Letter

$\mathbb{Z}_2$ topology of bismuth

Irene Aguilera,* Hyun-Jung Kim, Christoph Friedrich, Gustav Bihlmayer, and Stefan Blügel
Peter Grünberg Institute and Institute for Advanced Simulation, Forschungszentrum Jülich and JARA, 52425 Jülich, Germany

(Received 21 May 2021; accepted 23 August 2021; published 17 September 2021)

While first-principles calculations with different levels of sophistication predict a topologically trivial $\mathbb{Z}_2$ state for bulk bismuth, some photoemission experiments show surface states consistent with the interpretation of bismuth being in a topologically nontrivial $\mathbb{Z}_2$ state. We resolve this contradiction between theory and experiment by showing, based on quasiparticle self-consistent *GW* calculations, that the experimental surface states interpreted as supporting a nontrivial phase are actually consistent with a trivial $\mathbb{Z}_2$ invariant. We identify this contradiction as the result of a *crosstalk* effect arising from the extreme penetration depth of the surface states into the bulk of Bi. A film of Bi can be considered bulklike only for thicknesses of about 1000 bilayers ($\approx$ 400 nm) and more.

DOI: 10.1103/PhysRevMaterials.5.L091201

The topological classification of crystals is a very powerful and successful theoretical concept. It underlies, for example, the bulk-boundary correspondence, which relates the bulk topological invariant to the number of robust chiral edge or surface states localized at the boundaries of the material. Hence, experimental efforts have been dedicated to observing boundary localized gapless states to validate their topological character [1–4]. Along with the development of topological band theory, right from the beginning, Bi has been regarded as a prototypical material. The $Bi_{1-x}Sb_x$ alloys were among the first materials theoretically predicted [5], and the first experimentally detected, as nontrivial topological insulators in three dimensions [2]. Elemental Bi, however, was known for many years simply as a trivial semimetal in terms of a $\mathbb{Z}_2$ classification [6]. Therefore, no time-reversal symmetry-protected surface states were expected. On the other hand, a space-group symmetry-based $\mathbb{Z}_4$ classification unveiled a nontrivial topology, suggesting that bulk Bi presents a topological crystalline insulator phase [7] and a higher-order topological phase [8] that supports one-dimensional hinge states in a rotational-symmetry-preserving rod.

However, experimentally, the interpretation of Bi being in a topologically trivial $\mathbb{Z}_2$ state was challenged by angle-resolved photoemission spectroscopy (ARPES) used to investigate the dispersion of the surface states in the $\overline{\Gamma}$-$\overline{\mathrm{M}}$ direction of the Brillouin zone. In 2013, an experimental work [9] claimed a nontrivial $\mathbb{Z}_2$ topology for bulk Bi. In 2016, this work was followed by several other experiments [10–12] that also displayed surface states that connect to bulk valence and conduction bands consistent with a nontrivial interpretation [as shown in Fig. 1(b)]. Thus, the determination of the $\mathbb{Z}_2$ topological character of bismuth has been elusive. The reason for this is the tiny (∼11–15 meV) direct band gap of the Bi band structure at the L point of the bulk Brillouin zone, which projects onto the $\overline{\mathrm{M}}$ point of the surface Brillouin zone. An inversion of the valence and conduction bands at this **k** point would change the topological order of Bi. In other words, it is the symmetry (parity) of the wave functions of the highest valence band at the L point that determines the $\mathbb{Z}_2$ topological character of Bi.

First-principles calculations of bulk Bi, even when carried out with very different theoretical methods [7,13], tend to show that the electronic structure of Bi is described by a trivial $\mathbb{Z}_2$ invariant, and density functional theory (DFT) calculations of films of Bi(111) display trivial surface states [14–16], similar to those shown schematically in Fig. 1(a). The two surface states (red and blue), which are a Kramers pair, are degenerate at the time-reversal-invariant momentum $\overline{\mathrm{M}}$ of the two-dimensional surface Brillouin zone. The obvious contradiction between experiment and theory has remained unresolved to this day and is a result of a lack of understanding.

The discrepancies between theory and experiment have been attributed to the failure of DFT to predict the size of band gaps correctly. This is not entirely improbable given the very small band gap of bulk Bi at the L point. If one improves the DFT calculation by including many-body effects using the quasiparticle self-consistent *GW* (QS*GW*) method [17] and a consistent treatment of spin-orbit coupling, the band gap indeed decreases, but it does not change sign [13], thus again supporting the trivial character of bulk Bi [18].

Several theoretical works have focused on trying to resolve the contradiction between theory and experiment by searching for nontrivial $\mathbb{Z}_2$ phases of Bi: improving the theoretical approximations [13,19], adding strain [13,16,19], and doping [20]. But these works seem to disregard that there are also experiments, not based on photoemission, that support the theoretical predictions of a trivial $\mathbb{Z}_2$. For example, Kang *et al.* [21] recently showed with angle-dependent magnetoresistance experiments that Bi and Sb exhibit markedly different behaviors due to the different parities of the highest valence

*Corresponding author: i.aguilera@fz-juelich.de; Present address: IEK5-Photovoltaik, Forschungszentrum Jülich, 52425 Jülich, Germany.

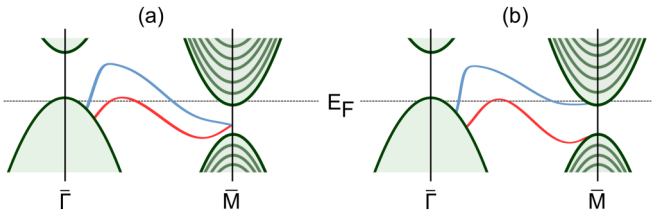

FIG. 1. The two controversial behaviors of the surface states of Bi films in the $\bar{\Gamma}$-$\bar{M}$ direction: (a) trivial surface states as predicted by DFT calculations and (b) experimentally observed connectivity of the surface states with the bulk at the $\overline{M}$ point.

band at the L point [22]. In other words, Bi and Sb must have different $\mathbb{Z}_2$ topological invariants (trivial and nontrivial, respectively). Thus, it is not sufficient to address only the contradiction between theory and ARPES because the confusion around the $\mathbb{Z}_2$ character of Bi goes beyond that.

In this Letter, we consider this issue from a more fundamental perspective that reconciles the very different conclusions of the experimental and theoretical literature. For that, we explore with QS*GW* calculations both finite films of Bi and the thermodynamic limit by comparison with semi-infinite calculations. In this way we intend to provide a consistent picture of the electronic structure of bulk and thin films of Bi where the seemingly contradictory theoretical and experimental results in the literature actually reconcile and both are consistent with a trivial $\mathbb{Z}_2$ topology.

Following the method described in Ref. [23], we parametrize a tight-binding (TB) Hamiltonian that accurately reproduces the *bulk* QS*GW* band structure of Ref. [13]. (The reader is referred to that reference for the convergence parameters of the QS*GW* calculation.) The accuracy of the method is shown for DFT in Fig. S1(a) of the Supplemental Material [24]. The interpolated QS*GW* bulk band structure is shown in Fig. S1(b). The TB Hamiltonian is expressed in a basis of Wannier functions [25] and thus calculated fully *ab initio* with the QS*GW* method without the need for adjustable parameters to fit experimental results. We subsequently use the tight-binding parameters obtained for the bulk to construct the QS*GW*-TB Hamiltonian of films of bismuth of different thicknesses.

The band structure of a film of Bi(111) of 200 bilayers (BLs; approximately 80 nm) is shown in Fig. 2. Unexpectedly, the resulting surface states do not look like in the previous DFT calculations, with the two states touching at the $\overline{M}$ point, and instead, they look very similar to the experimental ones (see Fig. 3(b) of Ref. [12] for a 202-BL sample measured with ARPES). Given this result, the inconsistency between theory and experiment seems to be resolved. The surface states in Fig. 2 display, indeed, a connectivity to the bulk states that suggests a nontrivial topology [in the manner of Fig. 1(b)]. This raises the question of whether the surface states of this film are topologically protected, even though the bulk is trivial.

This question is addressed by performing calculations of the band gap at $\overline{M}$. We show this gap in Fig. 3 as a function of the thickness of the film (top horizontal axis) and the inverse of the thickness (bottom horizontal axis), with zero corresponding to the bulk result. The two upper curves represent the energy difference between the second conduction band and the next to last valence band ($E^{\mathrm{CBM+1}} - E^{\mathrm{VBM-1}}$) obtained with TB based on DFT in the local density approximation (LDA-TB, red squares) and on QS*GW* (purple circles). This energy difference, defined by the quantum-well states at the $\overline{M}$ point (see Fig. 2), should tend to the bulk band gap value for an infinite thickness. The real bulk values, i.e., obtained in an *infinite* (three-dimensional-periodic) LDA or QS*GW* calculation, are shown as open symbols, and the thickness-dependent values indeed converge towards these bulk values, as expected. The small cyan circles represent the evolution of

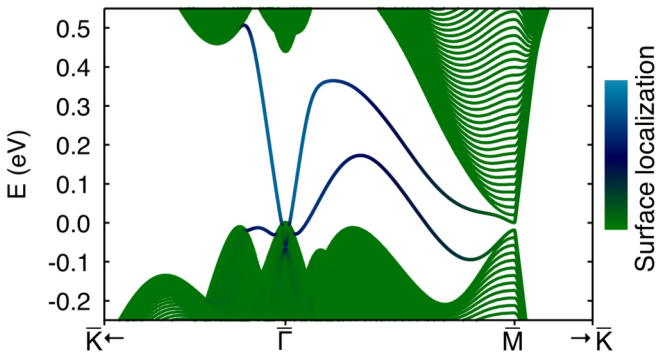

FIG. 2. Band structure of a 200-BL film of Bi(111) along some high-symmetry lines in the two-dimensional Brillouin zone obtained with the QS*GW*-TB method. The color scheme represents the degree of localization of a state at the surface, with blue representing surface states and surface resonances, and green representing the bulklike character.

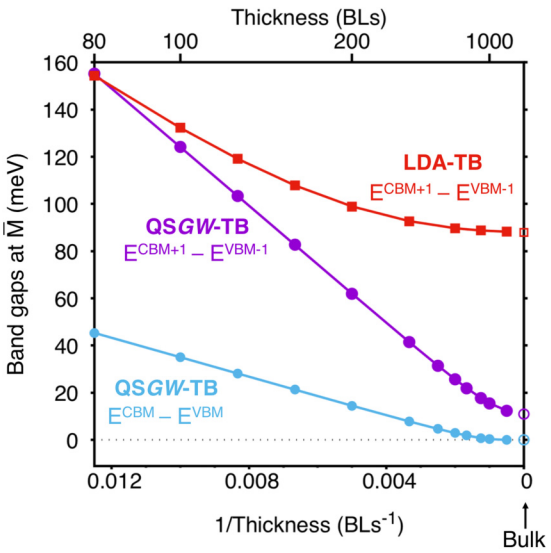

FIG. 3. Band gap at $\overline{M}$ as a function of the thickness (top horizontal axis) and its inverse (bottom horizontal axis) of the Bi films in units of bilayers. Red squares and big purple circles represent the energy difference between the second conduction band and the next to last valence band obtained with TB parameters from DFT (LDA) and QS*GW* calculations, respectively. The QS*GW*-TB results for the actual band gap at $\overline{M}$ are shown by small cyan circles. The open symbols represent the bulk values corresponding to infinite thickness (or zero at inverse thickness).

the actual band gap, i.e., the gap between the first conduction band and the last valence band, obtained with QS*GW*-TB.

If the surface states of the film in Fig. 2 were the result of a nontrivial topology, this band gap would not tend to zero but to the bulk band gap. In other words, the separation between quantum-well states would get smaller as the thickness increases, but both surface states would remain within the bulk projection, so the two curves (cyan and purple) would approach each other, and at infinite thickness, both would tend to the bulk value (see also Fig. S2 in the Supplemental Material [24] for a schematic visualization of this). That is not the case, and as seen in Fig. 3, the band gap indeed tends to zero. From our analysis in Ref. [13], we know that bulk Bi is a trivial semimetal when calculated with QS*GW*. From Fig. 3, we know that the surface states of films of any thickness are not the result of a nontrivial $\mathbb{Z}_2$ topology. This proves that the surface states in Fig. 2 are trivial in spite of their nontrivial appearance.

In the following, we discuss the origin of the seemingly nontrivial character. It has already been discussed in the literature, especially in Ref. [26], that the surface states of Bi have a very long decay length into the bulk. From topological insulators, we know that a long decay length of the surface states can give rise to a "crosstalk" effect: If the wave functions of the surface states at both surfaces of a film decay into the bulk deep enough to overlap with each other, the two surface states might hybridize, which can open a gap where there should otherwise be a degeneracy. The calculations of the decay length of the surface states of Bi have so far been performed with DFT. The most recent predictions of this length were 24 BLs [26] and 21 BLs [16]. This means that films of at least 42 BLs need to be studied to find a degeneracy at the $\overline{\text{M}}$ point as in Fig. 1(a). For thinner films, one should observe a hybridization gap between the surface states leading to a situation more similar to that in Fig. 1(b). But the experimental works in Refs. [11,12] observed such behavior for much thicker samples, typically considered bulklike (80 and 202 BLs, respectively), which led the authors to conclude in the end that the topology is nontrivial.

However, it should be kept in mind that the decay length depends inversely on the size of the band gap [27]. In the case of DFT with the standard approximations used in Refs. [16,26], inverted band gaps like the one of bismuth at the L point are overestimated [13]. In particular, for Bi, the DFT band gap at L is 6 times larger than the experimental one [13]. This leads to an underestimated decay length of the surface states. We have performed calculations with the QS*GW*-TB method for films of thicknesses up to 2000 BLs (approximately 800 nm). Even for a film of 1000 BLs, the two surface states are not yet degenerate at the $\overline{\text{M}}$ point, and there is still a small band gap ($\sim$0.3 meV) between them. For 2000 BLs, the two states can already be considered degenerate at $\overline{\text{M}}$, with a gap between them of only 0.005 meV. This means that when calculations of the decay length are performed with a method that can predict the band gap accurately, the decay length of the surface states of Bi must be larger than 500 BLs. This can explain why the samples in Refs. [11,12] present a connectivity of the surface states with bulk valence and conduction bands without having to assume a nontrivial topology.

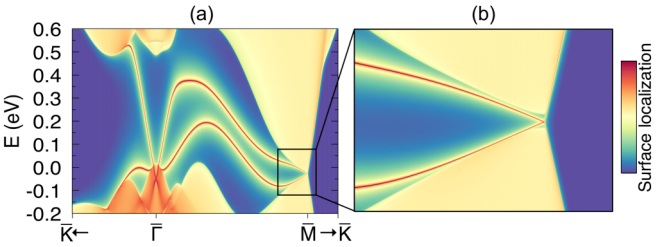

FIG. 4. (a) Spectral function of a semi-infinite calculation of Bi obtained with the iterative Green's function method and a TB Hamiltonian based on QS*GW* calculations. (b) A magnification of (a) around the $\overline{\text{M}}$ point.

Actually, the fact that a 200-BL film of Bi cannot be considered bulklike should not be surprising, given that one can experimentally clearly observe well-defined quantum-well states close to $\overline{\text{M}}$ (see Fig. 3 of Ref. [12]).

It should also be noted that in Refs. [9,10] the possibility of an interaction between the top and bottom surfaces was already mentioned. However, the authors interpreted that the "inter-surface interaction would re-invert the bulk bands at L" (projected to $\overline{\text{M}}$), resulting in a change in topology. As we have discussed above and as is supported by the measurements of Ref. [12], there is no evidence of such an effect as a function of thickness, and therefore, no change in topology occurs that is caused by the thickness or by the surface-surface interaction. So, although it may be counterintuitive, the spectra in Refs. [9–12] are perfectly consistent with a trivial situation. In the case of our calculation in Fig. 2, we are, indeed, certain that the system is trivial and the "topological-looking" connectivity is a consequence of crosstalk and not of a nontrivial topology. We should mention here that this picture does not change if the two surfaces are slightly inequivalent due to the presence of a substrate on one side. Although the dispersion of the upper and lower surface states changes slightly, the connectivity near the $\overline{\text{M}}$ point remains almost unchanged (see Fig. S3 of the Supplemental Material [24]).

To further support our explanation based on a crosstalk effect, we performed a semi-infinite calculation with the iterative Green's function method [28] as implemented in WANNIERTOOLS [29]. For that, we used again the TB Hamiltonian with parameters obtained using QS*GW* calculations. The results are shown in Fig. 4 and prove that in the limit of infinite thickness (where there is obviously no crosstalk), the surface states behave as expected according to the trivial fashion of Fig. 1(a). The surface states are degenerate at $\overline{\text{M}}$, and they do not connect separately to the bulk valence and conduction bands.

Making use of the fact that Bi undergoes a topological phase transition under strain [13], we can now also compare calculations for trivial and nontrivial systems. From Fig. 3(a) of Ref. [13], we can choose two strained states of Bi that have approximately the same band gap but opposite topologies. In Fig. 5, we show results for two systems under volume-conserving compressive strain [0.5% and 1.0% in-plane strain in Figs. 5(a) and 5(b), respectively]. These two strain values were chosen close to the phase transition on the trivial and nontrivial sides. The main (top) panels show band structures

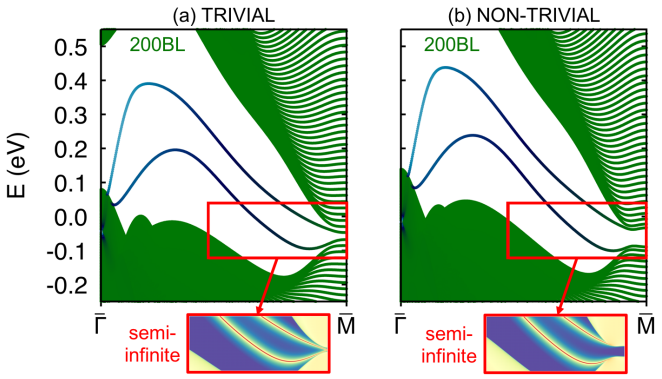

FIG. 5. Band structure of 200-BL films (top panels) of Bi and spectral function of a semi-infinite calculation (bottom panels). Results are obtained with a TB Hamiltonian based on one-shot *GW* calculations. The films have been strained to represent two cases very close to the topological phase transition (see main text). As in Figs. 2 and 4, the color schemes represent the localization on the surface.

of 200 BLs of Bi obtained with a TB Hamiltonian based on the one-shot *GW* method including off-diagonal elements of the self-energy matrix [30]. The surface states of the two 200-BL films in Fig. 5 look qualitatively the same, and this may lead to the conclusion that these films are both nontrivial. But when calculations are performed in a semi-infinite geometry (bottom panels), the differences become evident, and one can now clearly distinguish between the two phases: the trivial one in Fig. 5(a), with degenerate surface states at $\overline{\mathrm{M}}$, and the topological one in Fig. 5(b), with one surface state connecting to the valence band and the other one connecting to the conduction band. Figure 5 thus proves that one cannot distinguish between a trivial phase and a nontrivial phase of Bi only by looking at the connectivity of the surface states with the valence and conduction states unless the bulk limit has really been reached, which is not the case for 200 BLs.

To summarize, there is no inconsistency between theory and experiment about the surface states of Bi films. The surface states predicted by the advanced QS*GW* method are in very good agreement with the ARPES spectra. The theory predicts that bulk Bi is trivial. The surface states of Bi films might look nontrivial because of a crosstalk effect due to the extremely large decay length of the surface states into the bulk. We have discussed both finite films and the thermodynamic limit of semi-infinite systems, and we have shown that bismuth starts behaving like the bulk only after at least 1000 BLs. All this shows that one cannot draw conclusions about the topology of a given sample just by observing the connectivity of the surface states with the valence and conduction bands, and the surface states of Bi(111), claimed to be topological in Refs. [9–12], are fully consistent with a trivial phase. This resolves the long-standing controversy about the $\mathbb{Z}_2$ character of Bi. We would like to point out that according to Ref. [13], the strain needed to drive Bi into a topological semimetal is very small, and therefore, depending on the doping, choice of substrate, or growth conditions, one cannot rule out that some of the experimentally measured samples might actually be in a nontrivial state [31].

To conclude, irrespective of the possible existence of both trivial and nontrivial Bi samples, this work has shown that the observation of a seemingly nontrivial surface-bulk connectivity can be deceptive in the case of film geometry. If the film is not thick enough, a crosstalk effect of interacting surface wave functions, penetrating deep into the sample from both sides of the film, can explain the observation equally well, and no conclusive answer as to whether the respective bulk material is trivial or not can be given. Therefore, predictions from DFT calculations have to be interpreted with care as an overestimation of the band gap of a few meV can lead to a grossly underestimated penetration depth of the surface states. Since deeply penetrating surface states are not only characteristic of Bi but also common in topological insulators with their narrow band gaps, the crosstalk modification of the band connectivity is a general issue in topological matter. In fact, our conclusions are also applicable to the edge states of ribbons of two-dimensional topological insulators, i.e., quantum spin Hall as well as Chern insulators, if the ribbon is not wide enough to decouple the two edges.

Acknowledgments. S.B. gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Centers SFB 1238 (Project C01) and H.-J.K. acknowledges financial support from the Alexander von Humboldt Foundation. We acknowledge the computing time granted through JARA-HPC on supercomputer JURECA at Forschungszentrum Jülich.

[1] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, Science **314**, 1757 (2006).
[2] D. Hsieh, D. Qian, L.Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature (London) **452**, 970 (2008).
[3] I. K. Drozdov, A. Alexandradinata, S. Jeon, S. Nadj-Perge, H. Ji, R. J. Cava, B. A. Bernevig, and A. Yazdani, Nat. Phys. **10**, 664 (2014).
[4] Y. Zhang, K. He, C.-Z. Chang, C.-L. Song, L.-L. Wang, X. Chen, J.-F. Jia, Z. Fang, X. Dai, W.-Y. Shan, S.-Q. Shen, Q. Niu, X.-L. Qi, S.-C. Zhang, X.-C. Ma, and Q.-K. Xue, Nat. Phys. **6**, 584 (2010).
[5] L. Fu and C. L. Kane, Phys. Rev. B **76**, 045302 (2007).
[6] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).
[7] C.-H. Hsu, X. Zhou, T.-R. Chang, Q. Ma, N. Gedik, A. Bansil, S.-Y. Xu, H. Lin, and L. Fu, Proc. Natl. Acad. Sci. USA **116**, 13255 (2019).
[8] F. Schindler, Z. Wang, M. G. Vergniory, A. M. Cook, A. Murani, S. Sengupta, A. Y. Kasumov, R. Deblock, S. Jeon, I. Drozdov, H. Bouchiat, S. Gueron, A. Yazdani, B. A. Bernevig, and T. Neupert, Nat. Phys. **14**, 918 (2018).
[9] Y. Ohtsubo, L. Perfetti, M. O. Goerbig, P. L. Fèvre, F. Bertran, and A. Taleb-Ibrahimi, New J. Phys. **15**, 033041 (2013).
[10] Y. Ohtsubo and S. Kimura, New J. Phys. **18**, 123015 (2016).
[11] M.-Y. Yao, F. Zhu, C. Q. Han, D. D. Guan, C. Liu, D. Qian, and J. Jia, Sci. Rep. **6**, 21326 (2016).

[12] S. Ito, B. Feng, M. Arita, A. Takayama, R.-Y. Liu, T. Someya, W.-C. Chen, T. Iimori, H. Namatame, M. Taniguchi, C.-M. Cheng, S.-J. Tang, F. Komori, K. Kobayashi, T.-C. Chiang, and I. Matsuda, Phys. Rev. Lett. **117**, 236402 (2016).

[13] I. Aguilera, C. Friedrich, and S. Blügel, Phys. Rev. B **91**, 125129 (2015).

[14] Y. M. Koroteev, G. Bihlmayer, J. E. Gayone, E. V. Chulkov, S. Blügel, P. M. Echenique, and P. Hofmann, Phys. Rev. Lett. **93**, 046403 (2004).

[15] H.-J. Zhang, C.-X. Liu, X.-L. Qi, X.-Y. Deng, X. Dai, S.-C. Zhang, and Z. Fang, Phys. Rev. B **80**, 085307 (2009).

[16] T.-R. Chang, Q. Lu, X. Wang, H. Lin, T. Miller, T.-C. Chiang, and G. Bian, Crystals **9**, 510 (2019).

[17] S. V. Faleev, M. van Schilfgaarde, and T. Kotani, Phys. Rev. Lett. **93**, 126406 (2004).

[18] Of course, there is still the possibility that high-order self-energy diagrams may eventually flip the sign, but we show here that there is no need for more accurate calculations to reconcile the literature.

[19] C. König, J. C. Greer, and S. Fahy, Phys. Rev. B **104**, 035127 (2021).

[20] K.-H. Jin, H. W. Yeom, and F. Liu, Phys. Rev. B **101**, 035111 (2020).

[21] W. Kang, F. Spathelf, B. Fauque, Y. Fuseya, and K. Behnia, arXiv:2103.07296.

[22] O. J. Clark, F. Freyse, I. Aguilera, A. S. Frolov, A. M. Ionov, S. I. Bozhko, L. V. Yashina, and J. Sánchez-Barriga, Commun. Phys. **4**, 165 (2021).

[23] I. Aguilera, C. Friedrich, and S. Blügel, Phys. Rev. B **100**, 155147 (2019).

[24] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevMaterials.5.L091201 for bulk band structures, the evolution of trivial and nontrivial surface states with increasing thickness, and a simulation of an asymmetric Bi film.

[25] A. A. Mostofi, J. R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, Comput. Phys. Commun. **178**, 685 (2008).

[26] H. Ishida, J. Phys.: Condens. Matter **29**, 015002 (2016).

[27] J. Linder, T. Yokoyama, and A. Sudbø, Phys. Rev. B **80**, 205401 (2009).

[28] M. P. L. Sancho, J. M. L. Sancho, and J. Rubio, J. Phys. F **14**, 1205 (1984).

[29] Q. Wu, S. Zhang, H.-F. Song, M. Troyer, and A. A. Soluyanov, Comput. Phys. Commun. **224**, 405 (2018).

[30] I. Aguilera, C. Friedrich, G. Bihlmayer, and S. Blügel, Phys. Rev. B **88**, 045206 (2013).

[31] For example, in Ref. [32], the authors attributed the nontrivial topology observed to the strain induced in the vicinity of a dislocation. They concluded that their samples consisted of a nontrivial cylindrical volume (around the dislocation) embedded in a $\mathbb{Z}_2$ trivial bulk.

[32] A. K. Nayak, J. Reiner, R. Queiroz, H. Fu, C. Shekhar, B. Yan, C. Felser, N. Avraham, and H. Beidenkopf, Sci. Adv. **5**, eaax6996 (2019).