Many-body corrected tight-binding Hamiltonians for an accurate quasiparticle description of topological insulators of the Bi_2Se_3 family

Irene Aguilera,* Christoph Friedrich, and Stefan Blügel
Peter Grünberg Institute and Institute for Advanced Simulation, Forschungszentrum Jülich and JARA, D-52425 Jülich, Germany

(Received 18 June 2019; published 30 October 2019)

We generate many-body corrected tight-binding Hamiltonians for topological insulators of the Bi_2Se_3 family. To this end, we use *ab initio* calculated parameters extracted from *GW* calculations, thus capturing many-body exchange and correlation effects, in contrast to previous tight-binding models. We investigate the effect of many-body renormalizations on the electronic structure of bulk and surface states of semi-infinite systems as well as thin films of these materials. It is shown that the *GW* self-energy correction brings about profound changes not only in the band-gap values but also in the band dispersion around the inverted gaps with respect to standard density-functional theory (DFT). These changes substantially improve the agreement with experiment. We discuss the strong renormalization effect as being a result of the characteristic overestimation of inverted gaps by standard approximations of DFT (opposite to the underestimation of gaps in topologically trivial materials). In particular, we analyze the consequences that these renormalizations have on the dispersions of the topological surface states and on the surface resonances. For reference, the tight-binding Hamiltonians are provided in the Supplemental Material [30].

DOI: 10.1103/PhysRevB.100.155147

I. INTRODUCTION

Since the first publication [1] of the bulk and surface band structures of Bi_2Se_3, Bi_2Te_3, Sb_2Te_3, and Sb_2Se_3 in 2009, much has been learned about the electronic properties of these materials from theoretical calculations. However, the vast majority of first-principles calculations have relied, thus far, on standard density functional theory (DFT). Whereas DFT is an exact theory [2], approximations must be made in practice, the most common being the local-density (LDA) [3] or the generalized gradient (GGA) approximations to the exchange-correlation functional within the Kohn-Sham (KS) formalism [4]. It is well known that LDA and GGA miss important many-body exchange and correlation effects, which might have affected the results published so far. In fact, as we will discuss in this paper, closer analysis reveals that several results are at odds with experimental observations. Furthermore, comparison with LDA or GGA band structures might have led to erroneous interpretation of experimental spectra. From a theoretical point of view, the DFT-KS eigenstates are the solutions of a Schrödinger equation for a fictitious non-interacting system. The KS eigenvalues should therefore not be interpreted as the excitation energies of the real interacting system. In practice, this is nevertheless often done, and the KS eigenvalue spectrum is interpreted as the band structure of the real system. All DFT calculations in this paper have been carried out with the LDA approximation. The corresponding GGA results are very similar.

To go beyond DFT, few works have characterized the materials with the use of the quasiparticle many-body *GW* approach [5], which contains many-body exchange and correlation effects. Due to the high computational cost of the *GW* method, most studies only focused on bulk properties [6–15]. Only two works have also addressed the question of how the many-body effects would affect the surface states. The first [7] employed a "scissors operator" that was made to reproduce the gap opening seen in *GW* calculations of the bulk without spin-orbit coupling (SOC). Then, this scissors operator was applied to the films of Bi_2Se_3 and Bi_2Te_3 with up to five quintuple layers (QLs) and finally the SOC was included as a correction. The second work [16–18] presented *explicit GW* calculations of slabs of Bi_2Se_3, Bi_2Te_3, and Sb_2Te_3 up to six-QL thickness. The discussion was thus concentrated on the two-dimensional (2D) limit. These works showed very good agreement with spectroscopy measurements of the topological surface state (TSS) and are, to date, the most accurate calculations of this family of materials in the 2D limit. Such calculations are a real computational challenge and going to thicknesses larger than that order of magnitude (five to six QLs) is extremely costly.

However, most spectroscopy measurements published in the literature for this family of materials are made on single-crystal samples or thick films and several publications [19,20] discuss the surface resonances between the TSS and the bulk continuum. In addition, since the bulk properties of these materials are often substantially affected by many-body effects [8,13,15], the question arises as to how strongly the surface states would be affected, in particular, in comparison to the effects in the bulk continuum. To answer this question, two types of systems are investigated in this paper, very thick films and the semi-infinite limit. The systems are so large

*Current affiliation: IEK5-Photovoltaik, Forschungszentrum Jülich, 52425 Jülich, Germany.

that they cannot be studied with DFT or the *GW* method anymore. A way out is offered by the tight-binding (TB) approach employing parameters obtained from DFT or *GW*. We construct TB Hamiltonians based on Wannier functions [21] for the description of the surface electronic structure in a film geometry. This method is especially accurate for materials that present a van der Waals gap in their lattice structure and has often been used for this family of materials. Applications include calculations of heterostructures [22], topological invariants [23], topological phase transitions [24], semi-infinite systems [1] or very thick slabs and their spin texture [25], electrons dynamics [26], band bending at the surfaces [27], nonstoichiometric alloys [28], etc. In almost all calculations so far, the TB parameters have been obtained from DFT calculations. In this paper, the corresponding parameters are deduced, instead, from *GW* calculations of the bulk. In the following, we will use the notation *GW*-TB and DFT-TB to refer to results obtained with a *GW* and DFT parametrization, respectively. The present method should not be confused with the so-called DFTB approach and its version combined with *GW* [29]. These latter methods constitute simplifications in the numerical realization of the DFT and *GW* method using ideas of TB, which allow the treatment of systems larger than what would be possible with standard DFT or *GW* calculations. The method used in the present paper differs from those approaches in that it employs standard DFT and *GW* calculations of a small system (normally the bulk) without approximations, which can then be mapped to very large supercells (normally a thick film) using TB Hamiltonians. The *GW*-TB approximation has, thus, the same computational burden as DFT-TB, and both can be used to study equally large systems. *GW*-TB, of course, requires that a *GW* calculation of the bulk material is feasible. For the present family of materials, with five atoms in the unit cell, bulk *GW* calculations (also including the spin-orbit interaction [11] in both *G* and *W*) are easily affordable. As we will show, the *GW*-TB approximation offers then a relatively cheap method to study large system sizes including many-body renormalization effects.

For reference, we provide the bulk *GW*-TB Hamiltonians in the Supplemental Material [30] both including and excluding lifetime effects, leading to non-Hermitian and Hermitian Hamiltonians, respectively [31]. We analyze results of the four binary materials Bi_2Se_3, Bi_2Te_3, Sb_2Se_3, and Sb_2Te_3, as well as the ternary compound Bi_2Te_2Se. There is a growing interest in the latter because it shows a much smaller bulk doping level than its binary siblings. In practical applications, the bulk doping level can have very important consequences on the electron transport and dynamics [32,33].

References [6–15] highlighted the main changes caused by the *GW* many-body corrections on the bulk band structures when compared to DFT. DFT results often showed a very pronounced valence-band maximum (VBM) or conduction-band minimum (CBM) at an off-symmetry **k** point. We will discuss in Fig. 2 that this is the result of a systematic exaggeration of the band inversion in DFT calculations. As a consequence, the DFT inverted band gaps of topological insulators (TIs) tend to be overestimated, which might result in an incorrect identification of trivial semiconductors as TIs [34]. The combination of the DFT underestimation of normal band gaps and its overestimation of inverted band gaps results in very rich renormalization effects on the bands that can have a significantly different dispersion in DFT and *GW* (see, e.g., Fig. 2 of Ref. [9]). In particular, it changes the fundamental band gap of Bi_2Se_3 from indirect (DFT) to direct (*GW*), in agreement with experiments (see Sec. IV A).

All these deficiencies of the bulk DFT electronic structure were corrected by the *GW* calculations reported in the cited publications, discussed in more detail in Sec. II. In the present paper, we want to go beyond the bulk description and investigate the effects that the many-body corrections have on the surface electronic structure and the surface resonances. This will provide a revised and more reliable theoretical description of the TSS of these materials.

II. REVIEW OF PREVIOUS BULK RESULTS

Figures 1(a)–1(d) show the DFT-TB calculations of 100-QL films of the topological insulators Bi_2Se_3, Bi_2Te_3, Sb_2Te_3, and Bi_2Te_2Se. (In Sec. IV D, we discuss separately the special case of Sb_2Se_3, the identification of which as a trivial insulator has recently been debated.) The color scheme gives the surface character of the respective states. More precisely, it gives the projection of the states onto the Wannier functions associated to the atoms of the topmost QL. Dark purple thus refers to pure surface states, surface resonances appear in darker blue (visible, for example, when the Rashba states merge with the bulk continuum), and the light blue states have a bulklike character. Some of the bulk features visible in the band structures of Figs. 1(a)–1(d) have been shown in Refs. [8,13,15] to be inconsistent with experiments.

(i) Whereas it is well known that DFT (in the LDA and GGA approximations) tends to underestimate the *normal* gaps of semiconductors, the *inverted* gaps of TIs have turned out to be overestimated by DFT in the cases considered so far [7,14]. Usually, both types of gaps are improved when the *GW* approximation is used (schematically represented in Fig. 2). In the case of band inversion, the bands are inverted locally, i.e., at specific points in the Brillouin zone, whereas the bands are uninverted elsewhere and the energy gaps are *normal* (or *trivial*). This means that, within the same material, there can be over- and underestimation of the gap by DFT at different **k** points. It is easy to imagine that, in such a case, the dispersion of a band along the path that connects two of these **k** points must be inaccurate, and a simple rigid shift of the unoccupied states cannot correct the band structure. An example will be presented in Fig. 8.

(ii) DFT may wrongly identify trivial semiconductors as TIs [34], as a result of the overestimated band inversion.

(iii) The critical points for topological phase transitions are, in general, predicted inaccurately by DFT [35–37]. The phase transitions can be caused by varying strain, alloying, variation of the SOC strength, etc. Given that the predicted band-gap value of the unperturbed material is far from experiment, it is to be expected that the critical parameter (strain, etc.) needed to close and (un)invert it will be over- or underestimated, too. For example, the In concentration necessary to cause a topological phase transition in the alloys $(Bi_{1-x}In_x)_2Se_3$

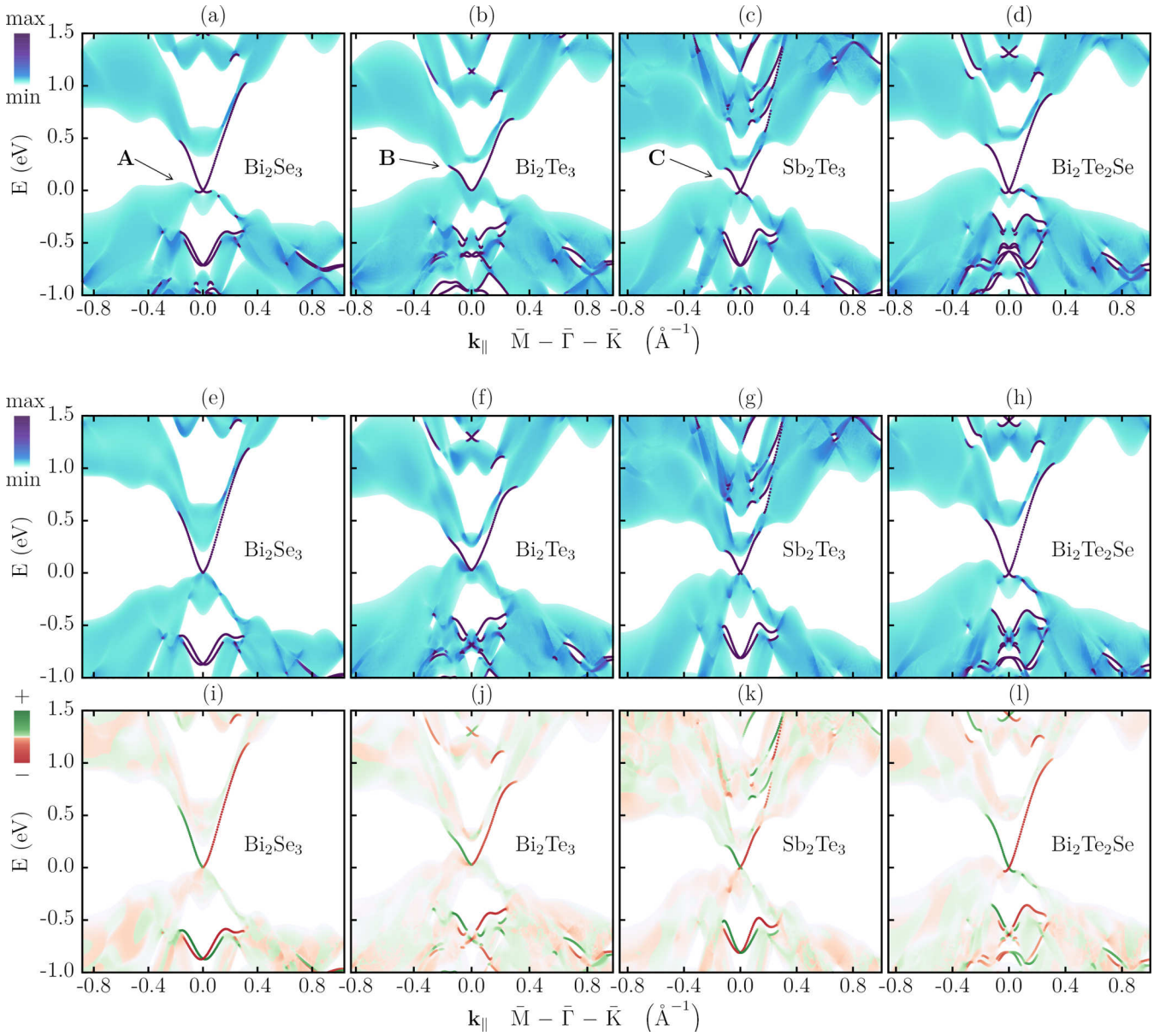

FIG. 1. Band structure of 100-QL films of Bi_2Se_3, Bi_2Te_3, Sb_2Te_3, and Bi_2Te_2Se obtained with a tight-binding model based on DFT (a)–(d) and with many-body effects at the *GW* level (e)–(l). The color map of panels (a)–(h) represents the degree of localization of the states on the topmost QL. The color coding in panels (i)–(l) represents the in-plane spin polarization. A, B, and C label bulk features characteristic of DFT calculations that were shown to be inconsistent with experimental measurements (see text). The Bloch momentum $\mathbf{k}_\parallel$ is given relative to the $\bar{\Gamma}$ point of the surface Brillouin zone. Energies are aligned such that the Dirac point appears at 0 eV.

is predicted by DFT to be around 17% [35], instead of the experimental value of ≈5% [37].

(iv) The band inversion in TIs is usually caused by a band reordering due to the SOC. The avoided crossing of the bands opens an inverted gap and gives rise in many cases to a so-called M shape or camelback shape. For materials where the band inversion is very "deep" (typically with stronger SOC), there is a more pronounced M shape, whereas for other materials a band inversion can occur without producing any M shape [9]. As DFT overestimates the inverted gaps, it often gives rise to exaggerated M shapes. In some cases, DFT thus predicts a local minimum in the band dispersion (the valley of the M shape) where there should be a maximum [7].

(v) This has a fundamental consequence in Bi_2Se_3 : Feature A in Fig. 1(a) shows the absolute bulk VBM of Bi_2Se_3 as predicted by DFT at an off-symmetry **k** point, whereas at the $\bar{\Gamma}$ point there is a minimum in the bulk states (light blue) due to the presence of an M shape. However, based on angle-resolved photoemission spectroscopy (ARPES) and *GW* calculations, Ref. [8] showed that Bi_2Se_3 is a *direct*-gap semiconductor with the VBM at the Γ point. For many years, in contrast, the gap of Bi_2Se_3 had been believed to be *indirect*. This belief resulted from a misinterpretation of ARPES measurements on the basis of DFT results. (The M shape observed in ARPES for photon energies corresponding to the bulk Γ point is not a bulk feature but a 2D one [8,27,38]. We will discuss this

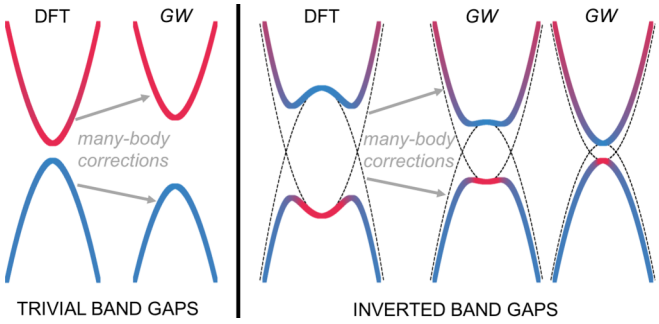

FIG. 2. Band inversion caused by SOC as described by $k \cdot p$ perturbation theory [1,7]. In the absence of an inverted gap (left), the general trend observed for *GW* corrections is to increase the band gap of semiconductors. In the case of an inverted gap (right), the effect is opposite and the value of the inverted gap is usually reduced by the *GW* correction. The last two panels illustrate the flattening or disappearance of the camelback shape due to many-body effects.

in Sec. IV A and Fig. 3.) Previous tight-binding Hamiltonians provided in the literature [39] do not predict the VBM at the Γ point or the direct nature of the band gap and, instead, display an M shape around the Γ point, in disagreement with photoemission, optics, and transport experiments [15,27,40–43].

(vi) In Ref. [15], the electronic structure of Sb_2Te_3 was thoroughly investigated using different exchange-correlation functionals and different sets of lattice parameters. The conclusion was that an off-symmetry VBM [like the one labeled C in Fig. 1(c)] significantly higher in energy than the one at $\bar{\Gamma}$ is not consistent with the experimental literature. *GW* predicts a maximum of the valence band at the center of the Brillouin zone. Depending on the lattice parameters and which functional was used as a starting point for the *GW* calculation, this maximum can be the true VBM or a local one. However, the energy difference between the local maximum and the VBM is very small, so that the two maxima can be considered degenerate. This scenario favors that the Dirac point (DP) of the TSS of Sb_2Te_3 lies above the bulk VBM. There is experimental evidence [44–46] that this is in fact the case.

(vii) Feature B in Fig. 1(b) represents the absolute CBM of Bi_2Te_3. Reference [13] showed that the location of the CBM shifted away from $\bar{\Gamma}$ is at odds with ARPES measurements. Instead, the experimental CBM of Bi_2Te_3 appeared along the $\bar{\Gamma}$ direction, in contrast to what is predicted by DFT. There is a second minimum that is almost energetically degenerate with the one at $\bar{\Gamma}$, but never dropping as low as the one of feature B.

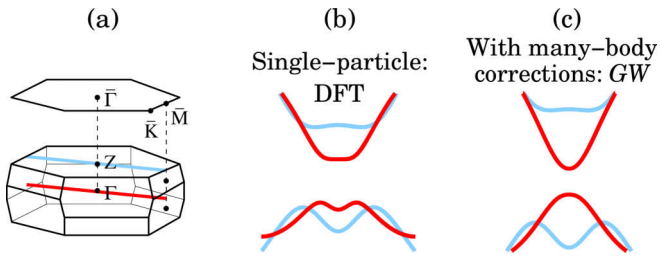

FIG. 3. (a) Bulk Brillouin zone of the materials of the Bi_2Se_3 family. (b) DFT-LDA and (c) *GW* bulk bands of Bi_2Se_3 in the vicinity of $\bar{\Gamma}$ for the path including Γ (red, where the band inversion takes place) and the path including Z (blue, no band inversion).

III. METHODS

The *GW* calculations for the bulk systems were performed with the same convergence parameters as the ones described in Ref. [9] except that the **k**-point sampling employed here is denser ($8 \times 8 \times 8$) in order to obtain well converged Wannier interpolations [47] of the *GW* bands. We used the DFT code FLEUR [48] and the *GW* code SPEX [49]. Both rely on the all-electron full-potential linearized augmented-plane-wave (FLAPW) formalism. The WANNIER90 library [21] was employed for the disentanglement procedure in the Wannier construction. The experimental lattice parameters were taken from Refs. [50] (Bi_2Se_3), [51] (Bi_2Te_3 and Bi_2Te_2Se), and [52] (Sb_2Te_3). Given the lack of experimental results for rhombohedral Sb_2Se_3, we used the theoretically optimized (relaxed) lattice structure from Ref. [53] (see Sec. IV D). As mentioned above, with DFT we refer to the LDA approximation, but results based on GGA show the same behavior. There is not much literature about the use of other functionals (e.g., hybrids [54]) for this family of TIs, but the B3PW91 functional has been shown [55] to yield "*GW*-quality results" for Bi_2Se_3 and Bi_2Te_3. Unpublished results for Bi_2Se_3 with the HSE functional [56] are in good agreement with *GW* when experimental lattice parameters are used. In contrast, Park and Ryu [57] concluded that hybrid DFT calculations are not superior to GGA for predicting the band structures of this family of materials. However, in this case, the comparison was made between GGA band gaps computed with the GGA-relaxed lattice parameters and HSE06 band gaps using the HSE06-relaxed structure. Since the band gap for this family of materials is very sensitive to the lattice parameters [8], the differences in the band gaps in [57] might arise from the different lattice parameters and not from the different functionals. Given the lack of systematic literature with hybrid functionals for topological insulators of this family, the question of whether these functionals provide an alternative to *GW* in terms of accuracy remains open.

For the *GW* calculations in this paper, the SOC was included already at the level of the reference one-particle band structure (the $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ approach from Refs. [11,58]). In this way, the SOC effects are fully included in the Green function G, the screened Coulomb potential W, and the self-energy Σ. The calculations were carried out on the supercomputer JURECA at the Jülich Supercomputing Centre [59].

In Sec. II, we have discussed the main features of the DFT electronic structures that do not fully agree with experimental observations and require the inclusion of many-body effects. The discussion was mostly focused on the bulk band structure but, in practice, to study topological insulators, one is also interested in the topological surface states and their interplay with the bulk continuum. As the surface states coexist and interact with the bulk states, an accurate theoretical description should include many-body effects in very thick films. Unfortunately, due to the high computational cost of *GW*, the direct calculation of very large supercells required for the study of thick films, in particular, taking into account SOC, is

currently not feasible. Therefore, we construct tight-binding Hamiltonians for the slab geometry as described below. The TB parameters are obtained from first-principles calculations of the bulk (*GW* or DFT). For the semi-infinite calculations discussed in Sec. IV E, we use the iterative Green-function method [60,61] as implemented in the code WANNIERTOOLS [62].

Wannier functions are linear combinations of the Bloch eigenfunctions and are localized in space. They can be viewed as Fourier transforms of the Bloch functions. Their centers coincide in most cases with the ionic positions. The Wannier functions thus provide a suitable set of functions to be used in TB studies. In the conventional TB approach, pure atomic orbitals are used as a basis set instead. For the extended bulk states, Wannier functions constitute in general a more natural and more accurate basis for TB Hamiltonians. Often, Wannier functions are defined in such a way that they are maximally spatially localized. For these compounds, however, we found that already the *projected* Wannier functions, that are often used as a "first guess" to start the maximal localization procedure, are sufficiently localized to yield an interpolated band structure as accurate as the one obtained from maximally localized Wannier functions (Fig. 12 in Appendix A). Wannier functions constructed in this way have the advantage that they remain very close to pure atomic orbitals. We have found the disentanglement procedure [63] to be necessary, however.

Apart from the construction of TB Hamiltonians, the Wannier functions also provide an efficient way to obtain band structures from calculations performed on a coarse $\mathbf{k}$ mesh. This so-called Wannier interpolation technique will naturally lead us to an explanation of how the TB Hamiltonians of this paper are constructed. A formal description of the Wannier interpolation can be found in Ref. [47] and also in Appendix A of the present paper. The actual first-principles calculation is carried out on a relatively coarse, uniform reciprocal-space $\mathbf{k}$ mesh ($8 \times 8 \times 8$), where the quantity of interest, in this case the Hamiltonian $H(\mathbf{k})$, is calculated from the Bloch eigenstates. A set of selected bands is then transformed into Wannier functions via a Fourier transformation to real space. The Hamiltonian $H(\mathbf{R})$ is now expressed in real space in a basis of localized Wannier functions. Thanks to the spatial localization of the Wannier functions, the Hamiltonian matrix elements in real space decay rapidly with $|\mathbf{R}|$, which invites truncation of long-range interactions. An inverse Fourier transformation then provides the Hamiltonian matrix elements at an arbitrary point $\mathbf{q}$ not necessarily contained in the original $\mathbf{k}$-point set. The so-obtained "interpolated" Hamiltonian gives rise to physical and surprisingly accurate band structures that are often practically indistinguishable from those obtained in the explicit first-principles DFT or *GW* calculation (Fig. 13 in Appendix A). In the case of *GW*, Wannier interpolation is often used to plot band structures, as it is computationally demanding to evaluate the *GW* self-energy explicitly for the many $\mathbf{k}$ points necessary to obtain a smooth band plot. In the examples of this paper, the interpolated Hamiltonian reproduces, upon diagonalization, the bulk *GW* band structure of the topological insulators very accurately.

In the step between the two Fourier transformations described above, the Hamiltonian is defined in real space and its matrix elements represent the on-site energies and hoppings between sites in the lattice. Given the spatial interpretation of the matrix elements, one can construct a new Hamiltonian for a system with different geometry by selecting the appropriate matrix elements from the bulk. In particular, we can construct slab Hamiltonians by "stacking" blocks of $H(\mathbf{R})$ in the direction perpendicular to the surface. We also need to make zero the hoppings that relate the atoms at the surface with their neighbors that would exist in the bulk but not in the slab. The method assumes that the surface does not suffer significant changes when the bulk is "cut." This approximation might be too drastic for certain materials, for example, if a surface structural reconstruction plays an important role. However, in the case of the topological insulators of the Bi_2Se_3 family, the van der Waals gap between QLs is a very natural cleavage plane, and this approximation gives excellent results.

IV. RESULTS

Figures 1(e)–1(h) show the *GW*-TB band structures of 100-QL films of the four compounds, analogous to Figs. 1(a)–1(d). (A version of this figure with a much larger energy range can be found in Fig. 14 of Appendix B.) In the top panels of Fig. 1 the color coding refers to the localization of the states on the surface. In the bottom panels [(i)–(l)], the color coding shows the degree of in-plane spin polarization, with red and green for opposite spin directions (white indicates no spin polarization). Not only are the TSSs clearly recognizable but also the Rashba states in the valence band. Whereas the spin polarization decreases very abruptly when the surface states merge into the bulk states, induced spin polarization can be found in bulk states in the vicinity of the points where the surface states disappear into the bulk. This is visible, for example, for the lower Dirac cone (DC) of Bi_2Se_3 in Fig. 1(i) and of Sb_2Te_3 in Fig. 1(k), or for the Rashba states [especially around $\bar{\Gamma}$ at ≈ -0.6 eV for Bi_2Te_3 and Bi_2Te_2Se in Figs. 1(j) and 1(l)]. By comparison with Figs. 1(a)–1(d), it becomes obvious that the *GW*-TB approximation has corrected those DFT-TB bulk features (light blue) that were listed in Sec. II as inconsistent with the experimental observations. And the correction has strongly affected the surface states (dark purple).

In Appendix B, we provide all the technical details of the Hamiltonians used for these calculations. These Hamiltonians have already been used for analyzing the ultrafast electron dynamics [26] and the photogalvanic effect spectra [64] of the present TIs, as well as understanding the topological phase transition in $(Bi_{1-x}In_x)_2Se_3$ alloys [37].

The most important changes in the surface states brought about by the *GW* renormalization are discussed in the following. Comparison with photoemission experiments from the literature is shown when possible.

A. Lower Dirac cone of Bi_2Se_3

The discovery of the direct nature of the band gap of bulk Bi_2Se_3 [7–9,65] changed our understanding of its electronic structure. With first-principles methods, this can only be correctly described with state-of-the-art theoretical techniques, beyond standard DFT, like *GW* calculations [7–9] or DFT with hybrid functionals [55,56]. The first experimental evidence of the direct nature of the band gap was provided by

ARPES measurements [8,65]. These experimental measurements indicated not only a direct band gap but also that the Dirac point lies very close to the bulk VBM. In addition, the *GW* calculations showed that both the bulk valence and conduction bands are parabolic [see light blue states in Fig. 1(e)] and, respectively, with maximum and minimum at the Γ point.

It is unfortunate that these crucial results went relatively unnoticed in the following years. The reason for this is the common understanding of the camelback shape observed in photoemission spectra of Bi_2Se_3 to originate from the inversion of the gap around the Γ point. However, this is a misconception, and the seeming agreement with DFT results has reinforced it. Let us try to understand this with the schematic representation of the bands in Fig. 3. The mechanism of how an inverted band gap might be related to a camelback-shaped valence band was very intuitively shown in Ref. [7] using $k \cdot p$ perturbation theory. But an inverted gap can also exist without a camelback shape [7,9], as is shown in the last panel of Fig. 2. And conversely, a band can present a dispersion that resembles a camelback shape, without it being related to an inverted gap (the avoided crossing of an occupied and an unoccupied band). We will see in the following that this is the case of the experimentally observed camelback shape of Bi_2Se_3 (i.e., it is not related to the inverted gap): Close to the Γ point, many-body calculations [red bands in Fig. 3(c)], photon-energy dependent ARPES studies, as well as transport and optical measurements [40–43,67] all agree on a direct gap and parabolic valence and conduction band (no camelback shape). Close to the Z point of the Brillouin zone [see Fig. 3(a)], where Bi_2Se_3 does not present an inverted band gap [68], the band has indeed a dispersion that resembles an M or a camelback shape [light blue valence band in Figs. 3(b) and 3(c)]. On this point, all measurements and calculations agree. But why does one observe a camelback shape close to Γ in many ARPES measurements of Bi_2Se_3 when the most accurate bulk calculations indicate that there should not be one? The camelback shape observed in the ARPES spectra for photon energies corresponding to Γ is not a bulk feature. It has been shown [8,27,38] to be a nondispersive 2D feature, which causes a resonant enhancement of the photoemission intensity of the camelback-shaped state of the bulk near the Z point. This 2D state could originate [27] from the quantization of the valence-band states in a potential well formed by the band bending of the bands at the surface [69]. The camelback shape is, however, still often attributed to the bulk band inversion given the presence of a camelback shape close to Γ in DFT results [red in Fig. 3(b)], whereas the alternative interpretation provided by *GW* results is overall more consistent with the experimental observations.

The parabolic bands predicted by *GW* invite us to compute the effective masses and compare to magneto-optic experiments. The experimental effective masses are $0.14m_0$ for electrons and $0.18m_0$ for holes [41], which compare very well with the values predicted by *GW* ($\approx 0.15m_0$ for electrons and $\approx 0.21m_0$ for holes). In addition, the experiments have found that the electron-hole asymmetry in Bi_2Se_3 is very small [41,43]. The electron-hole asymmetry can be characterized by the D_2 parameter in $k \cdot p$ Hamiltonians of the form proposed in Ref. [1]. Experimentally, it is found to be approximately 3 eV Å^2. The value given by *GW* (6 eV Å^2) is twice that value, but it is much smaller and closer to the experiment than the values reported from DFT (19.6, 26.3, and 30.4; see Refs. [1,70] and references therein). In summary, the parameters of the $k \cdot p$ Hamiltonian [1] that approximate best the *GW* bands are the following (numbers in brackets are values from magneto-optics experiments [41]): $2M = 210$ meV (200 ± 5), $A_1 = 2110$ meV Å, $A_2 = 3103$ meVÅ (3100 ± 100), $B_1 = 98$ meVÅ^2, $B_2 = 23\,864$ meV Å^2 ($22\,500 \pm 1000$), $C = -3$meV, $D_1 = 149$ meV Å^2, and $D_2 = 5992$ meV Å^2 (3000).

With respect to the value of the bulk band gap of Bi_2Se_3, a large range of values has been given in the literature. The ARPES community reports most often band gaps close to 300 meV, with some values close to 200. Some of the differences might arise from changes in the lattice structure of the samples. Nechaev *et al.* [8] showed that small differences in the lattice parameters used in the *GW* calculations can change the direct band gap of Bi_2Se_3 from 190 to 340 meV. The range of values of *GW* band gaps is thus similar to the range provided by ARPES. In the calculation of the present paper, the bulk band gap is 210 meV, well within the range of most reported experimental values. In the case of the ARPES measurements, the band gap depends in addition on what is interpreted as the bulk VBM. Some works use the "valley" or the "humps" of the camelback shape as VBM (although, as discussed above, the camelback is not related to the inverted gap of Bi_2Se_3). For example, Lou *et al.* [71] position the VBM in their spectra in such a way that the band-gap value ends up above 400 meV, significantly larger than the usually reported range [72]. The ARPES works that support a direct nature of the gap place the VBM very slightly below the Dirac cone and report the smallest values, around 200 meV (see, e.g., [73]). Therefore, the particular positioning of the VBM in the spectra strongly contributes to the ambiguities in the band-gap values.

On the other hand, the optical studies [41–43,67,74–76] report overall smaller band gaps, always closer to 200 meV. Some difference between the optical gap and the quasiparticle gap given by ARPES is expected. In optical experiments, there is a significant population of electrons in the conduction band and holes in the valence band. The attractive Coulomb interaction between electrons and holes can give rise to the formation of excitons, the binding energy of which effectively lowers the optical gap with respect to the fundamental band gap. The latter, on the other hand, is measured in photoemission experiments, where electrons are not excited to conduction bands but emitted from the sample, in which case excitons play a less important role. However, in the case of Bi_2Se_3 the difference in the band gaps cannot be fully attributed to excitonic effects: Most ARPES works report an indirect gap (since they place the VBM at the humps of the M shape), whereas the optical measurements insist on the direct nature of the gap and the parabolic bands (no M shape). Therefore, both types of measured band gaps cannot be directly compared to draw conclusions about excitonic effects. From a theoretical point of view, the *GW* band gaps should be compared to the ones measured in photoemission spectroscopy. The optical

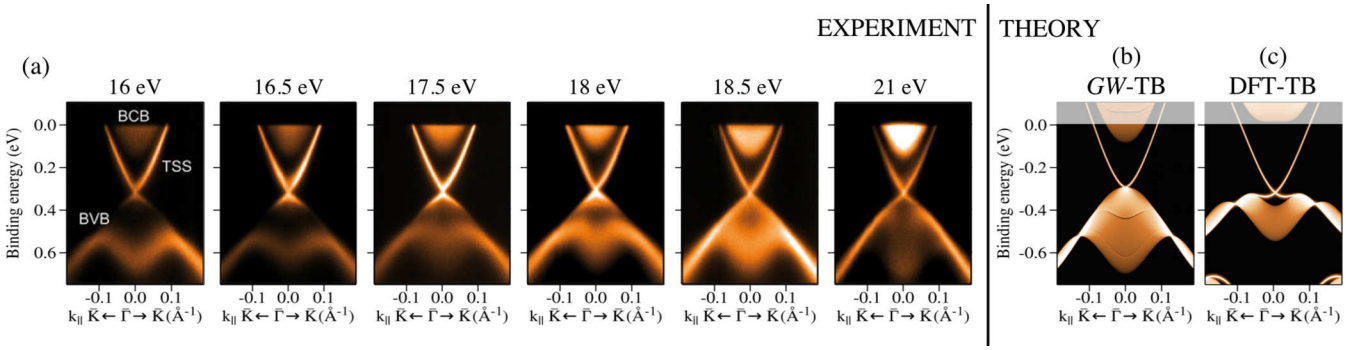

FIG. 4. (a) The photon-energy dependent ARPES spectra are reproduced with permission from Chiatti *et al.* [66]. The shaded areas in panels (b) and (c) above 0 eV correspond to energies above the experimental Fermi energy. The energy scale in the two theoretical panels has been shifted so that the energy of the Dirac point agrees with the experimental one.

gaps, on the other hand, require a different theoretical method. They can be deduced from the solution of the Bethe-Salpeter equation [77].

We now discuss how all the changes in the bulk dispersion, especially the parabolic bands, affect the dispersion of the surface states. A comparison of the theoretical spectra (DFT-TB and *GW*-TB) with ARPES for different photon energies is shown in Fig. 4. The lower DC obtained from DFT-TB is distinctly separate from the bulk valence states and has a very flat dispersion [panel (c)]. This form is consistent with previous LDA calculations (e.g., Refs. [16,78]) but clearly disagrees with what is observed in ARPES measurements [panel (a)]. A slightly more dispersive lower DC has been reported in the literature for a GGA functional instead of LDA, but the bulk camelback shape basically remains unchanged. The situation changes when many-body effects through the *GW* self-energy are included [panel (b)]. We see then that the direct nature of the band gap is predicted and, as a consequence, the lower DC strongly resonates with the bulk states. The dispersion of these resonant states is in excellent agreement with the photoemission spectra. In the case of ultrathin films, such as those studied in Refs. [7,16], the lack of a bulk continuum allows for the lower DC to be energetically separate from the valence-band quantum wells, preserving a purer 2D character. The surface resonance also affects the spin polarization, which appears significantly reduced for the lower DC [see Fig. 1(i)]. It decays rapidly as the TSS merges into the bulk valence band to values lower than 5% in the resonance region. The value increases to approximately 20% when the film is thinned down to ten QLs as a result of the partial decoupling of bulk and surface states in the ultrathin limit. The spin polarization of the upper DC, in contrast, reaches values of 70%. This result explains the much lower spin polarization of the lower DC observed in spin-integrated ARPES [79] and circular dichroic [80] ARPES, and it is consistent with the fact that the spin polarization of the lower DC is larger for six QLs than for single-crystal samples [79]. In the conduction band, surface resonances with low spin polarization are also observed when the upper DC merges with the bulk continuum.

The calculated energetic position of the Dirac point with respect to the bulk VBM can range between 3 meV (present calculation) and 35 meV, depending on the chosen lattice parameters, being in all cases much closer to the bulk valence band than to the middle of the band gap, the latter of which is sometimes deduced from DFT calculations [1,81]. This is consistent with the fact that scanning tunneling spectroscopy (STS) results show clear electronlike discrete Landau levels, but almost no holelike Landau levels [73]. The energy difference between the Dirac point and the bulk CBM is thus 207 meV in the present *GW* calculation. This is in much better agreement with experiments than the prediction by DFT (330 meV), which is almost twice as large as the experimental value (170–190 meV extracted from ARPES [65,82] or 184 meV from STS [83]).

The inclusion of many-body effects can thus explain several apparent disagreements between photoemission, optical, and transport measurements, which become consistent when interpreted on the basis of the *GW* results.

B. Warping of the upper Dirac cone of Bi_2Te_3

Whereas the DC of Bi_2Se_3 is very isotropic, the one of Bi_2Te_3 is subject to a strong warping due to the presence of the bulk states (see also Fig. 15 in Appendix B). The warping of the DC results in constant energy cuts having shapes different from a circle, for example, a hexagon or a snowflake [84]. The warping can also be inferred from the asymmetry of the DC in the $\bar{\Gamma}$-$\bar{K}$ and $\bar{\Gamma}$-$\bar{M}$ directions. In Fig. 5 we show these two directions compared to an ARPES spectrum from Ref. [13]. It is well known that the warping of the upper DC of Bi_2Te_3 is caused by the bulk off-symmetry CBM in the $\bar{\Gamma}$-$\bar{M}$ direction, indicated as B in Fig. 1(b). This bulk minimum appears too low in DFT calculations and, therefore, the DC [white in Fig. 5(b)] deviates too much from the experimental dispersion. The correct experimental dispersion of the DC is recovered when the calculations are performed with *GW* [white in Fig. 5(a)]. A similar improvement was reported in Ref. [17] for an ultrathin film of six QLs. The absolute VBM and the local CBM in the $\bar{\Gamma}$-$\bar{M}$ direction [Fig. 1(f)] are likely responsible for the onset of direct interband transitions observed in optical spectroscopy and discussed extensively in Ref. [42]. We also find excellent agreement between our calculated *GW* quasiparticle energies and experiment for states above the Fermi energy, namely, the TSS [26] and the two spin-polarized surface resonances [85] as measured in time-resolved ARPES (trARPES).

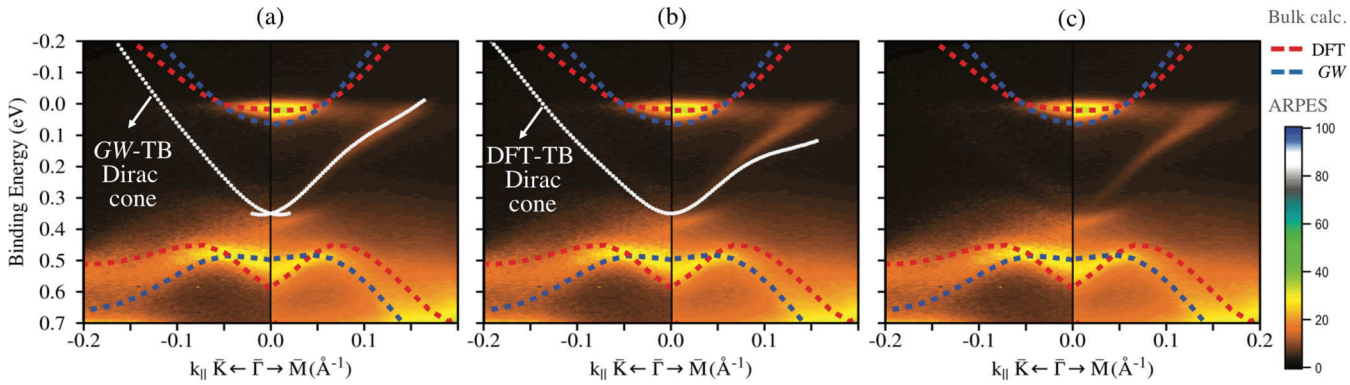

FIG. 5. Dirac cone (white) of a 100-QL film of Bi_2Te_3 obtained from (a) *GW*-TB and (b) DFT-TB calculations superposed on an ARPES spectrum. The ARPES measurements are adapted with permission from Michiardi *et al.* [13] (copyrighted by the American Physical Society) and are shown in panel (c) to observe better the experimental dispersion of the Dirac cone. The bulk bands corresponding approximately to the photon energy of the experimental spectrum are shown as dashed red (DFT) and blue (*GW*) curves in the three panels.

C. Valence band and Dirac cone of Sb_2Te_3

The bulk states and DC of Sb_2Te_3 show a similar quasiparticle renormalization as the ones of Bi_2Se_3. The maximum labeled C in Fig. 1(c) moves to lower energies—as did the one labeled A in Fig. 1(a). This maximum becomes almost degenerate with the maximum at $\bar{\Gamma}$, as discussed in Ref. [15]. For the ternary compound Bi_2Te_2Se, by comparison of Figs. 1(d) and 1(h), we see that a similar effect happens, and the bulk maximum in the $\bar{\Gamma}$-$\bar{M}$ direction is pushed down in energies. In that case, however, there is still a clear bulk VBM above the maximum at $\bar{\Gamma}$. The bulk VBM is found in ARPES measurements in the $\bar{\Gamma}$-$\bar{M}$ direction for a photon energy of 21 eV [82,86], and approximately 90 meV above the Dirac point. For this ternary compound, the dispersion of the DC does not change significantly between the DFT-TB and the *GW*-TB calculations, but the Dirac point is "buried" much deeper below the bulk VBM in the case of DFT-TB (210 meV). Again, the *GW*-TB value (117 meV) is much closer to the experimental one.

In contrast to Bi_2Se_3 and Bi_2Te_3, Sb_2Te_3 is naturally *p* doped and, therefore, the DC is not observed in static ARPES, which can only probe occupied states. In order to observe the DC of Sb_2Te_3 experimentally, one has to use a technique like trARPES, which is able to measure above the Fermi energy as well [88,89]. In Fig. 6 we show a comparison between our calculations for Sb_2Te_3 and trARPES measurements from [87]. When many-body effects are taken into account, the bulk bands in the vicinity of $\bar{\Gamma}$ are fairly parabolic. As in the case of Bi_2Se_3, this gives rise to surface resonances between the DC and the bulk states, which has also been concluded from experimental observations in Ref. [19]. In Fig. 6, the lower DC is very clearly seen as a linearly dispersing surface resonance in the experimental spectrum in panel (c). A red line is used as a guide to the eye for its linear dispersion. In the left parts of panels (a) and (b), we show our spectra calculated with *GW*-TB and DFT-TB, respectively. The experimental dispersion is indicated by the red line. Whereas the upper part of the DC has a very similar dispersion in *GW*-TB and DFT-TB, this is not the case for the lower part. The dispersion obtained from DFT-TB deviates strongly from a straight line and does not agree with the experimental one. In contrast, the *GW*-TB calculation shown in panel (a) reproduces very well the experimental band dispersion and surface-state velocity below the Dirac point.

So far, we have discussed features near the Fermi energy. Of course, the *GW* renormalization affects the whole band structure, and we show now that the agreement with experiment is indeed improved over a very wide energy range. Figure 7 shows a comparison of the *GW*-TB calculation (left) with an ARPES spectrum [90] (right) up to binding energies 3 eV below the Dirac point. Surface states appear in white and light yellow color, surface resonances appear in yellow, and bulk features appear in red. The theoretical spectrum agrees remarkably well with the experimental spectrum. A comparison of electron-energy loss spectra (EELS) on a wider energy range, up to 50 eV, was shown in Ref. [15].

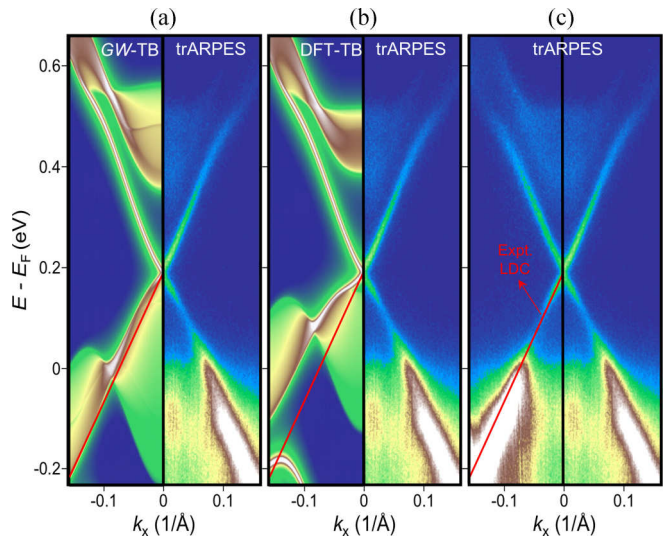

FIG. 6. On the left-hand sides of panels (a) and (b), we show the *GW*-TB and DFT-TB calculations of Sb_2Te_3, respectively. The results are compared with time-resolved ARPES shown on the right-hand sides of panels (a) and (b) and fully in panel (c). The trARPES spectra are reproduced with permission from Sumida *et al.* [87]. The experimental dispersion of the lower Dirac cone (LDC) is shown in red as a guide to the eye.

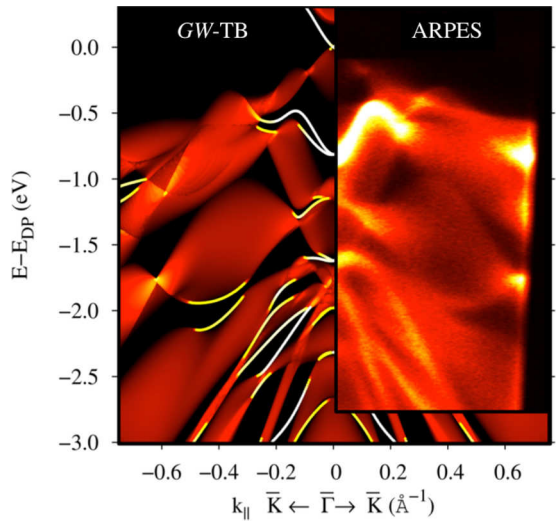

FIG. 7. *GW*-TB calculation of a 100-QL film of Sb_2Te_3 with an ARPES spectrum superimposed on the right-hand side and reproduced with permission from Eschbach and Plucinski [90]. The color coding allows surface states (white, and light yellow), surface resonances (yellow), and bulk features (red) to be distinguished. Energies are aligned such that the DP appears at 0 eV.

D. Topological characterization of Sb_2Se_3

The binary compound Sb_2Se_3 is a special case. Until very recently, rhombohedral Sb_2Se_3 was believed to be trivial, as suggested in the original paper [1] and subsequent DFT studies. Only very recently, Cao *et al.* claimed [53] in a comparative first-principles study that it is very likely a strong topological insulator, just like the rest of the family. (However, there is no experimental evidence so far.) Cao *et al.* argued that the false prediction by previous DFT studies was not related to the inherent problems of standard DFT but rather due to the usage of inaccurate structural parameters in these studies. Since the rhombohedral/hexagonal geometry is not the preferred structure for Sb_2Se_3 and its experimental growth in this geometry is extremely challenging [91], there is no literature on experimental structural parameters of this compound. Therefore, DFT calculations in the past were carried out using the DFT relaxed lattice parameters. Reference [53] showed that when DFT calculations are carried out with the lattice parameters relaxed with a functional that includes van der Waals interactions the material is predicted as topologically nontrivial.

It is uncommon that DFT falsely predicts a topological material to be trivial. DFT tends to do the opposite: Some trivial materials are predicted as nontrivial [34], a direct consequence of the overestimation of the DFT band inversion (Fig. 2). But the case of Sb_2Se_3 is different: Standard DFT can predict the correct topological character, but only if calculations are carried out with the correct lattice parameters. This is not surprising, since, after all, it is well known that strain can cause a topological phase transition [36,92] and an incorrect lattice parameter is, in a sense, equivalent to strain, potentially leading DFT to predict a different topological phase.

Figure 8 shows the DFT and *GW* band structures of bulk Sb_2Se_3, calculated with the relaxed lattice parameters from

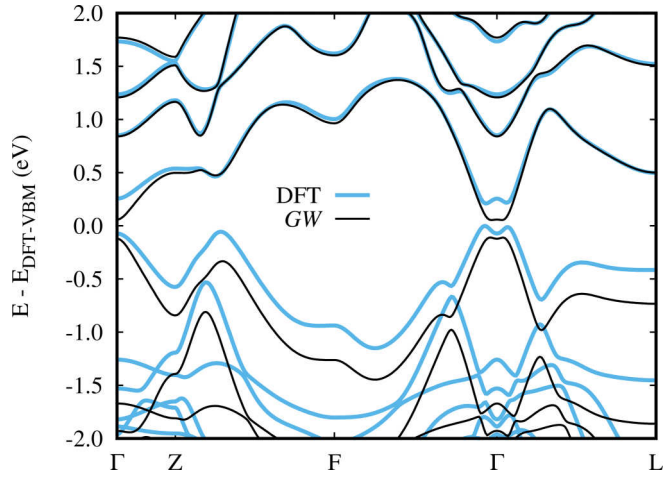

FIG. 8. *GW* and DFT band structure of bulk Sb_2Se_3. The zero energy is chosen at the VBM of the DFT calculation. The *GW* calculation shows the actual correction with respect to the DFT states, without an additional energy alignment.

Ref. [53]. One can see in this comparison a clear example of the effect discussed in Fig. 2: In the same material, both types of band gaps exist. For most of the **k** points (see, e.g., the Z, F, or L point in Fig. 8), the band gap is trivial (Fig. 2 left) and increases when many-body effects are included. At the Γ point and in its vicinity, the band gap is inverted (Fig. 2 right) and therefore decreases with the inclusion of many-body effects. The *GW* band structure is represented in Fig. 8 without any shift or alignment, showing the actual many-body corrections of the DFT states. We see that the effect on the valence states is significantly larger than that on the conduction states. This is likely due to the self-interaction correction: In LDA, the Se *p* states, which form the valence band (except around the band inversion at the Γ point), are subject to the self-interaction error and thus appear too high in energy. The *GW* approximation contains the exact exchange potential, which corrects this error and, as a consequence, pushes the states down in energy.

In Fig. 9, we present the *GW*-TB band structure of a 100-QL slab of Sb_2Se_3. It clearly shows a TSS with a strong

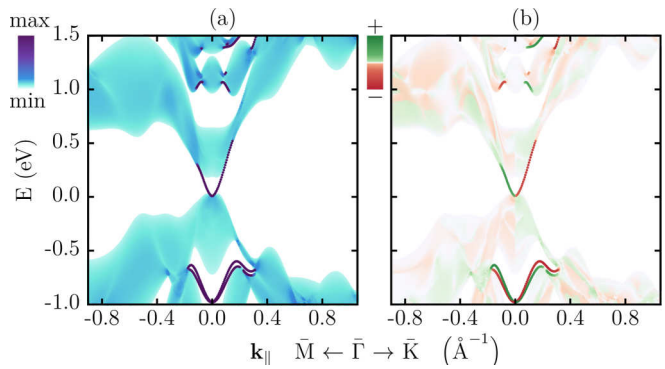

FIG. 9. *GW*-TB band structure of a 100-QL film of Sb_2Se_3. The color coding represents the degree of localization of the states on the topmost QL in panel (a) and the in-plane spin polarization in panel (b).

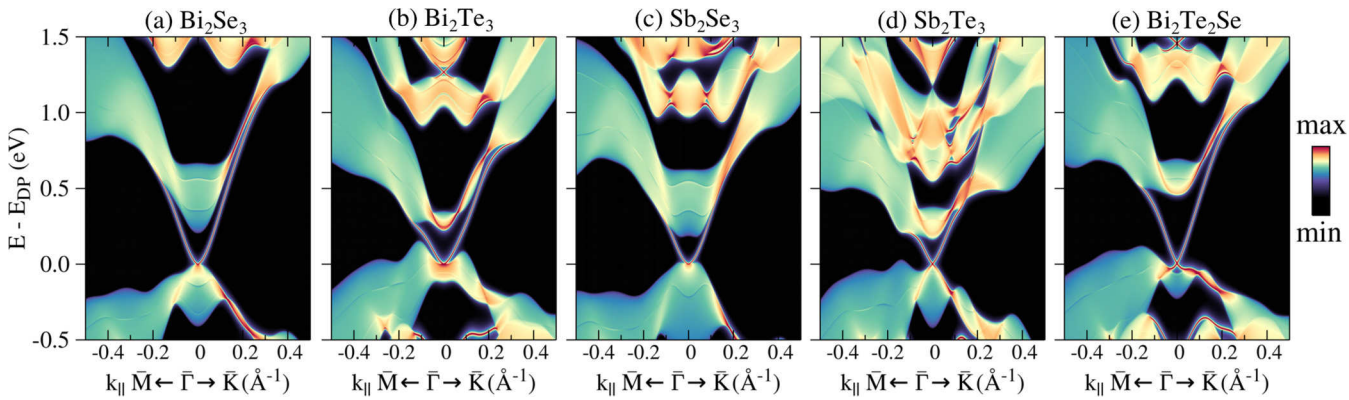

FIG. 10. *GW*-TB band structures of the five compounds in a semi-infinite geometry. Energies are aligned such that the DP appears at 0 eV.

spin polarization. To the best of our knowledge, experimental evidence of the topological nature of Sb_2Se_3, as well as its experimental lattice parameters in the rhombohedral structure, is still pending.

E. Semi-infinite systems

The Hamiltonians provided in the Supplemental Material [30] can be directly used in postprocessing programs like WANNIERTOOLS [62]. We have performed semi-infinite calculations with the iterative Green-function method [60,61] as implemented in WANNIERTOOLS. The spectral functions thus obtained for the five compounds are shown in Fig. 10. Surface states and surface resonances are clearly visible as red (darkest) and yellow (brightest) features. Bulk states appear in green and blue colors (intermediate shades when seen in grayscale).

V. SUMMARY AND CONCLUSIONS

In this paper, we have demonstrated that DFT calculations cannot explain, in a consistent manner, observations obtained with different experimental techniques and might have led to erroneous interpretations of experimental measurements in the past. We attribute this to the missing many-body effects in standard DFT functionals, which affects DFT calculations but also tight-binding calculations based on DFT parametrizations.

In the present paper, we provide improved tight-binding Hamiltonians for Bi_2Se_3, Bi_2Te_3, Sb_2Te_3, Sb_2Se_3, and Bi_2Te_2Se with *ab initio* parameters that are derived from *GW* calculations based on an all-electron description within the FLAPW method. The *GW* approach employs a diagrammatic expansion of many-body effects and thus goes beyond standard DFT calculations. We have presented calculations performed with these Hamiltonians and have discussed the most prominent differences to DFT by a comparison with photoemission experiments. We have compared states both close to the Fermi energy and at larger binding energies. In all cases, we found that the *GW* results are overall more consistent with experimental observations of photoemission, optics, EELS, and transport than DFT calculations.

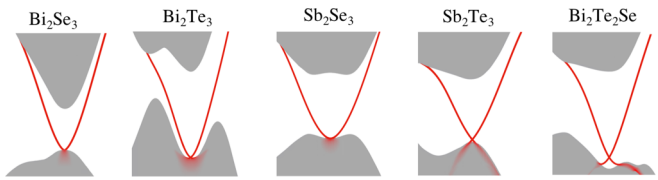

FIG. 11. Schematic summary of the behavior of the bulk states, surface states, and lower-DC resonances of the five compounds.

In contrast to previous works, we have discussed the electronic structure of bulklike systems, i.e., of systems with very many layers, and, thus, provided a revised electronic structure in which many-body effects are included and bulk states coexist with surface states and surface resonances. Figure 11 serves as a visual summary of these revised electronic structures. For most of the five stoichiometric compounds studied here, the pure surface character of the lower DC extends only a few meV below the Dirac point. The lower DC thus forms a surface resonance with the bulk valence bands. This was clearly seen in Fig. 10. These bulk valence bands are fairly parabolic in the cases of Bi_2Se_3 and Sb_2Te_3, and they present a camelback shape or M shape for Bi_2Te_3, Sb_2Se_3 (very flat), and the ternary compound Bi_2Te_2Se. In the case of DFT, on the other hand, camelback shapes are predicted for all materials, sometimes in disagreement with experiments [15,40,41,43]. We have discussed this failure of DFT to be a consequence of its underestimation of trivial band gaps and its overestimation of inverted band gaps. Our tight-binding film and semi-infinite calculations with first-principles parameters have confirmed the bulk results published in the literature and have, in addition, unveiled how they affect the topological surface states and the surface resonances with the bulk states.

The bulk *GW* tight-binding Hamiltonians of the five compounds are provided in the Supplemental Material [30] to the community. If used, the present paper should be cited. The technical details are described in Appendix B.

ACKNOWLEDGMENTS

We thank Anoop Chandran for turning our data into the three-dimensional plots of the Dirac cones in Fig. 15. We also thank Gustav Bihlmayer, Adam Dubroka, Philip Hofmann, Henriette Lund, Gregor Mussler, Ilya A. Nechaev, Milan Orlita, and Jalil A. Rehman for helpful discussions. We are

grateful to Akio Kimura, Matteo Michiardi, Lukasz Plucinski, and Jaime Sánchez-Barriga, as well as their groups, for providing or giving permission to use their photoemission data. The work was supported by the Helmholtz Association through the Virtual Institute for Topological Insulators. We gratefully acknowledge the computing time granted through JARA-HPC on the supercomputer JURECA at Forschungszentrum Jülich.

APPENDIX A: WANNIER INTERPOLATION

A review on Wannier functions and the Wannier interpolation can be found in Ref. [47]. For completeness, the method is briefly outlined here. We employ a transformation from Bloch states $\phi_{\mathbf{k}n}(\mathbf{r})$, which are defined on a finite $\mathbf{k}$-point set (e.g., $4 \times 4 \times 4$) to Wannier functions $w_{\mathbf{R}m}(\mathbf{r})$. Each Wannier function is localized on a specific atom of the corresponding supercell ($4 \times 4 \times 4$ in the above example), where $\mathbf{R}$ is the lattice vector of the unit cell that contains the atom. For simplicity, we consider m as a composite index composed of the atom index (if the unit cell contains more than one atom) and different Wannier orbitals (e.g., s-like, p_x-like, etc.). We first consider the unit cell at the origin for which $\mathbf{R} = \mathbf{0}$. The Wannier functions are given by linear combinations of Bloch states:

$$w_{\mathbf{0}m}(\mathbf{r}) = \frac{1}{N} \sum_{\mathbf{k},n} U_{mn}(\mathbf{k}) \phi_{\mathbf{k}n}(\mathbf{r}), \tag{A1}$$

where N is the number of $\mathbf{k}$ points, equivalently the number of unit cells in the supercell, and the coefficients $U_{mn}(\mathbf{k})$ are often determined from the condition that the Wannier functions be maximally localized and orthonormal in the supercell. (If there are as many Wannier functions as there are Bloch functions, this means that U is a unitary matrix.) Maximally localized Wannier functions can be obtained by the WANNIER90 library [21]. However, in Fig. 12 we show that for these materials (Bi_2Se_3 is shown as a representative) whether using maximally localized or first-guess Wannier functions (as defined in Sec. III) does not make visible differences in the DFT and *GW* interpolated band structures.

By virtue of the Bloch theorem, the Wannier functions in neighboring unit cells $\mathbf{R} \neq \mathbf{0}$ are generated by an additional exponential factor:

$$\begin{aligned} w_{\mathbf{R}m}(\mathbf{r}) &= w_{\mathbf{0}m}(\mathbf{r} - \mathbf{R}) \\ &= \frac{1}{N} \sum_{\mathbf{k},n} e^{-i\mathbf{k}\mathbf{R}} U_{mn}(\mathbf{k}) \phi_{\mathbf{k}n}(\mathbf{r}). \end{aligned} \tag{A2}$$

With the help of the "Wannier Bloch functions"

$$w_{\mathbf{k}m}(\mathbf{r}) = \sum_{n} U_{mn}(\mathbf{k}) \phi_{\mathbf{k}n}(\mathbf{r}), \tag{A3}$$

we can write Eq. (A2) as simple Fourier transforms:

$$w_{\mathbf{R}m}(\mathbf{r}) = \frac{1}{N} \sum_{\mathbf{k}} e^{-i\mathbf{k}\mathbf{R}} w_{\mathbf{k}m}(\mathbf{r}), \tag{A4}$$

the inverse Fourier transform of which is

$$w_{\mathbf{k}m}(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\mathbf{R}} w_{\mathbf{R}m}(\mathbf{r}). \tag{A5}$$

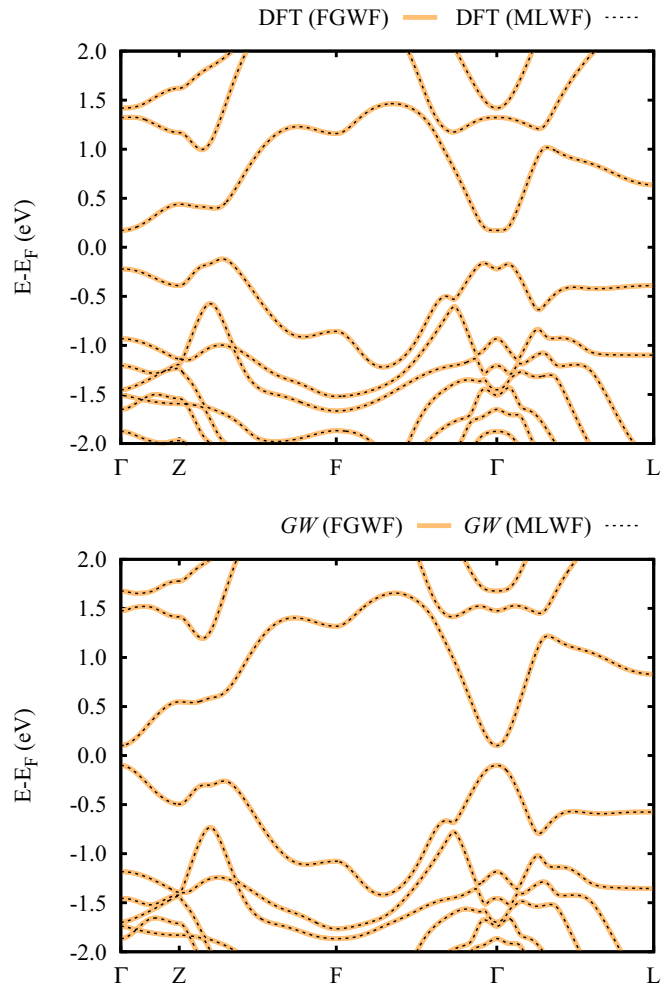

FIG. 12. DFT (top) and *GW* (bottom) interpolated band structures of Bi_2Se_3, comparing the ones obtained with maximally localized Wannier functions (MLWF) and with first-guess Wannier functions (FGWF).

We will generalize Eq. (A5) to an arbitrary $\mathbf{q}$ vector (i.e., one that may not be an element of the $\mathbf{k}$-point set) in the following. The Wannier functions [Eq. (A4)] are localized in a supercell but repeat themselves in all other supercells without a phase factor. This is a consequence of the Bloch theorem; the exponential factor in Eq. (A2) becomes 1 if $\mathbf{R}$ is a lattice vector that connects two supercells. Let us assume for now that the Wannier functions do *not* repeat themselves, that they are normalized (with respect to the whole space), and that $\mathbf{R}$ can run over all (i.e., infinitely many) lattice vectors. Then, we can define a general Wannier Bloch function:

$$w_{\mathbf{q}m}(\mathbf{r}) = \frac{1}{\sqrt{M}} \sum_{\mathbf{R}}^{\text{all}} e^{i\mathbf{q}\mathbf{R}} w_{\mathbf{R}m}(\mathbf{r}), \tag{A6}$$

where $\mathbf{R}$ runs over all lattice vectors and M is the (infinite) number of lattice vectors, equivalent to the (infinite) number of unit cells in the crystal. So, $w_{\mathbf{q}m}$ is a Bloch function for $\mathbf{q}$ by construction, and it is normalized with respect to the whole space by virtue of the prefactor $1/\sqrt{M}$. Now we can diagonalize the Hamiltonian at this $\mathbf{q}$ vector, which interpolates

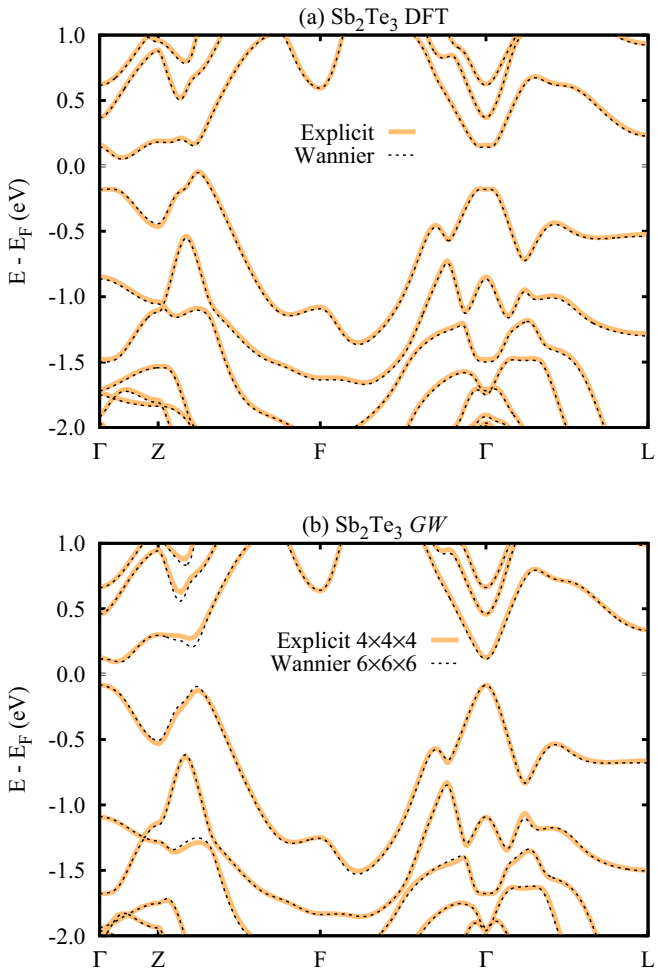

FIG. 13. Band structures of bulk Sb_2Te_3, comparing the ones obtained in an explicit (a) DFT and (b) *GW* calculation (light orange) with that obtained with the Wannier interpolation technique (dashed, black).

the band structure (Wannier interpolation). The Hamiltonian matrix written in terms of $w_{\mathbf{q}m}(\mathbf{r})$ is

$$\begin{aligned} H_{mm'}(\mathbf{q}) &= \langle w_{\mathbf{q}m}|H|w_{\mathbf{q}m'}\rangle_{\mathrm{ws}} \\ &= \frac{1}{M}\sum_{\mathbf{R},\mathbf{R}'}^{\mathrm{all}} e^{i\mathbf{q}(\mathbf{R}'-\mathbf{R})}\langle w_{\mathbf{R}m}|H|w_{\mathbf{R}'m'}\rangle_{\mathrm{ws}} \\ &= \sum_{\mathbf{R}}^{\mathrm{all}} e^{i\mathbf{q}\mathbf{R}}\langle w_{\mathbf{0}m}|H|w_{\mathbf{R}m'}\rangle_{\mathrm{ws}} \\ &\approx \sum_{\mathbf{R}} e^{i\mathbf{q}\mathbf{R}}\langle w_{\mathbf{0}m}|H|w_{\mathbf{R}m'}\rangle_{\mathrm{sc}} \\ &= \frac{1}{N}\sum_{\mathbf{R}}\sum_{\mathbf{k}} e^{i(\mathbf{q}-\mathbf{k})\mathbf{R}}\langle w_{\mathbf{k}m}|H|w_{\mathbf{k}m'}\rangle_{\mathrm{uc}} \\ &= \sum_{\mathbf{k}}\left[\frac{1}{N}\sum_{\mathbf{R}} e^{i(\mathbf{q}-\mathbf{k})\mathbf{R}}\right] H_{mm'}(\mathbf{k}), \end{aligned} \tag{A7}$$

where we have specified with "ws" (whole space), "sc" (supercell), and "uc" (unit cell) over which volume the integration in $\langle\cdot|H|\cdot\rangle$ is to be taken. The approximation ($\approx$) replaces the infinite sum over $\mathbf{R}$ by a finite one that is restricted to the supercell. (Here, the supercell should have the form of a Wigner-Seitz cell.) It utilizes the fact that $H_{mm'}(\mathbf{R})$ falls off exponentially with $\mathbf{R}$. Equation (A7) defines an approximated Hamiltonian matrix at an arbitrary $\mathbf{k}$ point $\mathbf{q}$ not necessarily contained in the original $\mathbf{k}$-point set for the *ab initio* calculation. Here, it becomes evident that taking the approximation ($\approx$) above is essential for the interpolation to a finer mesh, otherwise a back-and-forth Fourier transformation would be an identity operation without any effect.

Although we have specifically defined H to be a single-particle Hamiltonian operator, the derivation of Eq. (A7) works generally for arbitrary local operators H and even for nonlocal operators, as long as they fall off sufficiently fast in real space. In this sense, H may be the Hermitian Hamiltonian of a single-particle system, but it may also be the complex self-energy operator calculated in the *GW* method. When used with the latter, not only can the energies of the quasiparticle states be interpolated but also their lifetimes (imaginary parts of the quasiparticle energies).

Specifically, if H is the single-particle Hamiltonian, the matrix elements on the right-hand side of Eq. (A7) read

$$H_{mm'}(\mathbf{k}) = \sum_n U^*_{mn}(\mathbf{k})U_{m'n}(\mathbf{k})\epsilon_{\mathbf{k}n}, \tag{A8}$$

where $\epsilon_{\mathbf{k}n}$ are the energy eigenvalues. In the context of the one-shot *GW* method, the latter can also be taken as the complex quasiparticle energies, giving rise to non-Hermitian complex matrices $H_{mm'}(\mathbf{k})$ to be used in Eq. (A7). The resulting interpolated matrices $H_{mm'}(\mathbf{q})$ and their eigenvalues are then complex, too. In this way, Eq. (A7) interpolates both the quasiparticle energies and lifetimes.

For the calculations in this paper, we have used a scheme equivalent to that activated when the keyword `use_ws_distance` is used in the WANNIER90 code (see www.wannier.org/doc/user_guide.pdf). This improves in general the $\mathbf{k}$-point convergence of the Wannier interpolation and avoids a potential breaking of the Kramers degeneracy.

In Fig. 13, we show the DFT-LDA and *GW* band structures of Sb_2Te_3, comparing the interpolated and the explicitly calculated results. The agreement is excellent and only in the case of *GW* there are some small deviations between the two band structures for a few $\mathbf{k}$ points (mostly in the Z-F direction). These differences are not due to the Wannier interpolation technique itself, but due to the fact that the explicit *GW* band structure was obtained for a $4\times4\times4$ $\mathbf{k}$ mesh, whereas the Wannier interpolation was performed starting from a *GW* calculation based on a $6\times6\times6$ $\mathbf{k}$ mesh ($4\times4\times4$ $\mathbf{k}$ points were not enough to obtain a good interpolation). Due to the computational cost of *GW* calculations, obtaining an explicit *GW* band structure with the $8\times8\times8$ mesh used in the main text is computationally very demanding. Nevertheless, for illustration purposes, the agreement for the $6\times6\times6$ interpolated band structure is already excellent.

APPENDIX B: *GW* TIGHT-BINDING HAMILTONIANS

The format of the Hamiltonians provided in the Supplemental Material [30] is the same as the one used by the

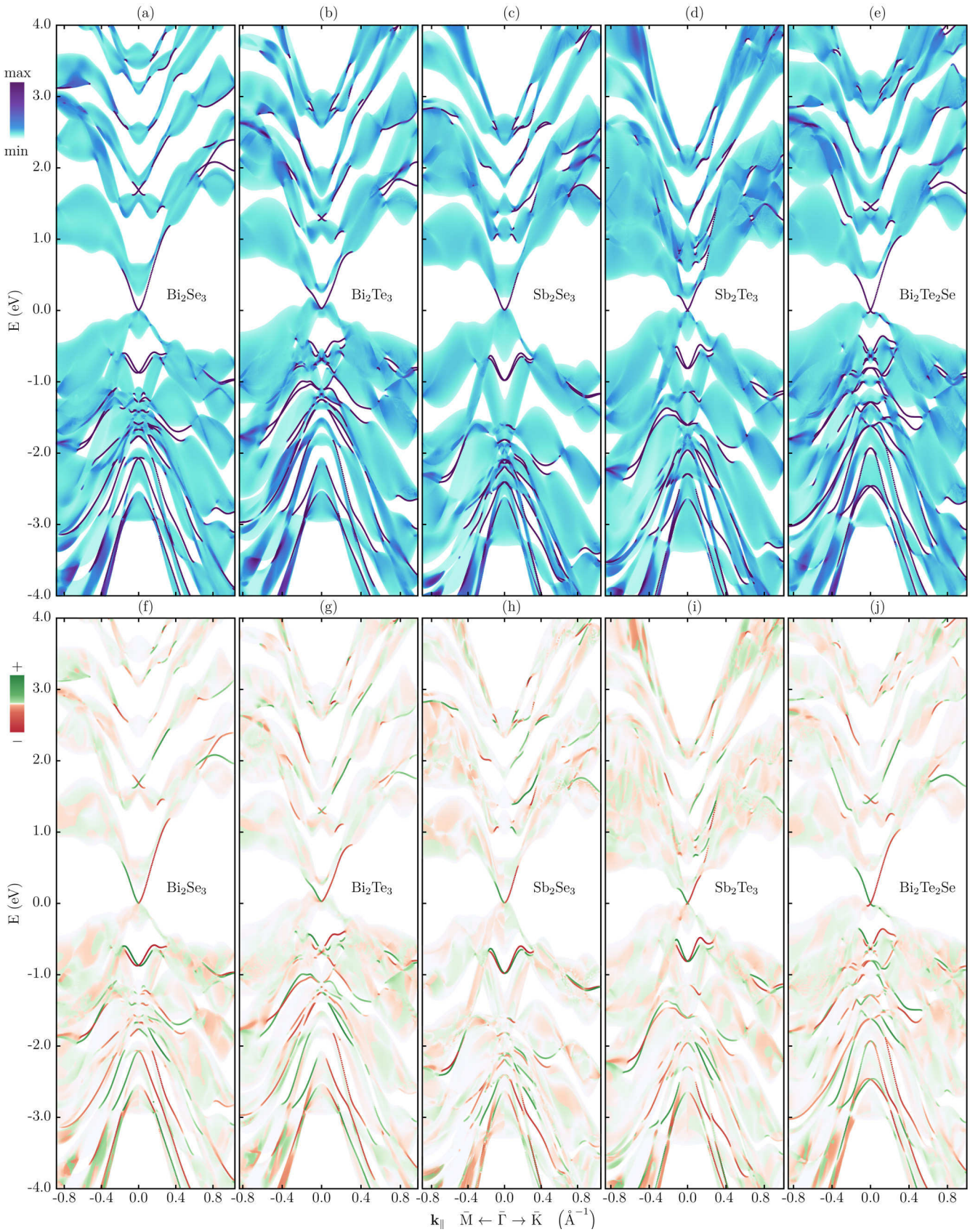

FIG. 14. Same as panels (e)–(l) of Fig. 1 for a larger energy range.

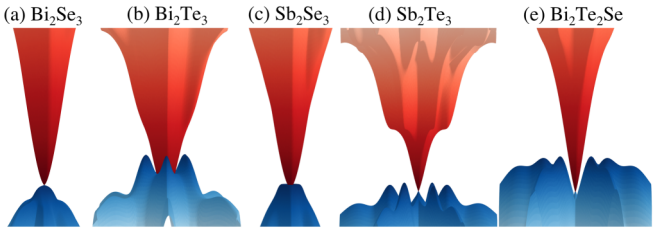

FIG. 15. Electronic band structure of the five compounds represented in three dimensions.

WANNIER90 package [21]: first line, a description of the data; second line, number of Wannier functions; third line, number of **R** points in the Wigner-Seitz cell. Then there is a block of integers that give the degeneracy of each of the **R** points. Each line contains 15 entries. These are followed by one line per Hamiltonian matrix element. Each line shows the components of the **R** vector in the basis of the lattice basis vectors, the i and j indices of the matrix element H_{ij}, followed by two real numbers giving the real and imaginary parts of H_{ij}. We chose this format because it is used by a number of post-processing programs like WANNIERTOOLS. The results of slab calculations performed with these Hamiltonians are shown in Fig. 14 for a large energy range. Another advantage of the use of the Wannier interpolation method is the possibility to map the whole Brillouin zone very densely. For example, this makes the explicit computation of Fermi surfaces and anomalous Hall conductivity of realistic materials possible, which, otherwise, cannot be carried out by first-principles calculations due to the huge set of **k** points needed to achieve convergence. It also allows the anisotropy of the Dirac cone to be visualized by representing it in the full two-dimensional Brillouin zone. The resulting three-dimensional representations of the *GW*-TB Dirac cones are shown in Fig. 15.

The Hamiltonians were obtained with first-guess Wannier functions (see Fig. 12), i.e., setting `num_iter=0` in the input file of WANNIER90. For all compounds, we used Wannier functions with orbital character p_z, p_x, p_y (the order is chosen for compatibility with the WANNIER90 code). The order of the basis functions is thus

$$p_z^{1\uparrow}\ p_x^{1\uparrow}\ p_y^{1\uparrow}\ p_z^{2\uparrow}\ p_x^{2\uparrow}\ p_y^{2\uparrow}\ \ p_z^{3\uparrow}\ p_x^{3\uparrow}\ p_y^{3\uparrow}\ \ p_z^{4\uparrow}\ p_x^{4\uparrow}\ p_y^{4\uparrow}\ \ p_z^{5\uparrow}\ p_x^{5\uparrow}\ p_y^{5\uparrow}$$
$$p_z^{1\downarrow}\ p_x^{1\downarrow}\ p_y^{1\downarrow}\ p_z^{2\downarrow}\ p_x^{2\downarrow}\ p_y^{2\downarrow}\ \ p_z^{3\downarrow}\ p_x^{3\downarrow}\ p_y^{3\downarrow}\ \ p_z^{4\downarrow}\ p_x^{4\downarrow}\ p_y^{4\downarrow}\ \ p_z^{5\downarrow}\ p_x^{5\downarrow}\ p_y^{5\downarrow},$$

where the number indicates the atom and the arrow refers to the spin. The Hamiltonian matrix elements in the files provided in the Supplemental Material [30] are given in (hartree) atomic units. We used the rhombohedral lattice structure with five atoms in the unit cell. The atomic positions and lattice vectors in the calculations are summarized in Table I. To improve the **k**-point convergence of the Wannier interpolation and to allow for a straightforward calculation of the slab, the atoms were backfolded such that they all belong to the same quintuple layer. The backfolding vectors were the following linear combinations of lattice vectors: atom 1, $-\mathbf{a}_1+\mathbf{a}_3$; atom 2, $-\mathbf{a}_1$; atom 3, $+\mathbf{a}_3$; atom 4, $-\mathbf{a}_1$; atom 5, $+\mathbf{a}_3$. (The Wannier centers are approximately on the backfolded atomic positions.)

TABLE I. Lattice vectors, atomic positions, and Fermi energy of the five compounds.

		Bi_2Se_3		
Lattice	$\mathbf{a}_1$	0.000000	−4.512025	18.024962
vectors	$\mathbf{a}_2$	3.907528	2.256012	18.024962
(bohr)	$\mathbf{a}_3$	−3.907528	2.256012	18.024962
Atomic	Se	0.0000	0.0000	0.0000
positions	Se	0.2109	0.2109	0.2109
(relative	Se	−0.2109	−0.2109	−0.2109
coordinates)	Bi	0.4006	0.4006	0.4006
	Bi	−0.4006	−0.4006	−0.4006
Fermi energy	0.213814	(hartree)		
		Bi_2Te_3		
Lattice	$\mathbf{a}_1$	0.000000	−4.782547	19.204021
vectors	$\mathbf{a}_2$	4.141807	2.391274	19.204021
(bohr)	$\mathbf{a}_3$	−4.141807	2.391274	19.204021
Atomic	Te	0.0000	0.0000	0.0000
positions	Te	0.2120	0.2120	0.2120
(relative	Te	−0.2120	−0.2120	−0.2120
coordinates)	Bi	0.4000	0.4000	0.4000
	Bi	−0.4000	−0.4000	−0.4000
Fermi energy	0.221526	(hartree)		
		Sb_2Se_3		
Lattice	$\mathbf{a}_1$	0.000000	−4.425957	17.982656
vectors	$\mathbf{a}_2$	3.832991	2.212979	17.982656
(bohr)	$\mathbf{a}_3$	−3.832991	2.212979	17.982656
Atomic	Se	0.0000	0.0000	0.0000
positions	Se	0.2120	0.2120	0.2120
(relative	Se	−0.2120	−0.2120	−0.2120
coordinates)	Sb	0.4000	0.4000	0.4000
	Sb	−0.4000	−0.4000	−0.4000
Fermi energy	0.210445	(hartree)		
		Sb_2Te_3		
Lattice	$\mathbf{a}_1$	0.000000	−4.651968	19.186050
vectors	$\mathbf{a}_2$	4.028722	2.325984	19.186050
(bohr)	$\mathbf{a}_3$	−4.028722	2.325984	19.186050
Atomic	Te	0.0000	0.0000	0.0000
positions	Te	0.2128	0.2128	0.2128
(relative	Te	−0.2128	−0.2128	−0.2128
coordinates)	Sb	0.3988	0.3988	0.3988
	Sb	−0.3988	−0.3988	−0.3988
Fermi energy	0.228224	(hartree)		
		Bi_2Te_2Se		
Lattice	$\mathbf{a}_1$	0.000000	−4.667741	18.808595
vectors	$\mathbf{a}_2$	4.042382	2.333871	18.808595
(bohr)	$\mathbf{a}_3$	−4.042382	2.333871	18.808595
Atomic	Se	0.0000	0.0000	0.0000
positions	Te	0.2117	0.2117	0.2117
(relative	Te	−0.2117	−0.2117	−0.2117
coordinates)	Bi	0.3961	0.3961	0.3961
	Bi	−0.3961	−0.3961	−0.3961
Fermi energy	0.272361	(hartree)		

The following files are provided—hamiltonian_Sb2Te3_GWreal.txt, hamiltonian_Sb2Te3_GWcomplex.txt, and wannier_centers_Sb2Te3.txt—for the Hamiltonians with real and complex eigenvalues, respectively (the real part of both being practically identical), as well as the list of Wannier centers.

[1] H. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, Nat. Phys. **5**, 438 (2009).

[2] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[3] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[4] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[5] L. Hedin, Phys. Rev. **139**, A796 (1965).

[6] E. Kioupakis, M. L. Tiago, and S. G. Louie, Phys. Rev. B **82**, 245203 (2010).

[7] O. V. Yazyev, E. Kioupakis, J. E. Moore, and S. G. Louie, Phys. Rev. B **85**, 161101(R) (2012).

[8] I. A. Nechaev, R. C. Hatch, M. Bianchi, D. Guan, C. Friedrich, I. Aguilera, J. L. Mi, B. B. Iversen, S. Blügel, P. Hofmann, and E. V. Chulkov, Phys. Rev. B **87**, 121111(R) (2013).

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

[10] I. A. Nechaev and E. V. Chulkov, Phys. Rev. B **88**, 165135 (2013).

[11] I. Aguilera, C. Friedrich, and S. Blügel, Phys. Rev. B **88**, 165136 (2013).

[12] I. P. Rusinov, I. A. Nechaev, and E. V. Chulkov, JETP Lett. **98**, 397 (2013).

[13] M. Michiardi, I. Aguilera, M. Bianchi, V. E. de Carvalho, L. O. Ladeira, N. G. Teixeira, E. A. Soares, C. Friedrich, S. Blügel, and P. Hofmann, Phys. Rev. B **90**, 075105 (2014).

[14] I. Aguilera, I. A. Nechaev, C. Friedrich, S. Blügel, and E. V. Chulkov, in *Topological Insulators: Fundamentals and Perspectives*, edited by F. Ortmann, S. Roche, and S. E. Valenzuela (Wiley, New York, 2015).

[15] I. A. Nechaev, I. Aguilera, V. De Renzi, A. di Bona, A. Lodi Rizzini, A. M. Mio, G. Nicotra, A. Politano, S. Scalese, Z. S. Aliev, M. B. Babanly, C. Friedrich, S. Blügel, and E. V. Chulkov, Phys. Rev. B **91**, 245123 (2015).

[16] T. Förster, P. Krüger, and M. Rohlfing, Phys. Rev. B **92**, 201404(R) (2015).

[17] T. Förster, P. Krüger, and M. Rohlfing, Phys. Rev. B **93**, 205442 (2016).

[18] C. Datzer, A. Zumbülte, J. Braun, T. Förster, A. B. Schmidt, J. Mi, B. Iversen, P. Hofmann, J. Minár, H. Ebert, P. Krüger, M. Rohlfing, and M. Donath, Phys. Rev. B **95**, 115401 (2017).

[19] C. Seibel, H. Bentmann, J. Braun, J. Minár, H. Maaß, K. Sakamoto, M. Arita, K. Shimada, H. Ebert, and F. Reinert, Phys. Rev. Lett. **114**, 066802 (2015).

[20] C. Jozwiak, J. A. Sobota, K. Gotlieb, A. F. Kemper, C. R. Rotundu, R. J. Birgeneau, Z. Hussain, D.-H. Lee, Z.-X. Shen, and A. Lanzara, Nat. Commun. **7**, 13143 (2016).

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

[22] T. Rauch, M. Flieger, J. Henk, and I. Mertig, Phys. Rev. B **88**, 245120 (2013).

[23] T. Rauch, M. Flieger, J. Henk, I. Mertig, and A. Ernst, Phys. Rev. Lett. **112**, 016802 (2014).

[24] T. Rauch, S. Achilles, J. Henk, and I. Mertig, Phys. Rev. Lett. **114**, 236805 (2015).

[25] Z.-H. Zhu, C. N. Veenstra, G. Levy, A. Ubaldini, P. Syers, N. P. Butch, J. Paglione, M. W. Haverkort, I. S. Elfimov, and A. Damascelli, Phys. Rev. Lett. **110**, 216401 (2013).

[26] M. Battiato, I. Aguilera, and J. Sánchez-Barriga, Materials **10**, 810 (2017).

[27] M. Bahramy, P. King, A. de la Torre, J. Chang, M. Shi, L. Patthey, G. Balakrishnan, P. Hofmann, R. Arita, N. Nagaosa, and F. Baumberger, Nat. Commun. **3**, 1159 (2012).

[28] E. Annese, T. Okuda, E. F. Schwier, H. Iwasawa, K. Shimada, M. Natamane, M. Taniguchi, I. P. Rusinov, S. V. Eremeev, K. A. Kokh, V. A. Golyashov, O. E. Tereshchenko, E. V. Chulkov, and A. Kimura, Phys. Rev. B **97**, 205113 (2018).

[29] T. A. Niehaus, M. Rohlfing, F. Della Sala, A. Di Carlo, and T. Frauenheim, Phys. Rev. A **71**, 022508 (2005).

[30] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.100.155147 for the parametrized TB Hamiltonians used in this paper and all files necessary to reproduce the present results.

[31] It should be noted that the *GW* method is a many-body method and gives rise to quasiparticle energies, which are complex, and their imaginary parts give information about the quasiparticle lifetimes. The Hamiltonians used for the results in this paper have thus complex eigenvalues (see Appendix A). To facilitate the use of these Hamiltonians by the community, we also provide Hamiltonians with real eigenvalues.

[32] J. Sánchez-Barriga, M. Battiato, E. Golias, A. Varykhalov, L. V. Yashina, O. Kornilov, and O. Rader, Appl. Phys. Lett. **110**, 141605 (2017).

[33] F. Freyse, M. Battiato, L. V. Yashina, and J. Sánchez-Barriga, Phys. Rev. B **98**, 115132 (2018).

[34] J. Vidal, X. Zhang, L. Yu, J.-W. Luo, and A. Zunger, Phys. Rev. B **84**, 041109(R) (2011).

[35] J. Liu and D. Vanderbilt, Phys. Rev. B **88**, 224202 (2013).

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

[37] J. Sánchez-Barriga, I. Aguilera, L. V. Yashina, D. Y. Tsukanova, F. Freyse, A. N. Chaika, C. Callaert, A. M. Abakumov, J. Hadermann, A. Varykhalov, E. D. L. Rienks, G. Bihlmayer, S. Blügel, and O. Rader, Phys. Rev. B **98**, 235110 (2018).

[38] M. Bianchi, D. Guan, S. Bao, J. Mi, B. B. Iversen, P. D. King, and P. Hofmann, Nat. Commun. **1**, 128 (2010).

[39] C. M. Acosta, M. P. Lima, A. J. R. da Silva, A. Fazzio, and C. H. Lewenkopf, Phys. Rev. B **98**, 035106 (2018).

[40] B. A. Piot, W. Desrat, D. K. Maude, M. Orlita, M. Potemski, G. Martinez, and Y. S. Hor, Phys. Rev. B **93**, 155206 (2016).

[41] M. Orlita, B. A. Piot, G. Martinez, N. K. Sampath Kumar, C. Faugeras, M. Potemski, C. Michel, E. M. Hankiewicz, T. Brauner, Č. Drašar, S. Schreyeck, S. Grauer, K. Brunner, C. Gould, C. Brüne, and L. W. Molenkamp, Phys. Rev. Lett. **114**, 186401 (2015).

[42] A. Dubroka, O. Caha, M. Hronček, P. Friš, M. Orlita, V. Holý, H. Steiner, G. Bauer, G. Springholz, and J. Humlíček, Phys. Rev. B **96**, 235202 (2017).

[43] A. Miyata, Z. Yang, A. Surrente, O. Drachenko, D. K. Maude, O. Portugall, L. B. Duffy, T. Hesjedal, P. Plochocka, and R. J. Nicholas, Phys. Rev. B **96**, 121111(R) (2017).

[44] C. Pauly, G. Bihlmayer, M. Liebmann, M. Grob, A. Georgi, D. Subramaniam, M. R. Scholz, J. Sánchez-Barriga, A. Varykhalov, S. Blügel, O. Rader, and M. Morgenstern, Phys. Rev. B **86**, 235106 (2012).

[45] H. Jiang, J. Phys. Chem. C **116**, 7664 (2012).

[46] L. Plucinski, A. Herdt, S. Fahrendorf, G. Bihlmayer, G. Mussler, S. Döring, J. Kampmeier, F. Matthes, D. E. Bürgler, D. Grützmacher, S. Blügel, and C. M. Schneider, J. Appl. Phys. **113**, 053706 (2013).

[47] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, Rev. Mod. Phys. **84**, 1419 (2012).

[48] www.flapw.de.

[49] C. Friedrich, S. Blügel, and A. Schindlmayr, Phys. Rev. B **81**, 125102 (2010).

[50] C. P. Vicente, J. L. Tirado, K. Adouby, J. Jumas, A. A. Touré, and G. Kra, Inorg. Chem. **38**, 2131 (1999).

[51] R. W. G. Wyckoff, *Crystal Structures 2* (Wiley, New York, 1964).

[52] T. L. Anderson and H. B. Krause, Acta Crystallogr. Sect. B **30**, 1307 (1974).

[53] G. Cao, H. Liu, J. Liang, L. Cheng, D. Fan, and Z. Zhang, Phys. Rev. B **97**, 075147 (2018).

[54] A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).

[55] J. M. Crowley, J. Tahir-Kheli, and W. A. Goddard, J. Phys. Chem. Lett. **6**, 3792 (2015).

[56] D. di Sante (private communication).

[57] S. Park and B. Ryu, J. Korean Phys. Soc. **69**, 1683 (2016).

[58] R. Sakuma, C. Friedrich, T. Miyake, S. Blügel, and F. Aryasetiawan, Phys. Rev. B **84**, 085144 (2011).

[59] D. Krause and P. Thörnig, J. Large-Scale Res. Facil. **4**, A132 (2018).

[60] F. Guinea, C. Tejedor, F. Flores, and E. Louis, Phys. Rev. B **28**, 4397 (1983).

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

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

[63] I. Souza, N. Marzari, and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2001).

[64] H. Plank, J. Pernul, S. Gebert, S. N. Danilov, J. König-Otto, S. Winnerl, M. Lanius, J. Kampmeier, G. Mussler, I. Aguilera, D. Grützmacher, and S. D. Ganichev, Phys. Rev. Mater. **2**, 024202 (2018).

[65] Y. L. Chen, J.-H. Chu, J. G. Analytis, Z. K. Liu, K. Igarashi, H.-H. Kuo, X. L. Qi, S. K. Mo, R. G. Moore, D. H. Lu, M. Hashimoto, T. Sasagawa, S. C. Zhang, I. R. Fisher, Z. Hussain, and Z. X. Shen, Science **329**, 659 (2010).

[66] O. Chiatti, C. Riha, D. Lawrenz, M. Busch, S. Dusari, J. Sánchez-Barriga, A. Mogilatenko, L. V. Yashina, S. Valencia, A. A. Ünal, O. Rader, and S. F. Fischer, Sci. Rep. **6**, 27483 (2016).

[67] H. Köhler and J. Hartmann, Phys. Status Solidi B **63**, 171 (1974).

[68] If there were an inverted gap at the Z point in addition to the one at Γ, Bi_2Se_3 would be topologically trivial.

[69] Since our calculations do not include a band-bending surface potential, the enhancement of the camelback shape due to this effect is not seen in our results.

[70] I. A. Nechaev and E. E. Krasovskii, Phys. Rev. B **94**, 201410(R) (2016).

[71] R. Lou, Z. Liu, W. Jin, H. Wang, Z. Han, K. Liu, X. Wang, T. Qian, Y. Kushnirenko, S.-W. Cheong, R. M. Osgood, H. Ding, and S. Wang, Phys. Rev. B **92**, 115150 (2015).

[72] Lou *et al.* [71] also reported a surprising sudden closure of the bulk band gap of $Bi_{2-x}In_xSe_3$ across a topological-to-trivial transition caused by the variation of the In concentration. If the measurements were interpreted based on the VBM being located very close to the Dirac point, such an anomalous step in the evolution of the gap would disappear or become much less noticeable.

[73] P. Cheng, C. Song, T. Zhang, Y. Zhang, Y. Wang, J.-F. Jia, J. Wang, Y. Wang, B.-F. Zhu, X. Chen, X. Ma, K. He, L. Wang, X. Dai, Z. Fang, X. Xie, X.-L. Qi, C.-X. Liu, S.-C. Zhang, and Q.-K. Xue, Phys. Rev. Lett. **105**, 076801 (2010).

[74] K. W. Post, B. C. Chapler, L. He, X. Kou, K. L. Wang, and D. N. Basov, Phys. Rev. B **88**, 075121 (2013).

[75] A. V. Galeeva, S. G. Egorova, V. I. Chernichkin, M. E. Tamm, L. V. Yashina, V. V. Rumyantsev, S. V. Morozov, H. Plank, S. N. Danilov, L. I. Ryabova, and D. R. Khokhlov, Semicond. Sci. Technol. **31**, 095010 (2016).

[76] G. Martinez, B. A. Piot, M. Hakl, M. Potemski, Y. S. Hor, A. Materna, S. G. Strzelecka, A. Hruban, O. Caha, J. Novak, C. D. A. Dubroka, and M. Orlita, Sci. Rep. **7**, 6891 (2017).

[77] E. E. Salpeter and H. A. Bethe, Phys. Rev. **84**, 1232 (1951).

[78] Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nat. Phys. **5**, 398 (2009).

[79] M. Neupane, A. Richardella, J. Sánchez-Barriga, S. Xu, N. Alidoust, I. Belopolski, C. Liu, G. Bian, D. Zhang, D. Marchenko, A. Varykhalov, O. Rader, M. Leandersson, T. Balasubramanian, T.-R. Chang, H.-T. Jeng, S. Basak, H. Lin, A. Bansil, N. Samarth, and M. Z. Hasan, Nat. Commun. **5**, 3841 (2014).

[80] Y. Wang and N. Gedik, Phys. Status Solidi Rapid Res. Lett. **7**, 64 (2013).

[81] P. Rakyta, B. Ujfalussy, and L. Szunyogh, New J. Phys. **17**, 123011 (2015).

[82] M. Michiardi, M. Bianchi, and P. Hofmann (private communication).

[83] L. A. Walsh, A. J. Green, R. Addou, W. Nolting, C. R. Cormier, A. T. Barton, T. R. Mowll, R. Yue, N. Lu, J. Kim, M. J. Kim, V. P. LaBella, C. A. Ventrice, S. McDonnell, W. G. Vandenberghe, R. M. Wallace, A. Diebold, and C. L. Hinkle, ACS Nano **12**, 6310 (2018).

[84] Y. L. Chen, J. G. Analytis, J.-H. Chu, Z. K. Liu, S.-K. Mo, X. L. Qi, H. J. Zhang, D. H. Lu, X. Dai, Z. Fang, S. C. Zhang, I. R. Fisher, Z. Hussain, and Z.-X. Shen, Science **325**, 178 (2009).

[85] J. Sánchez-Barriga, M. Battiato, M. Krivenkov, E. Golias, A. Varykhalov, A. Romualdi, L. V. Yashina, J. Minár, O. Kornilov, H. Ebert, K. Held, and J. Braun, Phys. Rev. B **95**, 125405 (2017).

[86] M. Neupane, S. Basak, N. Alidoust, S.-Y. Xu, C. Liu, I. Belopolski, G. Bian, J. Xiong, H. Ji, S. Jia, S.-K. Mo, M. Bissen, M. Severson, H. Lin, N. P. Ong, T. Durakiewicz, R. J. Cava, A. Bansil, and M. Z. Hasan, Phys. Rev. B **88**, 165129 (2013).
[87] K. Sumida, Y. Ishida, and A. Kimura (private communication).
[88] S. Zhu, Y. Ishida, K. Kuroda, K. Sumida, M. Ye, J. Wang, H. Pan, M. Taniguchi, S. Qiao, S. Shin, and A. Kimura, Sci. Rep. **5**, 13213 (2015).
[89] J. Sánchez-Barriga, E. Golias, A. Varykhalov, J. Braun, L. V. Yashina, R. Schumann, J. Minár, H. Ebert, O. Kornilov, and O. Rader, Phys. Rev. B **93**, 155426 (2016).
[90] M. Eschbach and L. Plucinski (private communication).
[91] J. A. Rehman and G. Mussler (private communication).
[92] W. Liu, X. Peng, C. Tang, L. Sun, K. Zhang, and J. Zhong, Phys. Rev. B **84**, 245105 (2011).