A striking prediction in topological insulators is the appearance of the quantized Hall resistance when the surface states are magnetized. The surface Dirac states become gapped everywhere on the surface, but chiral edge states remain on the edges. In an applied current, the edge states produce a quantized Hall resistance that equals the Chern number $C = \pm 1$ (in natural units), even in zero magnetic field. This quantum anomalous Hall effect was observed by Chang et al. With reversal of the magnetic field, the system is trapped in a metastable state because of magnetic anisotropy. We investigate how the system escapes the metastable state at low temperatures (10 to 200 mK). When the dissipation (measured by the longitudinal resistance) is ultralow, we find that the system escapes by making a few very rapid transitions, as detected by large jumps in the Hall and longitudinal resistances. Using the field at which the initial jump occurs to estimate the escape rate, we find that raising the temperature strongly suppresses the rate. From a detailed map of the resistance versus gate voltage and temperature, we show that dissipation strongly affects the escape rate. We compare the observations with dissipative quantum tunneling predictions. In the ultralow dissipation regime, two temperature scales ($T_1 \sim 70$ mK and $T_2 \sim 145$ mK) exist, between which jumps can be observed. The jumps display a spatial correlation that extends over a large fraction of the sample.

**INTRODUCTION**

In the semiconductors known as topological insulators (exemplified by $\text{Bi}_2\text{Se}_3$, $\text{Bi}_2\text{Te}_3$, and $\text{Bi}_2\text{Te}_2\text{Se}$), strong spin-orbit interaction, together with band inversion, leads to electronic surface states that display a Dirac-like linear dispersion. On each surface, the Dirac states have only one spin degree of freedom, with the spin locked transverse to the momentum. The node of the Dirac cone is protected against gap formation if time-reversal symmetry (TRS) prevails (this constrains the nodes to be pinned at time-reversal invariant momenta).

The unusual properties of the surface Dirac states have been intensively investigated by angle-resolved photoemission spectroscopy, scanning tunneling microscopy, and transport experiments. A natural question is what happens when TRS is broken by inducing a magnetization at the surface? Theory posits that breaking of TRS opens a gap at the Dirac node everywhere on the surface (involving the surface “bulk” state as opposed to the edge states) (1–4). However, a conducting edge state remains, which runs around the perimeter of the sample. The edge state is chiral (it is either right- or left-moving depending on the magnetization vector) and dissipationless (backscattering of the electrons is prohibited). As in the quantum Hall effect (QHE), the edge state displays a Hall resistance that is rigorously quantized ($R_{\text{xy}} = \hbar e^2/|e|$, where $\hbar$ is Planck’s constant and $e$ is the electron charge), whereas the longitudinal resistance $R_{\text{xx}}$ vanishes (1–5). However, the quantization of $R_{\text{xx}}$ occurs even if the external magnetic field $H$ is 0. This phenomenon is known as the quantum anomalous Hall (QAH) effect. The quantization of $R_{\text{xx}}$ in zero field was first established experimentally by Chang et al. (6) using ultrathin films of Cr-doped $(\text{Bi,SB})_2\text{Te}_3$. Similar results in related topological insulator films were subsequently reported by several groups (7–13). The topological nature of the Hall quantization produces a hysteretic loop in $R_{\text{xy}}$ that reflects the magnetic hysteresis when $H$ (the applied magnetic field) is slowly cycled beyond the coercive field $H_C$. If the system is prepared with magnetization $\mathbf{M} \parallel \mathbf{H} \parallel -z$, the chiral edge modes lead to $R_{\text{xy}} = -\hbar e^2/25.812$ kilohms. This state is characterized by a Chern number $C = -1$. As $H$ changes sign, the system becomes metastable; to exit the metastable state, it undergoes a very sharp transition to the state with $C = 1$ and $R_{\text{xy}} = +\hbar e^2/25.812$. Here, we focus on the nature of this transition at millikelvin temperatures $T$. In conventional magnets, the hysteretic transition reflects the gradual diffusive motion of domain walls. Surprisingly, the transition actually proceeds by large, discrete jumps in $R_{\text{xy}}$. We find that slightly increasing $T$ suppresses the jump probability. A detailed investigation of how the jump probability varies with $H$ and $T$ suggests that the jumps reflect quantum tunneling events in the presence of dissipation.

**RESULTS**

**Characterization and benchmarks**

In the experiment, we used samples of Cr-doped $(\text{Bi,SB})_2\text{Te}_3$ grown on a SrTiO$_3$ substrate by molecular beam epitaxy (MBE) to a thickness of 10 nm and cut into Hall bars ($0.5 \text{ mm} \times 1 \text{ mm}$) [see Materials and Methods and the study of Kandala et al. (11)]. As grown, the surface states are $n$-doped, with the chemical potential lying high above the surface Dirac node. A large, negative back-gate voltage of $V_g = -80$ V is required to lower $\mu$ to the surface Dirac node. Within the optimal gate window $-120 \text{ V} < V_g < -80$ V (the limits are nominal; they change by $\pm 10\%$ in successive experiments), the Hall resistance $R_{\text{xy}}$...
attains near-ideal quantization at a temperature of $T = 10$ mK. Figure 1A shows that, at $H_c \sim 0.14$ T, $R_{yx}$ undergoes narrow transitions between −1 and +1, with narrow peaks in the resistance $R_{xx}$ of the width $\Delta H \approx 20$ mT (hereafter, we quote resistances in units of $\hbar/e^2$).

An important question is the stability of the quantized values $R_{yx} = \pm 1$. If the magnetization is actually not spontaneous, $R_{yx}$ should relax away from the quantized values when $H$ is returned to 0, whereas a truly magnetized ground state with $C = \pm 1$ should display no sign of relaxation. To test this, we prepared the state in positive field $H \parallel z$. We then returned $H \rightarrow 0$ and monitored $R_{yx}$ for a period of 8 to 9 hours. As shown in Fig. 1B, we do not detect any sign of relaxation at 49 and 72 mK for $V_g = −80$ V. To us, this shows that, at the Hall plateaus, the system is in the true ground state with $C = \pm 1$. Moreover, it remains stable in the limit $H \rightarrow 0$ for at least 8 to 9 hours.

The expanded scale in Fig. 1C shows that $R_{yx}$ deviates from 1.0000 by about four parts in $10^4$ within the optimal gate window. Significantly, $R_{xx}$ at $H = 0$ falls to values $< 3 \times 10^{-4}$ at 10 mK, consistent with ultralow dissipation. These benchmarks are comparable to the best achieved to date (9, 10).

To provide a broader view of the transport behavior at 10 mK, we show the curves of $R_{xx}$ and $R_{yx}$ versus $H$ in Fig. 2, with $V_g$ set at values from 0 to −120 V. The Hall curves $R_{xx}$ (Fig. 2A) remain very close to 1.0000 in the optimal gate window but start to fall monotonically as $|V_g|$ decreases, reaching 0.03 when $V_g = 0$ (surface states strongly $n$-doped). Throughout the gate interval, −70 to 0 V, the sign of the $n$-type carriers imparts a gentle slope to $R_{yx}$ for fields $|H| > H_c$. Turning to the longitudinal resistance, we show in semilog scale the curves of $R_{xx}$ versus $H$ for the same gate voltage values (Fig. 2B). As noted above, $R_{xx}$ falls to values as low as $3 \times 10^{-4}$ (in terms of $\hbar/e^2$) within the field window $|H| > H_c$ (see curve at $V_g = −90$ V). The small narrow spikes at $H = \pm 10$ mT are caused by the quenching of superconductivity in the indium contacts (in an experiment in which indium is replaced by gold in all contacts, we find that the spikes are completely absent). When $|H|$ is increased beyond $H_c$, the dissipation increases exponentially. As we tune $V_g$ outside the optimal gate window, the curves of $R_{xx}$ also increase exponentially, eventually reaching $\sim 0.2$ when $V_g$ approaches 0. The pattern of an exponential increase in dissipation when either $V_g$ is tuned outside its optimal window or $|H|$ is increased above the coercive field implies the existence of a small energy gap that “protects” the dissipationless state with accurately quantized $R_{yx}$. We estimate the activation gap below (see section Activation across a small gap).

Jumps in $R_{xx}$ and $R_{yx}$

The main results (the observation of large jumps in $R_{xx}$ and $R_{yx}$) are described in this subsection. At each $T$, we prepare the system with Chern number $C = −1$ ($H \parallel −z$). To investigate the transition $C: −1 \rightarrow 1$, we scan $H$ very slowly (1 to 10 mT/min) across $H_c$ (left to right in all figures). Starting at our lowest $T$ (10 to 50 mK), $R_{xx}$ displays a nominally smooth profile as it changes from $−1$ to 1 (Fig. 3A). The smooth variation, which has a profile that is $T$-independent in width up to 580 mK, is likely associated with conventional domain wall motion.

In the interval of 50 to 65 mK, small precursor steps become discernible above an onset field $H_c$. At the characteristic temperature

Fig. 1. Quantization of the Hall resistance in a 10-nm-thick film of Cr-doped (Bi,Sb)$_2$Te$_3$ at millikelvin temperatures. (A) Hall plateaus in the Hall resistance $R_{yx}$ at 10 mK (blue curves). The transition between Chern states at the coercive field $H_c (∼0.14$ T) is nearly vertical. The longitudinal resistance $R_{xx}$ (red curves) displays ultralow dissipation, apart from the sharp peaks at $H_c$. (B) Testing of the stability of the Chern state $R_{yx} = \hbar/e^2$ over an extended period (8 to 9 hours), with $H$ fixed at 0 starting at the time indicated by vertical arrows. The two traces at 49 mK (gray trace) and 72 mK (green) are shown displaced for clarity. (C) The expanded scale shows that $R_{yx}$ deviates from $\hbar/e^2$ by a few parts in $10^4$ in the optimal gating window $−110 < V_g < −90$. (D) The expanded scale shows that $R_{xx}$ drops to $< 3 \times 10^{-4}$ $\hbar/e^2$ near $H = 0$. $R_{xx}$ is thermally activated with a gap $\Delta_c \sim 190$ mK.
The jumps are also prominent in the dissipative channel (curves of $R_{xx}$). At low $T$ (10 to 50 mK), $R_{xx}$ exhibits a nominally smooth, asymmetric profile (Fig. 4A). Above $T_1$, large jumps appear at the leading (left) edge, whereas the trailing edge remains smooth, as the width increases to 38 mT. Figure 4B follows these curves above $T_2$ to 216 mK.

As noted for $R_{xy}$, the escape field $H_{ij}(T)$ of the initial jump increases with $T$. However, in the window 125 to 138 mK, the jumps now have opposite signs (positive at the leading edge). Jumps observed in $R_{xy}$ and $R_{xx}$ are closely similar when $H$ is swept in the opposite direction (figs. S2 and S3).

Figure 4 (C and D) shows how the two important quantities (the jump magnitude and the initial field) vary with $T$. As noted, the jump magnitude $\Delta R_{xy}$ is initially small within the precursor interval but increases rapidly at $T_1$ to saturate at a plateau value of $\sim 1$ (Fig. 4C). As $T$ is further raised above $T_2$, $\Delta R_{xy}$ collapses to a value below our resolution. Above $T_2$ (up to 500 mK), the Hall curves vary smoothly throughout the hysteresis loop. In contrast to the nonmonotonicity of $\Delta R_{xx}(T)$, the escape field $H_i(T)$ increases steeply from $T_1$ to $T_2$ (solid symbols in Fig. 4D). It is interesting that $H_i$ is still rising with steepening slope when the jump magnitude falls below resolution above $T_2$. For comparison, we also plot the gentle decrease of $H_j$ and $H_i$ (open symbols) over the same interval of $T$ (obtained from curves in fig. S4).

### Spatial correlations

An interesting way to study the jumps is provided by simultaneous measurements of $R_{ij}$ across well-separated pairs of voltage contacts. The resistance $R_{ij,mm}$ is defined as $V_{mn}/I_{ij}$, where $V_{mn}$ is the voltage measured across contacts $(m, n)$ with the current $I_{ij}$ injected at the source contact $i$ and drained at $j$ (see inset, Fig. 5A). A typical pattern is that a jump in $R_{xy}$ measured across one pair of contacts barely affects the signal across a different pair, but at a slightly larger $H$, the reverse holds.

Figure 5 (A and B) shows traces of $R_{14,53}$ and $R_{14,62}$ versus $H$ at $T = 105$ and 120 mK, respectively. In both panels, the initial jump (at $H = 0.129$ T) is large across the downstream Hall pair $(3, 5)$, whereas the upstream Hall pair $(2, 6)$ is barely affected. At a larger $H$ (0.133 T), the second jump is prominent in the upstream pair but much smaller in the downstream pair. Simultaneously, the resistance $R_{14,23}$ in the right pair $(2, 3)$ first increases and then decreases, whereas in the left pair $(5, 6)$, $R_{14,65}$ shows the opposite sequence (Fig. 5, C and D).

Despite the stochastic nature of the jumps, the overall pattern is systematic and reproducible, especially for the first two jumps in a trace. In Fig. 6, we plot four successive scans of the traces taken successively, with $T$ fixed at 86 mK and $V_g = -80$ V. The sample is prepared in the same way in each scan, with the sweep rate fixed at 1 mT/min. Figure 6A reports the Hall resistances (light curves for the upstream pair and bold curves for the downstream pair). In Fig. 6B, we show the corresponding longitudinal resistance curves (bold curves for the left pair and light curves for the right pair). The first two jumps are reproducible, although the jump magnitudes and $H_j$ can vary by 10%. The variations from one jump to the next shown here are typical of the reproducibility across the whole temperature range investigated (10 to 580 mK).

How fast are the jumps? To minimize extraneous perturbations, all the traces shown above were taken at very slow field sweep rates (typically 1 mT/min). We have attempted to estimate the duration of the first jump $\Delta t$ by minimizing the capacitive loading in the measurement.
Fig. 3. Jumps in \( R_{xy} \) at the transition between Chern states. (A) Traces of \( R_{xy} \) at selected \( T \) (10 mK to 70 mK) as \( H \) is slowly swept (10 mT/min) past \( H_o \), as indicated by the arrows. The transition between Chern states \( C = \pm 1 \) is nominally smooth below 40 mK. Starting near \( T_s \sim 65 \) mK, small vertical jumps appear near the start of the transition. (B) \( R_{xy} \) versus \( H \) in the important interval 70 mK \( \leq T \leq 145 \) mK, in which large jumps are clustered. Starting at 83 mK, the system “escapes” the \( C = -1 \) state by a very large initial jump, followed by a cascade of smaller ones. Focusing on the initial jump, we see that field \( H(\mu) \) that triggers the jump (kink feature) increases steadily as \( T \) rises to 142 mK. Above 131 mK, the jump magnitude \( \Delta R_{xy} \) sharply decreases, becoming unresolved above \( T_s \sim 145 \) mK. From \( T_s \) to 580 mK, \( R_{xy} \) changes smoothly over the (now broadened) transition. (D) Schematic of how the \( R_{xy} \) curves change with \( V_g \) at 10 mK. At optimal gaiting \( (V_g = -90 \) V), the transition is smooth, but as \( V_g \) is increased to \(-10 \) V, the jumps reappear, implying that a small amount of dissipation is necessary to seed the jump at the lowest \( T \). The experimental time constants are 1 s in all panels, except in (D), where the constant is 5 s.

circuitry in the dilution refrigerator. To this end, we sequentially removed several stages of the filter elements in the circuit (which increased the noise in the traces) while slowing the field scan to 0.1 mT/min. The final trace in this process is measured with only the low-temperature filter present. The noisy trace (Fig. 7) shows that \( \Delta t \) is shorter than 1 ms, the time constant of the remaining filter.

**Activation across a small gap**

Extending the measurements over a broad range of \( V_g \) and \( T \), we find that the ultralow dissipation state is protected by a minigap \( \Delta_R \sim 190 \) mK. Setting \( V_g \) within the optimal gating window (\(-110 \) to \(-75 \) V) places \( \mu \) inside \( \Delta_R \) (these numbers can shift by 10% between runs because of charge trapping at the interface). Once \( \mu \) exits the minigap, dissipation rises exponentially.

Figure 8A shows the changes in \( R_{xx} \) and \( R_{xx} \) as \( V_g \) is varied from \(-120 \) to \( 0 \) V, with \( T \) fixed at 10 mK and \( H \) fixed at \(-0 \) T (solid triangles) or 1 T (solid circles). The Hall curves (red and magenta symbols) show that, as \( |V_g| \) is decreased, \( R_{xx} \) stays close to the quantum value \( h/e^2 \) (the optimal gating window) until \( V_g \) decreases below 70 V in zero field. Thereafter, \( R_{xx} \) decreases exponentially to 0.03 as \( V_g \to 0 \). [The lower limit of the optimal window (\(-70 \) V in Fig. 8A) varies from run to run because of hystereses caused by charge trapping in the SrTiO\(_3\) substrate.] The resistance \( R_{xx} \) at \( H = 0 \) (blue triangles) increases exponentially with a lower threshold than \( R_{xy} \). A finite field (1 T; black circles) lowers the threshold significantly. We identify the exponential changes in both \( R_{xx} \) and \( R_{xx} \) with raising of \( \mu \) above the upper limit of the gap \( \Delta_R \) by gating.

The magnitude of the minigap \( \Delta_R \) is determined by detailed measurement of the conductance \( \sigma_{xx} \) versus \( T \). From the semilog plot of \( \sigma_{xx} \) versus \( 1/T \) in Fig. 8B, we infer that \( \Delta_R \sim 190 \) mK for \( H = 33 \) mT (black squares). At 0.5 T, the equivalent plot gives a slightly smaller \( \Delta_R \) (155 mK), consistent with the steeper increase in the curve with black symbols in Fig. 8A. Both plots imply that the minigap decreases noticeably with increasing \( H \). This is also consistent with the gradual increase of the background dissipation discussed in conjunction with Fig. 2B.

The changes to the transport curves induced by varying either \( V_g \) or \( T \) provide firm evidence that the ultralow dissipation state in the
QAH state is protected by a minigap $\Delta_R \sim 190$ mK. An exponential increase in the dissipation is observed when $T$ is increased or when we move $\mu$ out of the optimal window by gating. The size of the gap is also decreased by increasing $H$ beyond 0.5 T. The gradual increase of the background with $H$, together with the negative slope of $R_{yx}$ outside the coercive field window, suggests that the bulk carrier population ($n$-type) is gradually increased by $H$ when $\mu$ falls outside the optimal window. To summarize the results, we schematically plot the behavior of the curves $\Delta_R$, $T_1$, and $T_2$ in the $V_g$-$T$ plane (Fig. 9).

**DISCUSSION**

**Metastability and the effect of $T$**

The reversal of $H$ from the direction in which the system is prepared ($\mathbf{H} \parallel -z$) traps the system in a metastable state (spin $\uparrow$) until escape occurs near the coercive field. We investigated the nature of the transition in the magnetic topological insulator at millikelvin temperatures. In a conventional ferromagnet, escape from the metastable state corresponds to gradual motion of domain walls separating up and down domains, which allows the spin-$\uparrow$ domain to expand. This leads to the familiar smooth hysteretic $M$-$H$ loop. However, here, the transition that involves both a change of the Chern number ($C = -1 \rightarrow 1$) and the average $\mathbf{M}$ displays a rich array of additional features, most notably the large jumps in $R_{yy}$ and $R_{xx}$ in the interval $T_1 < T < T_2$ (Figs. 3 and 4). The jumps occur on time scales shorter than 1 ms. The first jump involves reversal of $\mathbf{M}$ over roughly half the sample. (As mentioned, we identify the smooth $T$-independent variation of $R_{xx}$ above $T_2$ with conventional domain wall motion. Below $T_2$, the smooth variation persists as a background channel between the jumps.)

In two recent investigations (12, 13) of the transition $C = -1 \rightarrow 1$, the variation of $R_{xx}$ is much broader than the ones observed here. Instead of jumps in $R_{xx}$ a shoulder at $C = 0$ (zero-Hall conductivity feature) is observed. Compared to the present sample (thickness, 10 nm), the samples used are in the ultrathin limit [5 nm in the study of Kou et al. (12) and 6 nm in the study of Feng et al. (13)], where hybridization between the states on the two surfaces is expected. Both the presence of zero-Hall shoulder and the absence of jumps suggest that these ultrathin samples are in a distinctly different regime of QAH behavior from the present sample. The degree of disorder appears to be higher.
in the ultrathin regime. For example, in the zero-Hall sample (13), the dissipation remains high at 50 mK \((R_{xx} \sim 4\) at \(H = 0\)), whereas accurate Hall quantization is not attained \((R_{xx} \sim 0.9)\).

We start by applying the standard Cahn-Hilliard nucleation theory to the expansion of a bubble of the true ground state (1). The free energy of a two-dimensional (2D) bubble of radius \(R\) is

\[
F_b(R) = 2\pi R\sigma - \pi R^2 \Delta F \quad (\Delta F = F_{\uparrow\downarrow} - F_{\uparrow\downarrow})
\]

where \(\Delta F > 0\) is the free energy gain in the bulk of the bubble. The surface tension \(\sigma\), which expresses the cost of the wall, is proportional to the exchange energy. When \(R\) exceeds the critical radius \(R_c = \sigma/\Delta F\), the bubble expands without limit (if dissipation is absent), happily paying the cost of the wall to gain bulk energy. Conversely, if \(R < R_c\), the bubble shrinks to 0. At the critical radius, \(F_b\) has the critical value \(F^c_b = \pi\sigma^2/\Delta F\). For a purely classical (thermally activated) escape process, the lifetime of the metastable state is

\[
\tau = \tau_0 \exp(F^c_b / k_B T)
\]

where \(1/\tau_0\) is an attempt frequency. Clearly, if \(\sigma\) and \(\Delta F\) are both \(T\)-independent (which we argue is the case here at all \(T < 0.5 K\)), the lifetime of the metastable state decreases exponentially with increasing \(T\).

Trapping of the system in a metastable state can be represented as a particle located in a potential well \(U(M)\), where \(M\) is a collective coordinate representing the local magnetization \(M(r)\) (see sketch in Fig. 4C, inset). After the system is prepared with \(H\parallel -z\) at temperature \(T\sim 100\) mK, it remains trapped for extremely long times if \(H = 0\). As shown in Fig. 1B, the stability for \(H = 0\) exceeds 8 hours at 10 mK. Incrementing \(H\) above \(H_c\) leads to a lowering of the barrier and a decrease in \(t_0\). When \(H\) is equal to \(H_E\), we infer that \(t_0\) (and hence a smaller \(H_c\) of 0.122 T at 70 mK). This is opposite to the observed trend of \(H_c\) increasing steeply with \(T\). This trend is counterintuitive if we view the escape as a classical process. In such “over the barrier” processes, raising \(T\) invariably speeds up the escape event. Quite generally, in the Langer formulation (15), an escape occurs when a particular spin configuration creates a saddle point in the potential landscape \(U(M(r))\) that breaches the barrier. As the probability for this to occur rises exponentially with \(T\), raising \(T\) should lead to a sharp reduction of \(t_0\) (and hence a smaller \(H_c\)). This is opposite to the observed trend of \(H_c(T)\). By measuring how \(H_c(T)\) varies with \(T\), we can deduce the effect of \(T\) on the escape rate.

The curve of \(H_c(T)\) (Fig. 4D) shows that \(H_c\) increases steeply with \(T\). This trend is counterintuitive if we view the escape as a classical process. In such “over the barrier” processes, raising \(T\) invariably speeds up the escape event. Quite generally, in the Langer formulation (15), an escape occurs when a particular spin configuration creates a saddle point in the potential landscape \(U(M(r))\) that breaches the barrier. As the probability for this to occur rises exponentially with \(T\), raising \(T\) should lead to a sharp reduction of \(t_0\) (and hence a smaller \(H_c\)). This is opposite to the observed trend of \(H_c(T)\). Contrary to the general expectation that “escape” should become easier as more spin configurations become accessible at higher \(T\), we find that escape is more difficult.
Fig. 6. Four traces of the resistances $R_{14,nn}$ measured in succession at 86 mK with $V_g = -80$ V. For each trace, the sample is prepared in the $C = -1$ state, and the transition is induced by sweeping $H$ slowly through $H_c$ at a rate of 1 mT/min while monitoring the resistances $R_{14,nn}$. (A) Upstream Hall resistance $R_{14,26}$ (light curves) and downstream Hall resistance $R_{14,53}$ (bold). (B) Corresponding longitudinal left contact resistance $R_{4,65}$ (bold curves) and right contact resistance $R_{14,27}$ (light plots). Although the jumps are stochastic, the overall pattern for the first two jumps is fairly reproducible from one run to the next.

Next, we consider a hybrid classical quantum process in which escape results from thermally assisted tunneling. Absorption of bosons (phonons or spin waves) from the bath brings the system’s energy closer to the top of the barrier, thereby enhancing the tunneling rate. Clearly, raising $T$ increases the phonon or spin-wave population, leading to an enhanced escape rate, again in conflict with $H_f(T)$.

**Dissipative quantum tunneling**

One way to obtain an escape rate that increases with decreasing $T$ is dissipative quantum tunneling. Here, dissipation as monitored by $R_{xx}$ plays an important role.

The quantum tunneling of a magnetization configuration out of a metastable state is closely related to the decay of the false vacuum investigated by Coleman (16, 17) and Callan and Coleman (16, 17). Through an instanton process (see section S3), the system tunnels “under the barrier” to form a small bubble of the true ground state. Thereafter, the bubble expands rapidly without limit (if dissipation is absent) in the classical Cahn-Hilliard process.

Dissipation—introduced by a linear coupling to many classical oscillators—appears as the damping coefficient $\eta(\omega)$, which may depend on frequency $\omega$. They find that increasing the dissipation strongly suppresses the tunneling probability $\Gamma$. In effect, each sampling (measurement) of the particle’s position by an oscillator collapses its wave function; this resets the tunneling process. Whereas $\eta$ by itself decreases $\Gamma$, raising $T$ has the opposite effect (20). The opposing effects of dissipation and $T$ have been investigated experimentally (21).

Within the optimal gating window in our experiment, $R_{xx}$ is thermally activated with an energy gap $\Delta_g$. Hence, the dissipation increases faster than $T$ (by a factor of 17 versus 4 between 50 and 200 mK). We reason that $\Gamma$ should decrease, consistent with the observed trend in $H_f$.

At very low $T$, changes caused by increasing $\eta$ compete with purely statistical effects due to $T$. Following Grabert et al. (20), the tunneling rate $\Gamma(T,\eta)$ at low $T$ may be expressed as

$$\Gamma(T,\eta) = \Gamma_0(\eta) \exp[A(T,\eta)] \quad \Gamma_0(\eta) \sim \exp(-B(\eta,0)/\hbar)$$

where $B(\eta,0)$ is the $\eta$-dependent exponent at $T = 0$. The $T$-dependent exponent is given by $A(T,\eta) = f(\eta)T^2$ (for ohmic dissipation). Although the analytic expression for $B(\eta,0)$ is not known, numerical integration (22) shows that, at very large $\eta$, $B(\eta,0)$ is asymptotically linear in $\eta$, but it approaches a constant as $\eta \rightarrow 0$. Because they are additive and have opposite signs, $B(\eta,0)$ and $A(T,\eta)$ compete in the exponent of $\Gamma(T,\eta)$. Unfortunately, theory provides only limited guidance to their relative magnitudes. We reason that at low $T$ and small $\eta$, $A(T,\eta)$ is very small compared with $B(\eta,0)$ (Grabert et al. (20) show that $A(T,\eta)$ vanishes exponentially like $\exp(-\hbar\omega/k_B T)$ in...
Fig. 8. Gate voltage dependence of $R_{xx}$ and $R_{yx}$ and the activation gap at different magnetic fields. (A) Semilog plot of $R_{xx}$ and $R_{yx}$ as a function of $V_g$ with $T$ fixed at 10 mK and $H$ fixed at either 0 T (solid triangles) or 1 T (solid circles). As $|V_g|$ decreases, the exponential increase in $R_{xx}$ (blue and black symbols) implies that $\mu$ is being raised above the minigap $\Delta_0$. Comparison between $R_{yx}$ at $H=0$ T (blue) and 1 T (black) shows that a finite $H$ decreases $\Delta_0$. Similarly, the curves for $R_{yx}$ (red and magenta) show an exponential deviation from the quantum value $h/e^2$ as $|V_g|$ is decreased. The Hall curves persist longer at the quantized value than do the resistance curves. (B) The semilog plot of $\sigma_{xx}$ versus $1/T$ with $V_g=-80$ V reveals an activated conductivity at both $H=0.5$ T and $H=33$ mT. The straight line fits yield values of $\Delta_0 \approx 190$ mK (at $H=0$) and 155 mK (0.5 T).

the limit $\eta \to 0$, where $\omega_0$ is the characteristic frequency at the well minimum. Hence, $\Gamma(T, \eta)$, that is, $\Gamma_0(T, \eta)$ is an overall decreasing function of $\eta$, with negligible dependence on $T$ (except insofar as raising $T$ in our situation causes $\eta$ to increase exponentially). As we raise $T$ between $T_1$ and $T_2$, we expect $\Gamma(T, \eta)$ to decrease, consistent with the trend in $H_0(T)$. Tests of macroscopic quantum tunneling have been carried out in experiments involving Josephson junctions (21, 23, 24), molecular magnets (25–27), and domain wall dynamics (28–30).

A new feature in the present experiment is that both the jump magnitude $\Delta R_{yx}$ and its spatial correlation also provide information on the rapid wall expansion that occurs after tunneling. Because dissipation diverts some of the energy gain, a large dissipation will severely restrict the final bubble size, over-damping the expansion (31, 32). To us, this provides a persuasive explanation of the behavior of $\Delta R_{yx}(T)$ observed above 120 mK (Fig. 4C). The magnitude $\Delta R_{yx}$ provides a measure of the size of the spin-up regions after the expansion corresponding to the first jump. In the interval 80 mK < $T$ < 120 mK, the expansion of the ground state extends to roughly half the sample. As $T$ rises above 120 mK, dissipation increasingly limits the expansion. Above $T_2$, the final radius (as measured by the jump size) falls below resolution. Even if tunneling persists (as implied by the continued increase in $H_1$ near $T_2$), the final bubbles are too far too small to be detected by $R_{yx}$. Raising $T$ in this interval opens a dissipative channel that sharply limits the size of the true ground state bubbles. [Because the activation gap protecting the quantized state is similar in magnitude ($\Delta g \approx 190$ mK; section Activation across a small gap), we propose that the bulk electrons are the source of dissipation limiting the wall expansion.]

Above $T_2$, the smooth variations of $R_{yx}$ versus $H$ (Fig. 3C) imply a gradual increase of the spin-up domains associated with domain wall motion as in conventional magnets. At all $T$ investigated, this smooth evolution is present as a background process that exists in parallel with the jumps of interest.

At lower $T$, a major puzzle uncovered is the steady decrease of $\Delta R_{yx}$ to very small steps when $T$ decreases below 70 to 10 mK for $V_g=-80$ V (Figs. 3B and 4C). The nonmonotonic profile of $\Delta R_{yx}$ versus $T$ implies that a second process that suppresses the tunneling rate becomes increasingly important at very low $T$. We emphasize that this suppression is confined to the ultralow dissipation state (the region $-120 < V_g < -80$ V in Fig. 9). Once dissipation is activated by gating, the jumps reappear at 10 mK. As shown in Fig. 3D, the curve at $V_g=-90$ V does not show jumps, whereas jumps reappear in the curves when $V_g$ is changed to $(-40, -10)$ V (all curves are at 10 mK).

It is instructive to view this pattern in the $V_g$-$T$ plane (Fig. 9). The reappearance of the jumps for $V_g>-40$ V implies that the boundary $T_1$ decreases rapidly from 70 to <10 mK once $V_g$ leaves the optimal gating window. Hence, a large gap $\Delta g$ (which strongly suppresses activation of bulk carriers) favors a large $T_1$, whereas a small $\Delta g$ decreases $T_1$ to below 10 mK. The pattern implies that the absence of jumps below $T_1$ is intrinsic to the ultralow dissipation state.

The observation suggests the following (tentative) picture. Because zero dissipation leads to a steep increase in the correlation length $\xi$ of the spins in the limit $T \to 0$, we reason that the tunneling rate $\Gamma$ will again be suppressed if $N_s$ (the number of spins that tunnel coherently) increases to macroscopic values (for example, the size of the whole sample). In the Wentzel-Kramers-Brillouin (WKB) approximation, where $\Gamma \sim \exp(-B/h)$ with $B = 2l \sqrt{\Delta m T(q)} dq$, with $q$ as a collective coordinate, a large $N_s$ or large mass $m$ leads to strong suppression of $\Gamma$.

When $\mu$ lies within the minigap, the bulk electrons become negligible in the limit $T \to 0$ so that $\xi$ extends over macroscopic regions and $\Gamma$ becomes very small. Increasing the dissipation by changing $V_g$ to $(-40, -10)$ V serves to reduce the size of the coherent region. Hence, jumps can reappear at 10 mK. Although this argument is speculative, the experiment potentially provides a way to measure $\xi$ and its effect on $\Gamma$ in the limit $T \to 0$.

Nucleation in superfluid helium

Another known example in which the escape rate increases with decreasing $T$ is the metastable $A$ phase of superfluid $^3$He produced by supercooling below the $AB$ transition. If $^3$He is cooled from the normal state to the superfluid state at a finite pressure $P$ (say, 30 bar), it first undergoes a second-order transition to the anisotropic $A$ phase at the critical temperature ($T_c \approx 2.3$ mK). At a slightly lower $T$ ($\sim 2$ mK), a first-order transition to the isotropic $B$ phase occurs. Because of the
first-order nature of the AB transition, the A phase is metastable as a supercooled liquid until it reaches ~0.7 $T_c$ at ~1.4 mK when bubbles of the B phase expand to fill the whole sample.

The Cahn-Hilliard theory (Eq. 1) predicts that, in 3D, $R_c = 2\sigma_{AB}/\Delta F$ and $F_B = (2\pi/3)^3 R^3 \Delta F$, where $\sigma_{AB}$ is the surface tension and $F_B$ is the condensation energy of the B phase. Using known quantities, one calculates $F_B$ to be extremely large ($10^9\rightarrow10^{10}$ $k_B T$ at 0.7 $T_c$). The predicted lifetime of the A phase $\tau_A$ ought to be many times the age of the universe. The observed nucleation temperature posed a deep puzzle until Leggett proposed the theory (dubbed “baked Alaska”) (33) that nucleation is triggered by cosmic rays. The strongly stochastic and often irreproducible nature of the nucleation process made experimental tests difficult. Subsequently, it was shown that $\tau_A$ is reduced by factors of 10 to 100 if the superfluid is exposed to gamma rays from a $^{60}$Co source, in strong confirmation of the cosmic ray theory (34). Experimentally, $\tau_A(T)$ is observed to decrease steeply with decreasing $T$ with the empirical form $\tau_A \sim \exp(a(R/R_0)^b)$, with $a = 3$ to 5 ($a$ and $R_0$ are fit parameters).

Despite the similarity of the lifetime trend versus $T$, the two experiments are quite different. In $^3$He, the extremely large barrier $F_B$ precludes thermal activation as a relevant factor at 1 mK. The $T$ dependence of $\tau_A$ arises because, in the vicinity of the $T_c$ curve, the critical divergence of $\sigma_{AB}(T)$ leads to a divergent increase in $R_c$ (hence, of $F_B$) as $T$ approaches $T_c$ from below. The primary effect of decreasing $T$ is to strongly decrease the barrier height $F_B$. Clearly, this depends on being close to the second-order transition curve $T_c$ (to have a strongly $T$-dependent $R_c$). In addition, we need an intervening first-order (AB) transition for supercooling to occur.

In the QAH system, extensive measurements of $R_{ij}$ versus both $H$ and $T$ [reported here and in previous studies (6–10)] have found no evidence for thermodynamic phase transitions (first or second order) below 4 K (the critical temperature for the ferromagnetic transition is in the range of 15 to 30 K). The barrier height in the metastable state (which depends on $H_c$ and the exchange energy) is controlled by $H$ but is essentially $T$-independent below 0.5 K. Because the knob ($H$) that changes the barrier height is independent of the knob that controls the dissipation ($T$), we can infer that the observed trend of $H_f(T)$ reflects the effect of dissipation on the escape rate. Both reasons persuade us that the $^3$He supercooling theory is not applicable here.

**Multiple paths and network model**

The pattern of correlation between Hall signals observed across pairs of Hall contacts separated by 100 µm (or more) suggests that the jumps strongly perturb the magnetization pattern. As an illustration, we start with a one-wall model that allows $R_{14;mn}$ to be calculated for assumed positions of the wall using the Büttiker formalism (see section S4). We calculate that the downstream Hall pair $R_{14;53}$ undergoes a jump but not the upstream $R_{14;65}$ (fig. S5). $R_{14;23}$ jumps up, whereas $R_{14;65}$ jumps down, in agreement with Fig. 5 (C and D). Next, we assume that the wall is further displaced to lie between the pair (2, 6) and contact 1 at the second jump. The calculated upstream Hall pair undergoes a jump, and the resistance pairs now reverse their jump directions, consistent with observation. However, the single-wall model is inadequate in many ways. Its biggest difficulty is that the predicted jumps $\Delta R_{yx}$ are always 2 (in units of $h/e^2$), whereas the observed magnitudes are roughly 1 or much smaller. This flaw appears to be intrinsic to the one-wall assumption.

We generalize to a multidomain picture in which tunneling starts from many seeds distributed across the sample. A natural starting point is the closely related plateau transition in the conventional integer QHE (IQHE), which has received considerable attention (see section S5) (35–37). As $H$ is varied, the chemical potential $\mu$ crosses the center of a Landau level broadened by the disorder potential $V(x)$. The current is diverted from the dissipationless edge modes to a very small subset of extended but dissipative bulk states at the band center (states away are Anderson-localized). In the bulk, the centers of the cyclotron orbits describe chiral trajectories that track the energy contours of $V(x)$. Tunneling occurs between chiral trajectories that get close to each other (this leads to scattering). Hence, the transition of $R_{yx}$ from one Hall plateau to the next occurs concurrently with the appearance of dissipation (finite $\sigma_{yx}$).
The network model studied by Chalker and collaborators (section S5) (35, 36) provides an elegant way to treat the flow of current through multiply connected paths. The sample is modeled as a regular lattice of corner sharing loops defined by directed links. At the shared corners (the nodes), incoming currents are scattered to one of the two outgoing links, with amplitudes parametrized by $\beta$ (eqs. S10 and S11 in section S5). They find that the localization length $\xi(\beta)$ diverges as $\beta - \beta_c$ (with $\nu = 2.5$, where $\beta = \beta_c$ at maximal dissipation (where $\sigma_{xx}$ peaks).

For the Chern transition in the QAH problem (eq. S13 in section S5), Wang et al. (38) have established the existence of three types of domains with Chern number $C = 0, \pm 1$. Identifying the Chern number domain walls with the energy contours in the IQHE problem, Wang et al. map the Chern transition to the network model of Chalker and colleagues and obtain a divergence in $\xi$ with the same exponent $\nu$. Wang et al. infer that $\sigma_{xx}$ obeys the two-parameter scaling relationship (39). Evidence for this scaling in the QAH problem has been obtained by plotting $\sigma_{xx}$ versus $\sigma_{yy}$ in the vicinity of the Chern number transition (7, 8). Evidence for the $\nu = 0$ QAH state has also been reported by Yoshimi et al. (40).

We discuss our results in this context. Just before the first jump in $R_{xy}$ occurs, the sample is predominantly in the $C = -1$ state (because $R_{xx}$ is finite and very small, small regions with $C = 0$ exist). After the first jump, which involves (by assumption) expansion from multiple seeds spread over the sample, the $C = 0$ regions define a multiply connected network of conducting paths that extend across the width of the sample. The observation that $\Delta R_{xy}$ is close to 1 (but never 2) suggests to us that the converted regions have connectivity between the chiral modes is mediated by tunneling at the nodes as in the network model (the microscopic tunneling between chiral modes occurring at the nodes is to be distinguished from the macroscopic quantum tunneling that triggers the jumps). The bulk conduction is dominated by a few backbone paths (as in percolation networks) that extend across the sample width to give $\Delta R_{xy} \sim 1$. We expect the sudden appearance of a network of regions with $C = 0$ to strongly perturb the system away from the two-parameter scaling behavior. This is borne out in an interesting way.

In Fig. 9B, we show plots of $\sigma_{xx}$ versus $\sigma_{yy}$ at selected $T$ for $V_g$ fixed at $-80$ V (within the optimal gating window). Above $T_2$ (curves at 142 and 152 mK), the measured $\sigma_{ij}$ trace out a nearly ideal semicircle (the dome). Within the interval ($T_1$, $T_2$), the first jump and subsequent ones kick the orbit high above the dome—the orbit returns only when the transition is nearly complete ($\sigma_{xx} \to e^2/h$). A notable pattern is that the perturbed orbits extend above the dome (the exception is a small sliver inside the dome for prejump segments at $T < 100$ mK). This rich pattern suggests that the two-parameter scaling behavior is valid above $T_2$, but below $T_2$, the jumps take the system high above the scaling dome.

Although two-parameter scaling cannot be used to obtain $\sigma_{xy}$ when jumps occur, it seems possible to calculate $\sigma_{xy}$ and $\sigma_{xx}$ numerically from large networks, as carried out for $\xi$ by Chalker and colleagues (35, 36). Calculations clarifying the conditions under which $\Delta R_{xy}$ is smaller than 1 will add valuable insight into what happens at the jumps.

### Outlook and conclusions

The central features uncovered are the large jumps in the Hall resistance of magnitude $\Delta R_{xy} \sim e^2/h$ and in $R_{xx}$ when the QAH system is trapped in the metastable state with $C = -1$, within the interval $T_1 < T < T_2$. A key finding is that the escape rate is sharply decreased as $T$ increases. From the thermal activation of $R_{xx}$, we have inferred that $\Phi$ increases in dissipation overwhelms any purely thermal effect associated with increasing $T$; the increased dissipation then suppresses the tunneling probability $T$. Hence, the Caldeira-Leggett theory seems to be the most appealing explanation for the jumps.

Because the transitions involve tunneling of a field $\phi(x, t)$ with many degrees of freedom (as opposed to one collective degree of freedom), the present experiment uncovers additional features that invite further investigation. The most interesting seems to be the existence of a temperature scale $T_1$ below which jumps are not observed within the experimental time scale, which we tentatively identify as yet another effect of dissipation. As shown in Fig. 9, the ultralow dissipation state is protected by a small energy gap $\Delta_F$ from thermal excitation of bulk electrons. The latter opens up a dissipative channel that strongly affects the tunneling process. Above $T_{2s}$, the dissipation strongly limits the expansion of the bubble after tunneling occurs.

A recent scanning nano-squid investigation of the magnetization on closely similar QAH films at $T$ down to 300 mK (41) has not detected the bubble expansion corresponding to our jumps. However, 300 mK is twice our limit ($T_2$ for resolving the jumps. Extending the squid experiments to below 100 mK (where $\Delta R_{xy} \sim 1$) will allow imaging of the macroscopic bubbles corresponding to regions transformed by the first jump in $R_{xy}$. We anticipate that the static size of the spin-up regions will increase rapidly to macroscopic lengths (100 $\mu$m) as $T$ is lowered from $T_2$ to 100 mK. A related experiment revealing the coincidence of superparamagnetism and quantization in the QAH problem was recently reported (42). Investigation of the correlations between jumps can be feasibly extended to a large number (~20) of contacts to provide a more detailed map of the spatial displacement of the wall.

The role of dissipation in macroscopic quantum tunneling (for example, in determining $N_{dd}$) remains a challenge for both theory and experiment. The QAH system, in which the chemical potential, dissipation, field strength, and temperature can be varied independently, provides a powerful system to explore these issues quantitatively.

### MATERIALS AND METHODS

Samples were grown by MBE using thermally heated single elemental (at least 5N purity) sources. The SrTiO$_3$ substrates were annealed in an oxygen environment between 925° and 965°C for about 2.5 hours and were checked after annealing with atomic force microscopy for an atomically ordered surface with a root-mean-square roughness of <0.3 nm before growth. Substrates were indium-mounted on the MBE sample holder and outgassed at 500°C for an hour in the MBE chamber before growth. Cr-doped (Bi,Sb)$_2$Te$_3$ was grown at 240°C as measured by a pyrometer, with a growth rate of about 0.5 quintuple layer per minute. After growth, the sample was allowed to cool to room temperature and was capped with about 3 nm of Al. The sample was then exposed to atmosphere to allow the Al capping layer to oxidize, forming Al$_2$O$_3$. Finally, the sample was heated in an argon environment to melt the indium and free the sample from its substrate. The remaining indium on the back of the substrate was then used for the backgate contact. For further details, see the work of Kandala et al. (11).

In the experiment, the Hall bar (width, 0.5 mm; length, 1 mm) was made by scratching the thin film with a sharp needle controlled by programmable stepper motors. The ohmic contact was made by

---


---

10 of 12
firmedly pressing tiny pieces of indium onto the contact pads. The pressing action was believed to help the indium penetrate through the Al2O3 capping layer and form a good contact with the underlying Cr-doped (Bi,Sb)2Te3 layer. The measurements were performed with the sample immersed in the 3He-4He mixture of a top-loading wet dilution refrigerator (Oxford Instruments TLM400). To filter out room temperature radiation, three filter stages were added to the circuit. At the room temperature end, an LC π filter provided more than 80-dB attenuation above 100 MHz. Two cryogenic filters were installed in the mixing chamber. One was a powder filter that filters out radiation above 1 GHz, whereas the second was an RC filter with a cutoff frequency of 300 kHz. The overall time constant of the whole circuit was 10 ms. Nearly all the results reported were taken using a Keithley 6221 precision current source and a Keithley 580 data acquisition module. The room temperature filter was removed. The voltage signal was amplified by the SR560 and/or 5408 preamplifiers.

**SUPPLEMENTARY MATERIALS**

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/2/7/e1600167/DC1

---

**REFERENCES AND NOTES**

Acknowledgments: We thank D. Goldhaber-Gordon, J. Wang, Y. Wang, and S.-C. Zhang for valuable discussions. Funding: We acknowledge support from the Defense Advanced Research Projects Agency/Space and Naval Warfare (N66001-11-1-4110) and a Multidisciplinary University Research Initiative award for topological insulators (ARO W911NF-12-1-0461). N.P.O. acknowledges support from the Gordon and Betty Moore Foundation EPiQS Initiative through grant GBMF4539. Author contributions: M.L., N.S., A.Y., and N.P.O. designed the experiment. A.R.R. and N.S. grew the samples. M.L., W.W., and A.K. performed the measurements. M.L., W.W., J.L., A.Y., and N.P.O. carried out the analysis and modeling of the results. M.L. and N.P.O. wrote the manuscript with inputs from all authors. Competing interests: The authors declare that they have no competing interests. Data and materials availability: The raw data for the curves in all figures are available in Dryad Correspondence, and requests for materials should be addressed to M.L. (minhaol@princeton.edu) and N.P.O. (npo@princeton.edu).

Large discrete jumps observed in the transition between Chern states in a ferromagnetic topological insulator
Minhao Liu, Wudi Wang, Anthony R. Richardella, Abhinav Kandala, Jian Li, Ali Yazdani, Nitin Samarth and N. Phuan Ong

Sci Adv 2 (7), e1600167.
DOI: 10.1126/sciadv.1600167

ARTICLE TOOLS http://advances.sciencemag.org/content/2/7/e1600167
SUPPLEMENTARY MATERIALS http://advances.sciencemag.org/content/suppl/2016/07/25/2.7.e1600167.DC1
REFERENCES This article cites 41 articles, 3 of which you can access for free http://advances.sciencemag.org/content/2/7/e1600167#BIBL
PERMISSIONS http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service