Letter

Nonlocal sidewall response and deviation from exact quantization of the topological magnetoelectric effect in axion-insulator thin films

N. Pournaghavi,[1] A. Pertsova,[2] A. H. MacDonald,[3] and C. M. Canali[1]
[1]*Department of Physics and Electrical Engineering, Linnæus University, 391 82 Kalmar, Sweden*
[2]*Nordita, KTH Royal Institute of Technology and Stockholm University, Roslagstullsbacken 23, SE-106 91 Stockholm, Sweden*
[3]*Department of Physics, University of Texas at Austin, Texas 78712, USA*

(Received 4 July 2021; accepted 25 October 2021; published 5 November 2021)

Topological insulator (TI) thin films with surface magnetism are expected to exhibit a quantized anomalous Hall effect (QAHE) when the magnetizations on the top and bottom surfaces are parallel, and a quantized topological magnetoelectric effect (QTME) when the magnetizations have opposing orientations (axion-insulator phase) and the films are sufficiently thick. We present a unified picture of both effects that associates deviations from exact quantization of the QTME caused by finite thickness with nonlocality in the sidewall current response function. Using realistic tight-binding model calculations, we show that in Bi_2Se_3 TI thin films, deviations from quantization in the axion-insulator phase are reduced in size when the exchange coupling of tight-binding model basis states to the local magnetization near the surface is strengthened. Stronger exchange coupling also reduces the effect of potential disorder, which is unimportant for the QAHE but detrimental for the QTME, which requires that the Fermi energy lie inside the gap at all positions.

DOI: 10.1103/PhysRevB.104.L201102

Introduction. In magnetoelectric materials, an applied electric field **E** induces magnetization **M** and an applied magnetic field **B** induces electrical polarization **P** [1]. The magnetoelectric response is described by the linear magnetoelectric polarizability tensor α, whose diagonal components,

$$\alpha_{ii} = \left.\frac{\partial M_i}{\partial E_i}\right|_{\mathbf{B}=0} = \left.\frac{\partial P_i}{\partial B_i}\right|_{\mathbf{E}=0} = \frac{\theta}{2\pi}\frac{e^2}{h}, \tag{1}$$

are pseudoscalars. [θ in Eq. (1) is dimensionless]. Since **B** and **M** are odd under time reversal and **E** and **P** are odd under space inversion, the magnetoelectric response normally occurs in insulators that break both time-reversal and inversion symmetry (TRS and IS, respectively) and is typically characterized by a small value of θ. We are interested here in the orbital magnetoelectric response [2–4] of three-dimensional (3D) topological insulators (TIs) [5,6], which is special in the sense that it is nonzero even when TRS is not broken in the sample bulk [7]. The magnetoelectric response is instead related to a nontrivial topological invariant of the bulk bands [2,3,6] and is quantized at $\theta = \pi$ ($\alpha_{ii} = e^2/2h$).

The quantized topological magnetoelectric effect (QTME) is realized only when the TI surface magnetization adopts an axion-insulator configuration [2,3], one in which all facets and hinges (where facets meet) of the bulk TI crystal surface are insulating [8]. For thin films this requires that the top and bottom surface magnetizations have opposite orientations [9,10], as shown in Fig. 1. When the magnetizations have the same orientation, the film displays [11] the quantum anomalous Hall effect (QAHE), and the side walls (hinges) are not gapped. Both the QAHE and the QTME can be understood qualitatively [9] by considering the limit of weakly gapped Dirac-cone surface states, since these give rise to half-quantized intrinsic anomalous Hall conductances $\sigma_H = \pm e^2/2h$ with a sign determined by the magnetization orientation [2,6,12]. The QAHE has been observed in uniformly doped magnetic TI thin films [13], in modulation-doped TI films [14–18] that have surface magnetism only, and recently also in TI films with proximity-induced 2D magnetism [19]. On the other hand, the QTME not yet been directly [20] confirmed experimentally, even though successful realization of the axion-insulator configuration is strongly suggested in some experiments [15–18,21] by the absence of a Hall effect in states with oppositely oriented top and bottom surface magnetizations. Novel intrinsic antiferromagnetic TIs, such as the van der Waals layered $MnBi_2Te_4$ [22–24] and $Mn_4Bi_2Te_7$ families [25], and their heterostructures with nonmagnetic TIs [25,26] have also been found to display the QAHE. These systems do not suffer from the intrinsic disorder of doped TIs and typically possess relatively larger magnetic gaps at their Dirac points.

In this Research Letter we employ a unified description of the QAHE and the QTME in TI thin films, by expressing the magnetization in terms of sidewall currents that respond nonlocally to electric potentials that vary slowly across the film. In this picture, perfect quantization of the QTME requires a sidewall response that is localized near the top and bottom surfaces, whereas the QAHE requires only bulk state localization. We characterize the nonlocality of the sidewall response by calculating finite-size corrections to the QTME theoretically using a realistic tight-binding model, demonstrating that they are smaller for stronger exchange coupling $J(\mathbf{r})$ between the surface magnetization and tight-binding model basis states localized near the surface. By increasing surface-state gaps,

stronger exchange coupling not only reduces finite-size corrections but also reduces the effect of disorder, which can be tolerated in QAHE measurements but is deleterious for the QTME.

QAHE, QTME, and sidewall response. We consider the linear response of the $\hat{z}$-direction orbital magnetization of a thin film with a quasi-2D bulk gap to an electric potential that varies slowly across the film. Since the bulk is time-reversal invariant and insulating, the magnetization response must originate from changes in currents that circulate around the film sidewalls: $M_z = (1/d)\sum_l I_{\rm sw}(l)$. Here, d is the thickness of the film, and we have anticipated our use of a tight-binding model by specifying the vertical position along the sidewall using a discrete layer index l. By allowing the l-dependent bulk electric potentials to turn on slowly upon entering the sample bulk [27] so that lateral electric fields are present only near the sidewall, we can relate the current response to bulk Hall conductivity:

$$I_{\rm sw}(l) = \frac{1}{e}\sum_{l'} \sigma_H(l,l')[V(l') - \mu], \tag{2}$$

where $V(l)$ is a layer-dependent electric potential and $\sigma_H(l,l')$ is the thin-film Hall conductivity generalized [28] to allow for nonlocality in the $\hat{z}$ direction [29],

$$\sigma_H(l,l') = \frac{2e^2}{\hbar}\sum_{n'\neq n} f_n \int \frac{d^2\boldsymbol{k}}{(2\pi)^2} \times \frac{{\rm Im}[\langle u_{n\mathbf{k}}|P_l v_x(\mathbf{k})|u_{n'\mathbf{k}}\rangle\langle u_{n'\mathbf{k}}|P_{l'} v_y(\mathbf{k})|u_{n\mathbf{k}}\rangle]}{(E_{n\mathbf{k}} - E_{n'\mathbf{k}})^2}. \tag{3}$$

In Eq. (2) we have introduced a chemical potential to allow for a unified discussion of the QAHE and QTME. In Eq. (3), f_n is a band occupation number, P_l is a layer projection operator, $|u_{n\mathbf{k}}\rangle$ is a band state of the 2D Bloch Hamiltonian $H(\mathbf{k})$, $E_{n\mathbf{k}}$ is the corresponding band energy, and $v_i(\mathbf{k}) = \partial H(\mathbf{k})/\partial k_i$, $i = x, y$, is the velocity operator. When summed over l and l', $\sigma_H(l,l')$ yields e^2/h times the total Chern number C of all occupied 2D bands and is quantized.

The QAHE measures [30] the response of the total sidewall current to a uniform chemical potential shift, and its quantization is therefore explained simply by the quasi-2D band Chern numbers. The QTME is a zero-temperature property of a state that is fully insulating, and is a response not to chemical potential but to electric potential. Its quantization can nevertheless be understood in terms of Chern quantization by the following argument. Define $\sigma_{t(b)} = \sum_{l,l'\in\, t(b)} \sigma_H(l,l')$, where each layer is classified by proximity as belonging to the top or bottom layer subset. For thick films the sidewall response must be localized where time-reversal symmetry is broken, i.e., near the top or bottom surface. It follows that for any configuration of the surface magnetism, $\sigma_H = \sigma_t + \sigma_b$. Since σ_H is quantized, its value must be independent of small variations in local properties, including variations in the strength of the exchange coupling, which occur only at one surface and can change only σ_t or σ_b. It follows that σ_t and σ_b must be separately universal. Since both must change sign when the magnetization is reversed at their surface, $\pm\sigma_t \pm \sigma_b$ must be quantized. In the special case of an axion insulator ($\sigma_H = 0$) it follows that $\sigma_b = -\sigma_t$, $2\sigma_t = ne^2/h$, and that for $n = 1$, $\delta M \equiv M(E_z) - M(0) \approx (1/ed)\sigma_t[V(l_t) - V(l_b)] = (e^2/2h)E_z$ when an electric field E_z is applied across the sample in the insulating state. $V(l_{t(b)})$ is the electric potential at the top (bottom) layer with layer index $l_{t(b)}$. (Note that for asymmetric exchange fields at the two surfaces, the magnetization is nonzero even when $E_z = 0$). Because the sidewall response has a finite localization length, the magnetization response has a finite-size correction that varies inversely with the number of layers N in the film and is characterized by the dimensionless number $m_{\rm corr}(d) \equiv |[\delta M(d=\infty) - \delta M(d)]/\delta M(d=\infty)| = (h/4e^2)\sum_{(l')\in t} \sigma_H(l_t - l')/N$.

Finite-thickness corrections in Bi_2Se_3. Finite-size corrections in thin films distinguish the QTME from the QAHE. To estimate their size in realistic systems, we have added a uniform electric field E_z applied across finite-thickness quasi-2D TI films to a realistic tight-binding (TB) model with surface magnetism and explicitly evaluated the magnetization carried by the distorted bands using [31,32]

$$M_z(E_z) = -\frac{e}{\hbar}\sum_{n,n'} f_n \int_{\rm BZ} \frac{d^2k}{(2\pi)^2} \frac{(E_{n\mathbf{k}} + E_{n'\mathbf{k}} - 2\mu)}{(E_{n\mathbf{k}} - E_{n'\mathbf{k}})^2} \times {\rm Im}[\langle u_{n\mathbf{k}}|v_y(\mathbf{k})|u_{n'\mathbf{k}}\rangle\langle u_{n'\mathbf{k}}|v_x(\mathbf{k})|u_{n\mathbf{k}}\rangle]. \tag{4}$$

Note that Eqs. (4) and (2) agree in the case of constant $V(l)$ since a constant electric potential shifts band energies without changing wave functions. Equation (4) gives the 2D bulk magnetization for an *infinite*-cross-sectional-area TI slab with broken TRS at the surfaces. The physical origin of the response of this magnetization to an electric field is the changes in the sidewall currents discussed above. This example of bulk-edge correspondence is closely analogous to that of the QAHE [30,33].

We focus on Bi_2Se_3 thin films [34], whose electronic structure can be described by an sp^3 TB model with parameters obtained by fitting to *ab initio* electronic structure calculations [35,36]. We apply this TB model to thin films with finite numbers of van der Waals coupled quintuple layers (QLs) and model broken time reversal at the top and bottom surfaces by adding exchange fields of strength $J_{\rm t}$ and $J_{\rm b}$, oriented orthogonal to the (111) surface, that couple to electron spin. We will consider two types of exchange fields: (i) a homogeneous field applied to the entire first surface QL, modeling magnetic modulation doping [14–18], and (ii) a homogeneous field applied only to the very top and bottom atomic monolayers (MLs), modeling the exponentially evanescent proximity effect of an adjacent magnetic layer [19,21,37].

In Figs. 1(b) and 1(c) we plot the band structures of 15-QL TI films for two strengths of exchange fields of type (i), both in the parallel (P) QAHE configuration, corresponding to the Chern-insulator phase. For symmetric exchange fields $J_{\rm t(b)} = 0.1$ eV (smaller than the bulk gap of $\approx$0.3 eV), the in-gap states on the two surfaces are essentially degenerate Dirac cones with exchange gaps $\Delta \approx J$ at the Dirac point (DP). In the following, we will refer to the surface states below the exchange gap as valence-band states and to those above as conduction-band states. An analysis of the wave functions [38] shows that, around the Γ point, these states are localized at either the top or the bottom surface, decaying exponentially

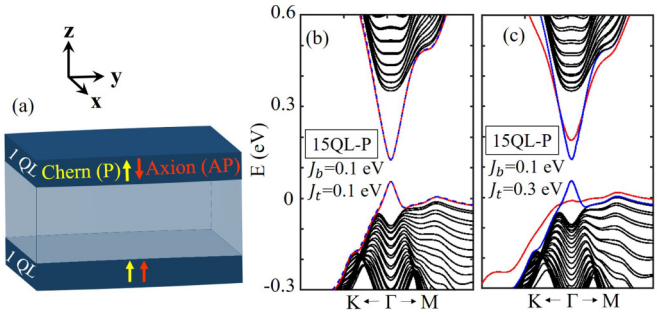

FIG. 1. (a) Schematic of a TI thin film with surface magnetizations at top and bottom in either parallel (P; Chern-insulator phase) or antiparallel (AP; axion-insulator phase) configurations. (b) and (c) Band structure of a 15-QL Bi_2Se_3 TI thin film in the P configuration for two choices of the top and bottom surface exchange fields J_b and J_t. The gapped Dirac surface states are indicated by red and blue lines. In the symmetric case (b) these states are nearly doubly degenerate and approach exact degeneracy in the thick-film limit.

within the first two QLs, just like the Dirac surface states of a nonmagnetic TI film [36]. For larger $\mathbf{k}$, the surface-state bands merge with bulk bands, and the corresponding wave functions are delocalized across the film [38]. On the other hand, when the exchange field at one surface is of the order of the bulk gap, as in Fig. 1(c), the Dirac-cone band structure at that surface is strongly modified; in particular, the valence band no longer resembles a gapped Dirac cone even near the Γ point. In fact, the band flattens but remains separated from the bulk bands for most $\mathbf{k}$ values. The corresponding wave functions are now localized at the strongly magnetized surface for a larger region of the Brillouin zone (BZ) [38]. For exchange fields of type (ii) (not shown in the figure), the structure of the gapped Dirac cones is robust and, apart from the increase in the exchange gap with $J_{t(b)}$, remains unmodified even for $J_{t(b)} > 0.3$ eV. These results demonstrate that the spatial distribution of $J(\mathbf{r})$ plays a separate role from its strength in influencing how the Dirac surface-state electronic structure is modified by surface magnetism.

We now consider the implications of these electronic structure properties for the magnetoelectric response. We compute M_z numerically as a function of E_z, represented in the TB Hamiltonian as an on-site energy varying linearly from the bottom to the top of the TI film [38]. By keeping $eE_z d$ smaller than the surface-state gap Δ in such a way that M_z depends linearly on E_z, we extract the θ parameter defined in Eq. (1). The results are shown in Fig. 2. From Fig. 2(a) we can see that, starting from five QLs, θ versus d is well described by the relation $\theta = \theta_{d\to\infty}(1 - w/d)$ [9], where $w \sim 2$ nm is a nonuniversal length scale that can be extracted from this figure, and measures the localization of the sidewall current response. Finite-size corrections are larger than 10% for film thicknesses below ∼20 nm. Figure 2(b) demonstrates that for all choices of the exchange strength $J_{b/t}$ and position dependence, $\theta_{d\to\infty} = \pi$ to within numerical accuracy ∼1%. That is, the magnetoelectric coefficient extrapolated to infinite thickness is, as expected, exactly quantized. Importantly, as shown in Fig. 2(a), the length scale w decreases with increasing $J_{t(b)}$. Finite-size corrections are reduced when the surface magnetization is strengthened.

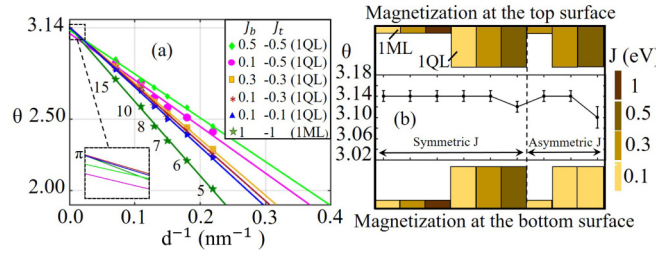

FIG. 2. Magnetoelectric coefficient θ extracted from the calculation of the magnetization in Bi_2Se_3 TI thin films of different thicknesses d and different values of the top and bottom surface exchange fields. "(1QL)" and "(1ML)" indicate that the exchange field is applied uniformly to the first surface QL or to the first surface monolayer, respectively. (a) θ vs $1/d$. (b) Asymptotic value ($d \to \infty$) of θ.

The quantization of the magnetoelectric response is consistent with the properties of the nonlocal Hall conductivity shown in Fig. 3, where $\sigma_H(l, l')$ is plotted as a function of the QL indices l and l' for two values of the exchange constants. As anticipated, $\sigma_H(l, l')$ is localized near the top and bottom surfaces, where TRS is broken. Furthermore, a careful numerical evaluation of $\sigma_{t(b)} = \sum_{l,l' \in t(b)} \sigma_H(l, l')$ shows that the larger $J_{t(b)}$ is, the more localized $\sigma_{t(b)}$ is at the surfaces.

To shed further light on the finite-size corrections, we define a QL-projected total Chern number $\mathcal{C}(l)$ in which the Berry curvature of each state is weighted by the projection of that state onto the lth QL:

$$\mathcal{C}(l) = \frac{1}{2\pi} \int_{\mathrm{BZ}} d^2k\, \Omega_{xy}^{(l)}(\mathbf{k}), \tag{5}$$

where

$$\Omega_{xy}^{(l)}(\mathbf{k}) = -2\mathrm{Im} \sum_{\substack{n=\mathrm{occ} \\ n'=\mathrm{unocc}}} \frac{\langle u_{n\mathbf{k}}|v_y(\mathbf{k})|u_{n'\mathbf{k}}\rangle\langle u_{n'\mathbf{k}}|v_x(\mathbf{k})|u_{n\mathbf{k}}\rangle}{(E_{n\mathbf{k}} - E_{n'\mathbf{k}})^2} W_{n\mathbf{k}}^{(l)}. \tag{6}$$

In Eq. (6), $W_{n\mathbf{k}}^{(l)} \equiv \sum_{s\in l} |\langle s|u_{n\mathbf{k}}\rangle|^2$ is the weight of $|u_{n\mathbf{k}}\rangle$ when projected on orbitals $|s\rangle$ centered at the sites s of the lth QL. The total Berry curvature does not single out the contribution of a particular quasi-2D band. However, when projected on a given QL, $\Omega_{xy}^{(l)}(\mathbf{k})$ and $\mathcal{C}(l)$ are a good measure of the contribution of the surface states relative to the contribution of the bulk states. The total Ω and $\mathcal{C}$, obtained by summing

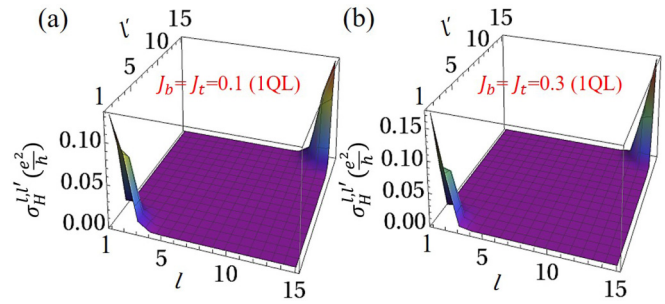

FIG. 3. Nonlocal response of Hall conductivity in 15-QL TI thin films for (a) $J_b = J_t = 0.1$ and (b) $J_b = J_t = 0.3$.

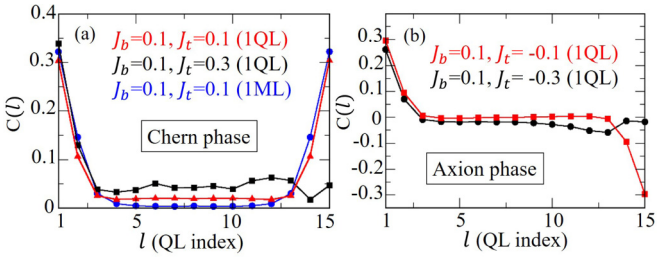

FIG. 4. QL-projected total Chern number for 15-QL TI thin films with different choices for the surface exchange fields J_b and J_t. (a) Chern-insulator configuration; (b) axion-insulator configuration. The total Chern number is dominated by sums over the three QLs closest to either surface only when J_b and J_t are both smaller than the bulk gap of ≈0.3 eV.

Eqs. (5) and (6) over l, yield $C = 1$ in the Chern-insulator state and $C = 0$ in the axion-insulator state, regardless of the number of QLs and the value of J.

In Fig. 4 we plot $\mathcal{C}(l)$ versus the QL index l for a 15-QL film for different values of $J_{b/t}$. Consistent with the band structures in Fig. 1, these results show that for $J_{b/t}$ substantially smaller than the bulk gap, the only states with substantial Berry curvature are those that derive from the nonmagnetic state Dirac cones, which are strongly localized in the first three QLs [38]. Note that the first three QLs represent the typical localization region of the evanescent surface states in nonmagnetic TI films [36]. In this case, we can operationally define a surface QAHE conductance by $\tilde{\sigma}_{H,t(b)} \equiv e^2/h \sum_{l=t(b)} \mathcal{C}(l)$, and we find that it is closer to $\pm e^2/2h$; in other words, the explanation for the topological magnetoelectric effect (TME) in terms of surface-localized half-quantized Hall conductivities applies literally. Only under these conditions do the definitions of a top and bottom surface Hall conductivity in terms of a projected Chern number $\mathcal{C}(l)$ and a nonlocal Hall conductivity $\sigma_{\rm H}(l, l')$ coincide. For very strong exchange potentials across a full quintuple layer, our calculations show that states with large Berry curvature have substantial weight deep in the bulk of the film [38]. This is particularly evident in the asymmetric case $J_b = 0.1$, $J_t = 0.3$ eV. Our explicit calculation of the magnetization response to electric fields nevertheless shows that finite-size corrections to the magnetoelectric response are actually smaller in this case, implying that the sidewall current response is even more concentrated in the layers of the film that have broken time-reversal symmetry, as evidenced by Fig. 3. For strong exchange interactions the localized sidewall response cannot be understood simply in terms of the Hall response of states localized at the surface.

Discussion. We have presented a unified analysis of the QAHE, the response of sidewall current to changes in chemical potential, and the QTME, the response of the magnetization associated with sidewall currents to changes in electric potential across the width of the film. The QAHE and the QTME are both of fundamental importance because of the direct relationship of these observables to Bloch state topology. Because the electric potentials in which we are interested are independent of lateral position, we can introduce them by adding layer-dependent lateral electric fields that are nonzero only near the film sidewalls. In this way we arrive at Eq. (2), which relates the magnetization and the sidewall currents to both electric potentials and chemical potentials via a Hall conductivity $\sigma_H(l, l')$ that is nonlocal across the film.

We have argued that in the thick-film limit the nonlocal Hall conductivity $\sigma_H(l, l') \approx 0$ when either l or l' are far from both the top and bottom surfaces and therefore in a region that is locally time-reversal invariant. Numerical evaluation of $\sigma_H(l, l')$ for the realistic TB model of a TI thin film reported on in Fig. 3 supports this statement, which allows us to separate the total Hall conductivity σ into top and bottom surface contributions $\sigma = \sigma_t + \sigma_b$. It follows from time-reversal symmetry that $\sigma_t(J_t, J_b) = -\sigma_t(-J_t, -J_b)$ and $\sigma_b(J_t, J_b) = -\sigma_b(-J_t, -J_b)$, where J_t, and J_b specify the exchange couplings at the top and bottom surfaces. We have argued that in the thick-film limit a locality condition also applies, namely, that σ_t depends only on J_t and σ_b depends only on J_b. This locality condition is certainly satisfied in the limit of weak time-reversal symmetry breaking where the exchange gap is considerably smaller than the bulk gap and the Hall conductivity is contributed by weakly gapped Dirac-cone surface states. It is less obvious, perhaps, that the locality condition is satisfied in the strong-(J_t, J_b) limit where we have shown that states with large Berry curvature are extended across the sample—although even here nonlocal response would be very surprising. The assumption of locality is supported by the fact that, when combined with quantization of the total Hall conductivity and time-reversal symmetry properties, it naturally accounts for the expected quantization value of the QTME $\sigma_t = \sigma_b = e^2/2h$, which we confirm numerically, and for the $(1 - w/d)$ form of finite-size corrections, with w being a nonlocality length of the sidewall response and d being the film thickness.

We have established the $(1 - w/d)$ finite-size law using explicit calculations of magnetization in the presence of an electric field applied across a Bi_2Se_3 topological insulator which yield $w \sim 2$ nm. If this is correct, the part in 10^6 quantization accuracy routinely achieved for quantum Hall systems would require films of approximately millimeter thickness. Stronger surface magnetism generates larger quasi-2D gaps, which in turn imply greater robustness of the quantized response against potential disorder that is inevitable and can invalidate quantization by inducing surface electron or hole puddles. It is therefore encouraging for QTME measurement efforts that finite-size corrections to the QTME are smaller for surface magnetism that is stronger. This can be achieved either in the sense of coupling more strongly to electron spins near the Fermi level or in the sense of being present over more near-surface layers of the film.

Acknowledgments. This work was supported by the Faculty of Technology at Linnæus University and by the Swedish Research Council under Grant No. 621-2014-4785. A.H.M. was supported by the Army Research Office under Grant No. W911NF-16-1-0472. We acknowledge valuable interactions with David Vanderbilt and Peter Armitage. Computational resources were provided by the Swedish National Infrastructure for Computing (SNIC) at Lunarc partially funded by the Swedish Research Council through Grant Agreement No. 2018-05973.

[1] M. Fiebig, J. Phys. D: Appl. Phys. **38**, R123 (2005).

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

[3] A. M. Essin, J. E. Moore, and D. Vanderbilt, Phys. Rev. Lett. **102**, 146805 (2009).

[4] A. Malaschevich, I. Souza, S. Coh, and D. Vanderbilt, New J. Phys. **12**, 053032 (2010).

[5] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).

[6] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).

[7] N. P. Armitage and L. Wu, SciPost Phys. **6**, 046 (2018).

[8] If only the top and bottom surfaces were magnetized, the sidewalls of a finite heterostructure would not be insulating since they would host conducting surface states. These states can, however, be gapped by (i) designing geometries in which an orthogonal exchange field is also present on the sidewall or (ii) exploiting energy level quantization by tapering off the thin-film thickness near the sidewalls.

[9] J. Wang, B. Lian, X.-L. Qi, and S.-C. Zhang, Phys. Rev. B **92**, 081107(R) (2015).

[10] T. Morimoto, A. Furusaki, and N. Nagaosa, Phys. Rev. B **92**, 085113 (2015).

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

[12] L. Fu and C. L. Kane, Phys. Rev. B **76**, 045302 (2007).

[13] 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).

[14] M. Mogi, R. Yoshimi, A. Tsukazaki, K. Yasuda, Y. Kozuka, K. S. Takahashi, M. Kawasaki, and Y. Tokura, Appl. Phys. Lett. **107**, 182401 (2015).

[15] M. Mogi, M. Kawamura, R. Yoshimi, A. Tsukazaki, Y. Kozuka, N. Shirakawa, K. S. Takahashi, M. Kawasaki, and Y. Tokura, Nat. Mater. **16**, 516 (2017).

[16] M. Mogi, M. Kawamura, A. Tsukazaki, R. Yoshimi, K. S. Takahashi, M. Kawasaki, and Y. Tokura, Sci. Adv. **3**, eaao1669 (2017).

[17] D. Xiao, J. Jiang, J.-H. Shin, W. Wang, F. Wang, Y.-F. Zhao, C. Liu, W. Wu, M. H. W. Chan, N. Samarth, and C.-Z. Chang, Phys. Rev. Lett. **120**, 056801 (2018).

[18] M. Allen, Y. Cui, E. Y. Ma, M. Mogi, M. Kawamura, I. C. Fulga, D. Goldhaber-Gordon, Y. Tokura, and Z. X. Shen, Proc. Natl. Acad. Sci. U. S. A. **116**, 14511 (2019).

[19] R. Watanabe, R. Yoshimi, M. Mogi, A. Tsukazaki, X. Z. Yu, K. Nakajima, K. S. Takahashi, M. Kawasaki, and Y. Tokura, Appl. Phys. Lett. **115**, 102403 (2019).

[20] The QTME is related to the quantum Faraday effect measured by L. Wu, M. Salehi, N. Koirala, J. Moon, S. Oh, and N. P. Armitage, Science **354**, 1124 (2016); by K. N. Okada, Y. Takahashi, M. Mogi, R. Yoshimi, A. Tsukazaki, K. S. Takahashi, N. Ogawa, M. Kawasaki, and Y. Tokura, Nat. Commun. **7**, 12245 (2016); and by V. Dziom, A. Shuvaev, A. Pimenov, G. V. Astakhov, C. Ames, K. Bendias, J. Böttcher, G. Tkachov, E. M. Hankiewicz, C. Brüne, H. Buhmann, and L. W. Molenkamp, *ibid.* **8**, 15197 (2017); Faraday effect measurements do not measure the TME directly. For related discussions, see C. W. J. Beenakker, Journal Club of Condensed Matter, April 2016, https://www.condmatjclub.org/?p=2816; and Ref. [7].

[21] Q. L. He, G. Yin, L. Yu, A. J. Grutter, L. Pan, C.-Z. Chen, X. Che, G. Yu, B. Zhang, Q. Shao, A. L. Stern, B. Casas, J. Xia, X. Han, B. J. Kirby, R. K. Lake, K. T. Law, and K. L. Wang, Phys. Rev. Lett. **121**, 096802 (2018).

[22] M. M. Otrokov, T. V. Menshchikova, M. G. Vergniory, I. P. Rusinov, A. Y. Vyazovskaya, Y. M. Koroteev, G. Bihlmayer, A. Ernst, P. M. Echenique, A. Arnau, and E. V. Chulkov, 2D Mater. **4**, 025082 (2017).

[23] J. Li, Y. Li, S. Du, Z. Wang, B.-L. Gu, S.-C. Zhang, K. He, W. Duan, and Y. Xu, Sci. Adv. **5**, eaaw5685 (2019).

[24] M. M. Otrokov, I. I. Klimovskikh, H. Bentmann, D. Estyunin, A. Zeugner, Z. S. Aliev, S. Gaß, A. U. B. Wolter, A. V. Koroleva, A. M. Shikin, M. Blanco-Rey, M. Hoffmann, I. P. Rusinov, A. Y. Vyazovskaya, S. V. Eremeev, Y. M. Koroteev, V. M. Kuznetsov, F. Freyse, J. Sánchez-Barriga, I. R. Amiraslanov *et al.*, Nature (London) **576**, 416 (2019).

[25] T. Hirahara, M. M. Otrokov, T. T. Sasaki, K. Sumida, Y. Tomohiro, S. Kusaka, Y. Okuyama, S. Ichinokura, M. Kobayashi, Y. Takeda, K. Amemiya, T. Shirasawa, S. Ideta, K. Miyamoto, K. Tanaka, S. Kuroda, T. Okuda, K. Hono, S. V. Eremeev, and E. V. Chulkov, Nat. Commun. **11**, 4821 (2020).

[26] T. Hirahara, S. V. Eremeev, T. Shirasawa, Y. Okuyama, T. Kubo, R. Nakanishi, R. Akiyama, A. Takayama, T. Hajiri, S.-i. Ideta, M. Matsunami, K. Sumida, K. Miyamoto, Y. Takagi, K. Tanaka, T. Okuda, T. Yokoyama, S.-i. Kimura, S. Hasegawa, and E. V. Chulkov, Nano Lett. **17**, 3493 (2017).

[27] D. Xiao, M.-C. Chang, and Q. Niu, Rev. Mod. Phys. **82**, 1959 (2010).

[28] This form assumes that the in-plane current and velocity projection operators commute.

[29] A. H. MacDonald, 2021, https://www.youtube.com/watch?v=JLi2hoinSg0&list=PLZee5Mirzz8XBsXnhWykt_Et1kC-fbvru&index=13.

[30] A. H. MacDonald, in *Les Houches, Session LXI, 28 Juin - 29 Juillet 1994: Physique Quantique Mésoscopique*, edited by E. Akkermans, G. Montambaux, J.-L. Pichard, and J. Zinn-Justin (North-Holland, Amsterdam, 1995), p. 659.

[31] T. Thonhauser, D. Ceresoli, D. Vanderbilt, and R. Resta, Phys. Rev. Lett. **95**, 137205 (2005).

[32] D. Xiao, J. Shi, and Q. Niu, Phys. Rev. Lett. **95**, 137204 (2005).

[33] P. Středa, J. Phys. C: Solid State Phys. **15**, L717 (1982).

[34] D. Hsieh, Y. Xia, D. Qian, L. Wray, J. H. Dil, F. Meier, J. Osterwalder, L. Patthey, J. G. Checkelsky, N. P. Ong, A. V. Fedorov, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature (London) **460**, 1101 (2009).

[35] K. Kobayashi, Phys. Rev. B **84**, 205424 (2011).

[36] A. Pertsova and C. M. Canali, New J. Phys. **16**, 063022 (2014).

[37] M. Mogi, T. Nakajima, V. Ukleev, A. Tsukazaki, R. Yoshimi, M. Kawamura, K. S. Takahashi, T. Hanashima, K. Kakurai, T. H. Arima, M. Kawasaki, and Y. Tokura, Phys. Rev. Lett. **123**, 016804 (2019).

[38] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.104.L201102 for a detailed description of the band structure and the surface states for different choices of the exchange field, and the calculation of the Chern number and magnetoelectric response.