zero-voltage plateau is seen up to ≈ 100 mK, while a linear I-V relation is slowly restored at higher temperatures. The temperature dependence of I_c, which tends to saturate towards lower T, is plotted in Fig. 6(c).

Figure 6(d) shows that the thermal activation of charge carriers determines the conductance above ≈ 100 mK with a small activation energy of $k_B T_0 \approx 40$ μeV (i.e., $T_0 \approx 0.5$ K). This value is comparable to the values ≈ 17–121 μeV found in previous studies [8,10,11,21–23]. It is worthwhile to note that this activation energy is much smaller than the exchange gap of ≈ 14–28 meV observed in scanning tunneling spectroscopy [24,25], hinting at the role of disorder.

IV. DISCUSSION

We have demonstrated that the current-induced breakdown of the QAHE, to be understood as the loss of dissipationless edge transport, is most likely caused by the large transverse electric field E_{yx}, giving rise to 2D diffusive transport across the width of the Hall-bar device in the local transport region. In the following, we discuss the three most probable breakdown mechanisms: bootstrap electron heating, electric-field-driven percolation of 2D charge puddles, and Zener tunneling between 2D charge puddles.

A. Bootstrap electron heating

We first discuss the role of electron heating in the breakdown process. It is obvious that an abrupt increase in the longitudinal resistance will lead to heating and that the resulting increase in the electron temperature would accelerate the breakdown process. However, the assumption that electron heating itself lies at the origin of the breakdown of the QAHE is doubtful. In the bootstrap electron heating (BSEH) model, initially proposed for the integer QHE [26], the breakdown is attributed to runaway electron heating and described by the following balance equation:

$$\rho_{xx}(T_{el})j^2 = \frac{\epsilon(T_{el}) - \epsilon(T_L)}{\tau}, \tag{8}$$

with j the current density, $\epsilon(T)$ the energy of the system at temperature T, T_{el} and T_L the electron and lattice temperature, respectively, and τ the relaxation time of the heated electrons [26]. Upon increasing the current, the energy gained by electrons, $\rho_{xx}(T_{el})j^2$, causes Eq. (8) to become unstable and a new equilibrium is found at a higher T_{el}. The model is in good agreement with the experimentally observed critical current values of the QHE [26,27]. Moreover, it is quite generic and independent of the microscopic details of the samples. Hence, BSEH can be easily employed to describe the breakdown of the QAHE as well [11].

However, by comparing the critical current density j_c at which the breakdown occurs for both the QHE (≈ 1 A/m [28]) and the QAHE (≈ 1 mA/m), it is clear that the heating effect differs by several orders of magnitude. Even if one considers the much smaller excitation energy of the QAHE (≈ 40 μeV in our samples) compared to the QHE ($\hbar\omega_c \approx 10$ meV at 10 T, with ω_c the cyclotron frequency), runaway electron heating seems unlikely. Moreover, if one compares the shape of the breakdown curve of the QHE to that of the QAHE, the large vertical jump in the longitudinal resistance at j_c in the QHE case, attributed to the jump in T_{el}, is absent in the QAHE breakdown curves. Hence, while electron heating would accelerate the breakdown of the QAHE, it does not seem to be its origin.

B. Electric-field-driven percolation of 2D charge puddles

To understand the origin of the breakdown as well as the strongly reduced activation energy, it is useful to consider the role of charge puddles appearing in compensated TI materials [29–31]. It has been established that puddle formation in three-dimensional (3D) compensated TIs is an unavoidable consequence of the long-range nature of the Coulomb
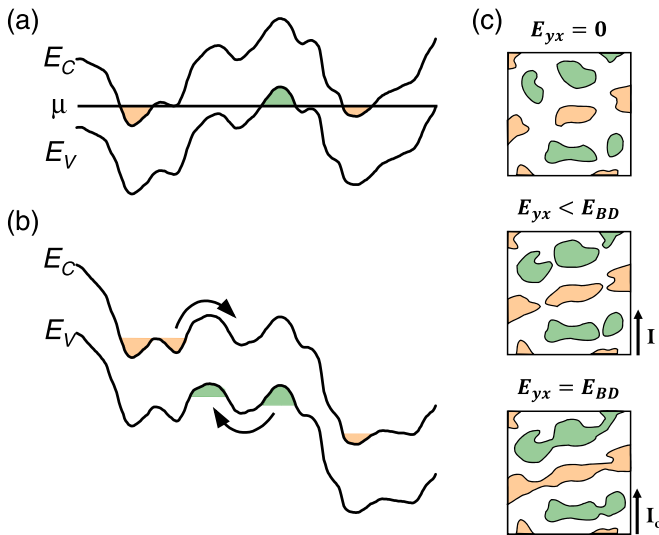
FIG. 7. Illustration of puddles in compensated QAHI films and the response to electric fields. (a) Spatial variation of the energy spectrum of the gapped 2D surface state. Meandering lines represent the 2D conduction and valence band edges, E_C and E_V, respectively, in the presence of Coulomb disorder. When the Fermi level μ crosses the band edges, electron and hole puddles (shaded regions) are created. (b) Situation close to the breakdown. Arrows depict thermally activated or hopping transport. (c) Growth of puddles driven by increasing electric field E_{yx} until breakdown occurs at $E_{yx} = E_{\text{BD}}$, based on Ref. [40]. The critical current I_c is reached with the critical source potential producing E_{BD} between two opposing edge states. The shaded regions correspond to electron and hole puddles in an insulating background.

interaction [32–34]. While the 3D bulk puddles are strongly suppressed near the surface due to the screening by the metallic surface states [35,36], 2D surface puddles are predicted to show up in compensated TI thin films [37]. In this regard, $(\text{Bi}_x\text{Sb}_{1-x})_2\text{Te}_3$ is a solid solution of *n*-type Bi_2Te_3 and *p*-type Sb_2Te_3, achieving a compensation to result in vanishing 2D surface carriers at low temperature [38]. In the QAHI films, the tendency to form 2D puddles would be strong, because the averaged chemical potential is tuned into the gap opened at the charge neutrality point. In such a case, little surface carriers are available to screen the Coulomb potential and the screening can only occur nonlinearly through the formation of 2D electron and hole puddles [32], as illustrated in Fig. 7(a). In addition, the large dielectric constant of TI films slows down the decay of the Coulomb potential in space and greatly enhances the puddle formation [37]. Indeed, signatures of puddle formation have been observed in the resistivity of ultrathin films of $(\text{Bi}_x\text{Sb}_{1-x})_2\text{Te}_3$ [39].

In the light of the likely existence of 2D puddles in compensated TI films, we propose that the breakdown occurs via the formation of metallic percolation paths connecting these 2D puddles across the width of the sample. The QAHI films can be thought of as an insulating background containing isolated metallic puddles, as shown in Fig. 7(c). In analogy to Ref. [40] for the QHE, we propose that at high enough source potential, the insulating regions separating two adjacent electron or hole puddles break down due to the high electric field created between source and drain potentials. Since the local potential is constant within the metallic puddles, the electric field is confined to the insulating regions. As the puddles grow with increasing source potential, the local electric field in between the puddles increases rapidly, facilitating further puddle growth in a nonlinear manner; see Figs. 7(b) and 7(c). At the critical value of the source potential corresponding to the critical current, the growth becomes unstable and leads to an avalanche process [40], so that the metallic paths percolate from one edge of the sample to the other and causes an abrupt onset of dissipation.

The puddle breakdown mechanism proposed here also sheds a new light on the very low temperature required to observe the QAHE, which is much lower than that expected from the Curie temperature T_C ($\approx$15–20 K) or the spectroscopically resolved exchange gap ($\approx$14–28 meV) [24,25]. In the presence of charge puddles, electrons are not excited across the 2D exchange gap; rather, electrons and holes are thermally excited from the puddles to the percolation levels [32]. As demonstrated already for 3D bulk puddles in compensated TIs, this reduces the activation energy for thermally-activated transport [32–34]. Moreover, at low temperature the electrons and holes may hop or tunnel directly between puddles, possibly giving rise to a crossover from activated transport to variable-range-hopping (VRH) behavior [32], as was observed in some transport studies on QAHI films [10,11]. The puddle scenario is also consistent with the bulk dissipation observed in Corbino geometry [14].

C. Zener tunneling between 2D charge puddles

Lastly, we comment on Landau-Zener tunneling between neighboring electron-hole puddles as a possible breakdown mechanism. Estimates of the tunneling probability for such a process in ultrathin TI films possessing a hybridization gap in the 2D surface state spectrum were made in Refs. [37,39]. Assuming a defect density of $N_{\text{def}} \approx 10^{19}$ cm^{-3}, an insulating state in our QAHI films is expected to be realized for a gap $\Delta > 10$–60 meV at the Dirac point. This is exactly the range of the exchange gap found for magnetically doped $(\text{Bi}_x\text{Sb}_{1-x})_2\text{Te}_3$ [24,25]. Note that in the presence of charge puddles, the application of an electric field is not required to induce Zener tunneling, because the disorder potential provides the required local electric field and band bending. In other words, in the presence of puddles, Zener tunneling would provide a finite bulk short even at infinitesimally small currents. As a result, no sudden onset of Zener tunneling at some critical current is expected.

It is prudent to mention that an estimate based on Refs. [37,39] would predict a sizable Zener tunneling in our films. However, a near-dissipationless QAHI state has been experimentally observed, at least for low probe currents, with a reported longitudinal resistivity value as low as 1.9 mΩ [11]. This speaks against any major role of Zener tunneling in compensated QAHIs. Nevertheless, Zener tunneling might be relevant to the small, nonvanishing resistance in the pre-breakdown regime.

V. CONCLUSION

We demonstrated that the breakdown of the QAHE, to be understood as the loss of dissipationless edge transport, occurs in the region with the shortest separation between the high- and low-potential branches of the chiral edge state, while it is absent in nonlocal transport regions. This indicates that the transverse electric field is responsible for the breakdown and gives a guiding principle for minimizing the breakdown effect. Moreover, we propose that charge puddles play a key role in the breakdown mechanism for the QAHE and govern the diffusive transport through the 2D bulk states.

Note added in proof. After submission, we became aware of the work by Fijalkowski *et al.* reporting nonlocal transport measurements on V-doped $(Bi_xSb_{1-x})_2Te_3$ in multiterminal Corbino devices [41].

[1] R. Yu, W. Zhang, H.-J. Zhang, S.-C. Zhang, X. Dai, and Z. Fang, Science **329**, 61 (2010).

[2] C.-Z. Chang, J. Zhang, X. Feng, J. Shen, Z. Zhang, M. Guo, K. Li, Y. Ou, P. Wei, L.-L. Wang, Z.-Q. Ji, Y. Feng, S. Ji, X. Chen, J. Jia, X. Dai, Z. Fang, S.-C. Zhang, K. He, Y. Wang *et al.*, Science **340**, 167 (2013).

[3] C.-Z. Chang, W. Zhao, D. Y. Kim, H. Zhang, B. A. Assaf, D. Heiman, S.-C. Zhang, C. Liu, M. H. W. Chan, and J. S. Moodera, Nat. Mater. **14**, 473 (2015).

[4] O. Breunig and Y. Ando, Nat. Rev. Phys. **4**, 184 (2022).

[5] X.-L. Qi, T. L. Hughes, and S.-C. Zhang, Phys. Rev. B **78**, 195424 (2008).

[6] X.-L. Qi, T. L. Hughes, and S.-C. Zhang, Phys. Rev. B **82**, 184516 (2010).

[7] X. Kou, S.-T. Guo, Y. Fan, L. Pan, M. Lang, Y. Jiang, Q. Shao, T. Nie, K. Murata, J. Tang, Y. Wang, L. He, T.-K. Lee, W.-L. Lee, and K. L. Wang, Phys. Rev. Lett. **113**, 137201 (2014).

[8] C.-Z. Chang, W. Zhao, D. Y. Kim, P. Wei, J. K. Jain, C. Liu, M. H. W. Chan, and J. S. Moodera, Phys. Rev. Lett. **115**, 057206 (2015).

[9] J. Wang, B. Lian, H. Zhang, and S.-C. Zhang, Phys. Rev. Lett. **111**, 086803 (2013).

[10] M. Kawamura, R. Yoshimi, A. Tsukazaki, K. S. Takahashi, M. Kawasaki, and Y. Tokura, Phys. Rev. Lett. **119**, 016803 (2017).

[11] E. J. Fox, I. T. Rosen, Y. Yang, G. R. Jones, R. E. Elmquist, X. Kou, L. Pan, K. L. Wang, and D. Goldhaber-Gordon, Phys. Rev. B **98**, 075145 (2018).

[12] M. Götz, K. M. Fijalkowski, E. Pesel, M. Hartl, S. Schreyeck, M. Winnerlein, S. Grauer, H. Scherer, K. Brunner, C. Gould, F. J. Ahlers, and L. W. Molenkamp, Appl. Phys. Lett. **112**, 072102 (2018).

[13] Y. Okazaki, T. Oe, M. Kawamura, R. Yoshimi, S. Nakamura, S. Takada, M. Mogi, K. S. Takahashi, A. Tsukazaki, M. Kawasaki, Y. Tokura, and N.-H. Kaneko, Appl. Phys. Lett. **116**, 143101 (2020).

[14] L. K. Rodenbach, I. T. Rosen, E. J. Fox, P. Zhang, L. Pan, K. L. Wang, M. A. Kastner, and D. Goldhaber-Gordon, APL Mater. **9**, 081116 (2021).

[15] C. W. J. Beenakker, P. Baireuther, Y. Herasymenko, I. Adagideli, L. Wang, and A. R. Akhmerov, Phys. Rev. Lett. **122**, 146803 (2019).

[16] I. Adagideli, F. Hassler, A. Grabsch, M. Pacholski, and C. W. J. Beenakker, SciPost Phys. **8**, 013 (2020).

[17] F. Hassler, A. Grabsch, M. J. Pacholski, D. O. Oriekhov, O. Ovdat, I. Adagideli, and C. W. J. Beenakker, Phys. Rev. B **102**, 045431 (2020).

[18] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.106.045419 for additional measurement data and discussion.

[19] M. Büttiker, Phys. Rev. B **38**, 9375 (1988).

[20] M. Kayyalha, D. Xiao, R. Zhang, J. Shin, J. Jiang, F. Wang, Y.-F. Zhao, R. Xiao, L. Zhang, K. M. Fijalkowski, P. Mandal, M. Winnerlein, C. Gould, Q. Li, L. W. Molenkamp, M. H. W. Chan, N. Samarth, and C.-Z. Chang, Science **367**, 64 (2020).

[21] A. J. Bestwick, E. J. Fox, X. Kou, L. Pan, K. L. Wang, and D. Goldhaber-Gordon, Phys. Rev. Lett. **114**, 187201 (2015).

[22] Y. Ou, C. Liu, G. Jiang, Y. Feng, D. Zhao, W. Wu, X.-X. Wang, W. Li, C. Song, L.-L. Wang, W. Wang, W. Wu, Y. Wang, K. He, X.-C. Ma, and Q.-K. Xue, Adv. Mater. **30**, 1703062 (2018).

[23] I. T. Rosen, E. J. Fox, X. Kou, L. Pan, K. L. Wang, and D. Goldhaber-Gordon, npj Quantum Mater. **2**, 69 (2017).

[24] I. Lee, C. K. Kim, J. Lee, S. J. L. Billinge, R. Zhong, J. A. Schneeloch, T. Liu, T. Valla, J. M. Tranquada, G. Gu, and J. C. S. Davis, Proc. Natl. Acad. Sci. USA **112**, 1316 (2015).

[25] Y. X. Chong, X. Liu, R. Sharma, A. Kostin, G. Gu, K. Fujita, J. C. S. Davis, and P. O. Sprau, Nano Lett. **20**, 8001 (2020).

[26] S. Komiyama and Y. Kawaguchi, Phys. Rev. B **61**, 2014 (2000).

[27] G. Nachtwei, Physica E **4**, 79 (1999).

[28] B. Jeckelmann and B. Jeanneret, Rep. Prog. Phys. **64**, 1603 (2001).

[29] N. Borgwardt, J. Lux, I. Vergara, Z. Wang, A. A. Taskin, K. Segawa, P. H. M. van Loosdrecht, Y. Ando, A. Rosch, and M. Grüninger, Phys. Rev. B **93**, 245149 (2016).

[30] T. Knispel, W. Jolie, N. Borgwardt, J. Lux, Z. Wang, Y. Ando, A. Rosch, T. Michely, and M. Grüninger, Phys. Rev. B **96**, 195135 (2017).

[31] O. Breunig, Z. Wang, A. Taskin, J. Lux, A. Rosch, and Y. Ando, Nat. Commun. **8**, 15545 (2017).

[32] B. Skinner, T. T. Chen, and B. I. Shklovskii, J. Exp. Theor. Phys. **117**, 579 (2013).

[33] B. Skinner, T. Chen, and B. I. Shklovskii, Phys. Rev. Lett. **109**, 176801 (2012).

[34] T. Chen and B. I. Shklovskii, Phys. Rev. B **87**, 165119 (2013).

[35] B. Skinner and B. I. Shklovskii, Phys. Rev. B **87**, 075454 (2013).

[36] T. Bömerich, J. Lux, Q. T. Feng, and A. Rosch, Phys. Rev. B **96**, 075204 (2017).

[37] Y. Huang and B. I. Shklovskii, Phys. Rev. B **103**, 165409 (2021).

[38] J. Zhang, C.-Z. Chang, Z. Zhang, J. Wen, X. Feng, K. Li, M. Liu, K. He, L. Wang, X. Chen, Q.-K. Xue, X. Ma, and Y. Wang, Nat. Commun. **2**, 574 (2011).

[39] D. Nandi, B. Skinner, G. H. Lee, K.-F. Huang, K. Shain, C.-Z. Chang, Y. Ou, S.-P. Lee, J. Ward, J. S. Moodera, P. Kim, B. I. Halperin, and A. Yacoby, Phys. Rev. B **98**, 214203 (2018).

[40] V. Tsemekhman, K. Tsemekhman, C. Wexler, J. H. Han, and D. J. Thouless, Phys. Rev. B **55**, R10201 (1997).

[41] K. M. Fijalkowski, N. Liu, P. Mandal, S. Schreyeck, K. Brunner, C. Gould, and L. W. Molenkamp, Nat. Commun. **12**, 5599 (2021).