# Spin-orbit coupling in quasiparticle studies of topological insulators

Irene Aguilera, Christoph Friedrich, and Stefan Blügel
*Peter Grünberg Institute and Institute for Advanced Simulation, Forschungszentrum Jülich and JARA, D-52425 Jülich, Germany*
(Received 30 July 2013; published 25 October 2013)

We present one-shot $GW$ calculations of the bulk electronic structure of the topological insulators $Bi_2Se_3$ and $Bi_2Te_3$ within the all-electron full-potential linearized augmented-plane-wave formalism. We compare three different ways of treating the spin-orbit interaction in calculating the quasiparticle energies: (i) The spin-orbit coupling (SOC) is already incorporated in the noninteracting system that serves as starting point for the quasiparticle correction. (ii) The SOC is added in a second-variation approach only after the quasiparticle calculation has been performed in the absence of SOC. We found that the approximate treatment (ii) yields most quasiparticle bands with reasonable accuracy but does fail in the important band-gap region, where the SOC gives rise to a band inversion relevant for the topological properties of these materials. For example, $Bi_2Se_3$ is just on the brink of becoming a trivial semiconductor within this approximate approach, while it maintains its topological properties in the case of the consistent treatment (i). Finally, we consider another approach (iii), in which the SOC is included in the Green function $G$ as in (i), but neglected in the calculation of the screened Coulomb potential $W$. This approach gives results in very good agreement with the full treatment (i), but with a smaller numerical effort. We conclude that, in the high-symmetry directions studied, bulk $Bi_2Se_3$ is a direct-gap and $Bi_2Te_3$ an indirect-gap semiconductor with band gaps of 0.20 and 0.19 eV, respectively.

DOI: 10.1103/PhysRevB.88.165136 PACS number(s): 71.10.−w, 71.15.Mb, 71.20.−b, 71.70.Ej

## I. INTRODUCTION

Recently, many-body calculations within the $GW$ approximation[1] have started to emerge in the theoretical study of topological insulators.[2–10] This field was so far dominated by calculations based on model Hamiltonians or parameter-dependent tight-binding descriptions,[11–14] and density functional theory (DFT) employing either the local-density (LDA) or generalized gradient (GGA) approximation.[15–24]

Topological insulators are a kind of material that show new electronic phenomena and great potential for applications in spintronics, quantum computing, thermoelectrics, and green IT due to the possibility of dissipationless currents.[25–27] Strong spin-orbit interactions cause an inverted band gap in these materials, which is of the order of the spin-orbit strength (up to a few hundreds of meV). Such small band gaps require a reliable description of the electronic structure in order to get accurate results. The characteristic inverted band structure gives rise to nontrivial edge or surface states that, by symmetry considerations, are required to be metallic. For potential applications, it is desirable that these metallic surface states form a Dirac cone separated from the bulk bands, and that the Fermi level lies at the Dirac point. This situation is known as the topological transport regime.[28] The correct description of the band gaps of the bulk semiconductors thus becomes of great importance, as they have to host the surface states. An accurate picture of the bulk band gaps allows us to determine the possible overlap between the bulk band structure and the surface Dirac cones and can therefore help in predicting whether topological transport can be realized in the material of interest.

In practice, DFT is restricted to the calculation of ground-state properties. Properties that rely on excited states are therefore often inadequately described within the Kohn-Sham formalism of DFT. In fact, previous studies[2,3,7,10] have shown that the nature of the band gap (indirect or direct) of $Bi_2Se_3$ and $Bi_2Te_3$, its magnitude, and the dispersions of the bands involved in the band inversion are not described correctly. It has also been demonstrated recently[5] that DFT can lead to the wrong prediction of some trivial insulators as topological insulators. Hence, the recent quest to perform beyond-DFT studies (e.g., with the $GW$ approximation) to achieve a more reliable *ab initio* description of the electronic structure of these materials.

$Bi_2Se_3$ and $Bi_2Te_3$ are among the most widely studied topological insulators due to the simplicity of their surface states consisting of a single Dirac cone at the $\Gamma$ point,[19] which is observed in angle-resolved photoemission spectroscopy measurements.[28–31] Both compounds crystallize in a rhombohedral structure with the space group $R\bar{3}m$ and five atoms in the unit cell. They consist of covalently bonded quintuple layers in the sequence $Se_1$–Bi–$Se_2$–Bi–$Se_1$ (or, equivalently, with Te). The layers are bound to the adjacent ones through weak van der Waals forces. Their experimental inverted band gaps, between approximately 100 and 300 meV,[7,15,32–37] make them good candidates for experimental studies of topological effects and for room-temperature applications. In addition, these materials and some of their alloys are nowadays commonly used in thermoelectric refrigeration and power generation.[15,38]

The coupling of the orbital motion with the electronic spin is responsible for the peculiar topological properties of the electronic band dispersion in topological insulators. In particular, it gives rise to a band inversion at the zone center in $Bi_2Se_3$ and $Bi_2Te_3$. The band inversion can be imagined as two parabolic bands, one with a minimum at $-\Delta/2$ and an inverted one with a maximum at $\Delta/2$. The two bands—formed in the present materials by the Bi $6p$ and Se (Te) $5p$ states, respectively—hybridize so that a direct crossing is avoided and an energy gap opens. Depending on the strength of the hybridization relative to $\Delta$, an M-shaped valence and a W-shaped conduction band may or may not form. The presence of this characteristic *camelback* shape in the band dispersion is a clear indication for inverted bands. But the reverse is not

 

true: Band inversion may be present even without a *camelback* shape.

As the electron moves through the crystal field, it sees, in its reference frame, a changing electric field that, according to Maxwell's equations, gives rise to an intrinsic magnetic field, with which the electron spin magnetic moment interacts. The spin-orbit coupling (SOC) term in the Hamiltonian can be interpreted as the magnetic energy of this interaction. In this sense, the SOC is a one-electron phenomenon that is readily treated within an independent-particle approach, such as the Kohn-Sham formalism of DFT. A treatment within many-body perturbation theory seems straightforward: The SOC can already be taken into account in the independent-particle reference system that is used as the starting point for the quasiparticle renormalization. However, this leads to very expensive calculations because the SOC gives rise to spin-off-diagonal blocks in the Hamiltonian, the Green function $G$, and the self-energy $\Sigma$. Therefore, the first $GW$ calculations with SOC present in the literature[2,3] for $Bi_2Se_3$ and $Bi_2Te_3$ were carried out in a second-variation approach: After performing a one-shot $GW$ calculation without SOC, a Hamiltonian is constructed with the quasiparticle energies on the diagonal and all matrix elements of the SOC operator obtained in a conventional DFT calculation; diagonalization then yields new quasiparticle energies. However, this mixing of a single-particle operator (the SOC) and excitation energies of the many-body system (the quasiparticle energies) is inconsistent and must be considered an approximation. In fact, Sakuma *et al.*[4] have recently shown that quasiparticle calculations with a full treatment of SOC, employing the four-component spinor wave functions, improve the spin-orbit splittings of Hg chalcogenides with respect to the values given by LDA. This is because, by incorporating the SOC already in the reference system, the $GW$ self-energy acquires terms that couple the two spin channels, which enables a many-body renormalization of the spin-orbit split bands, an effect that is absent otherwise.

The second-variation approach used in Refs. 2 and 3 is based on a $GW$ calculation in the absence of SOC, thus making calculations much more computationally efficient than those with a full treatment of SOC (which are about 10 times more time consuming). This is very promising, as the variety and the size of the systems that can be studied with this simpler approach is larger. However, the validity of this approximation for topological insulators, and in particular for the family of $Bi_2Se_3$ and $Bi_2Te_3$, remains to be assessed.

In this work we compare quasiparticle results calculated with the full treatment of SOC with those obtained from the second-variation approach. While the two approaches yield for the most part very similar quasiparticle band dispersions, we show that the latter does not reproduce correctly the dispersion of the important bands involved in the band inversion at the zone center. We have also performed calculations including SOC in a full way, but only in the calculation of the Green function $G$, whereas the screened interaction $W$ (whose calculation is computationally demanding) was calculated without spin-orbit interactions [i.e., it is approximated by the scalar-relativistic (SR) approach]. As we will see, this approach gives a reliable and good approximation at a smaller computational expense. In addition, it allows us to investigate the relative importance of changes induced by the SOC in $G$ and $W$, whose product gives the $GW$ self-energy. We find that the changes in $W$ affect the quasiparticle energies only a little, whereas a much bigger effect is seen from SOC-induced changes in the Green function.

## II. METHODS AND COMPUTATIONAL DETAILS

The DFT and $GW$ calculations are carried out within the all-electron full-potential linearized augmented-plane-wave (FLAPW) formalism as implemented in the DFT code FLEUR[39] and the $GW$ code SPEX.[40] For the self-consistent determination of the electronic density within DFT we employ the LDA exchange-correlation functional.[41] The use of the GGA functional has been shown to yield similar results.[7,10] We use an angular momentum cutoff of $l_{\max} = 10$ in the muffin-tin spheres, a plane-wave cutoff of 4.5 bohrs$^{-1}$ in the interstitial region, and an $8\times8\times8$ **k**-point grid to sample the Brillouin zone (BZ). Semicore $d$ states of Se, Te, and Bi are treated as valence states by the use of local orbitals. We use the experimental lattice structures of Refs. 42 and 43 for $Bi_2Se_3$ and $Bi_2Te_3$, respectively.

The mixed product basis[40,44] used in the $GW$ calculations is constructed with an angular momentum cutoff of $L_{\max} = 5$ and a plane-wave cutoff of 2.9 bohrs$^{-1}$. We have thoroughly converged the excitation energies with the number of unoccupied states and have found that with 500 bands (corresponding to states about 80 eV above the Fermi energy) we obtain band gaps that are converged to within less than 10 meV. We use a $4\times4\times4$ **k**-point sampling of the BZ for the $GW$ calculations. Two additional local orbitals per angular momentum up to $l = 3$ are included for each atom to describe high-lying states accurately and to avoid linearization errors.[45,46] We do not use any interpolation technique to represent the $GW$ band structures shown in this work. Instead, the self-energy is evaluated for each represented **k** point explicitly. The Fermi level is placed in the middle of the band gap in each case.

In the present work we compare three different approaches to include spin-orbit interactions in $GW$ calculations. We label them by "$GW$ + SOC," "$G^{\mathrm{SOC}}W^{\mathrm{SOC}}$," and "$G^{\mathrm{SOC}}W$" in the following.

The first one ($GW$ + SOC) incorporates the SOC using second variation[47] after the quasiparticle correction to the Kohn-Sham energies has been performed, a technique that is often used in LDA calculations to include SOC (referred to as "LDA + SOC"). In this approach, the relativistic Hamiltonian $H$ is represented in the scalar-relativistic Kohn-Sham eigenstates. So, it consists of the diagonal scalar-relativistic Hamiltonian $H_{\mathrm{SR}}$ with the scalar-relativistic LDA eigenvalues on the diagonal and the full SOC matrix $H_{\mathrm{SO}}$. The diagonalization of $H = H_{\mathrm{SR}} + H_{\mathrm{SO}}$ then yields the relativistic LDA eigenstates, i.e., including SOC. When applied to $GW$ one replaces the scalar-relativistic LDA eigenvalues by the quasiparticle energies obtained in the absence of SOC, which is the approach used in Refs. 2 and 3 for $Bi_2Se_3$ and $Bi_2Te_3$.[48] While the application of the second-variation procedure in the conventional DFT self-consistency loop only corresponds to a change of basis on the LDA level and, thus, does not constitute an approximation (apart from a possible reduction of the basis set), it must be considered an approximation when applied to the $GW$ quasiparticle correction: (1) The quasiparticle

amplitudes cannot be interpreted as single-particle states. Thus, using them to set up a SOC Hamiltonian with the single-particle SOC operator is inconsistent. (2) The self-energy operator depends on the reference system through the Green function and the screened interaction. Omitting the SOC in the noninteracting reference system amounts to neglecting spin-orbit induced screening effects. (3) Since the SOC is only considered after the quasiparticle correction, a many-body renormalization of spin-orbit split bands is not possible. (4) The quasiparticle equation [Eq. (5)] is nonlinear in the quasiparticle energy $E_{\mathbf{k}\mu}$. The *a posteriori* inclusion of SOC changes $E_{\mathbf{k}\mu}$, thus *detuning*, in a sense, the nonlinear quasiparticle equation solved previously without SOC.

We note that the $GW+\mathrm{SOC}$ approximation, in the way described here, is only applicable in cases where the Kohn-Sham states are reasonable approximations for the quasiparticle amplitudes (without SOC). This is correct for the two systems considered in this work, but not for another topological insulator of the same family, $Sb_2Te_3$, where the LDA calculation in the absence of SOC yields a "negative" band gap, while the $GW$ gap (also without SOC) is positive. In Ref. 9 we showed that off-diagonal elements of the self-energy $\Sigma$ become crucial in this case to describe the hybridizations close to the band-gap region correctly. But when $GW$ calculations are performed in such a way (including the off-diagonal elements of $\Sigma$), the resulting quasiparticle amplitudes deviate in general from the Kohn-Sham states—and they do so strongly in $Sb_2Te_3$—impeding a straightforward application of the $GW+\mathrm{SOC}$ method.

The second approach used ($G^{\mathrm{SOC}}W^{\mathrm{SOC}}$) is the one proposed in Ref. 4, in which the SOC is fully taken into account in a one-shot $GW$ formalism already at the level of the noninteracting reference system, i.e., it is naturally contained in the single-particle Green function $G$, the screened interaction $W$, and thus also in the self-energy $\Sigma$. In this way, the formal inconsistency mentioned above is avoided, and spin-orbit screening effects as well as renormalization effects in spin-orbit split bands are taken into account. This approach is consistent with the original formulation by Hedin[1] and also with a recent generalized formulation by Aryasetiawan and Biermann.[49,50] Using an explicit spin-dependent notation, the $GW$ self-energy is written as

$$\Sigma_{\alpha\beta}(\mathbf{r},\mathbf{r}';\omega)=\frac{i}{2\pi}\int G_{\alpha\beta}(\mathbf{r},\mathbf{r}';\omega+\omega')\times W(\mathbf{r},\mathbf{r}';\omega')e^{i\eta\omega'}d\omega', \tag{1}$$

where $\alpha$ and $\beta$ are the spin indices and $\eta$ is a positive infinitesimal. The Green function reads

$$G_{\alpha\beta}(\mathbf{r},\mathbf{r}';\omega)=\sum_{\mathbf{k},\mu}\frac{\phi^{\alpha}_{\mathbf{k}\mu}(\mathbf{r})\phi^{\beta *}_{\mathbf{k}\mu}(\mathbf{r}')}{\omega-\epsilon_{\mathbf{k}\mu}+i\eta\,\mathrm{sgn}(\epsilon_{\mathbf{k}\mu}-\epsilon_{\mathrm{F}})}, \tag{2}$$

where $\epsilon_{\mathrm{F}}$ is the Fermi energy, $\mathbf{k}$ is a Bloch vector, $\mu$ is the band index, and $\epsilon_{\mathbf{k}\mu}$ are the eigenvalues corresponding to the spinor Bloch functions

$$\phi_{\mathbf{k}\mu}(\mathbf{r},s)=\begin{cases}\phi^{\uparrow}_{\mathbf{k}\mu}(\mathbf{r}) & \text{for } s=\uparrow\\ \phi^{\downarrow}_{\mathbf{k}\mu}(\mathbf{r}) & \text{for } s=\downarrow\end{cases} \tag{3}$$

with the spin variable $s$. In the muffin-tin spheres, i.e., close to the atomic nuclei, the fully relativistic Dirac equation is employed so that each of the two functions separate further into a large and a small component and Eq. (3) becomes a four-component spinor wave function. The screened Coulomb interaction $W$ is calculated from the polarization function $P$ via the equation $W=v+vPW$, where $v$ is the bare Coulomb interaction and

$$P(\mathbf{r},\mathbf{r}';\omega)=\frac{-i}{2\pi}\sum_{\alpha,\beta}\int G_{\alpha\beta}(\mathbf{r},\mathbf{r}';\omega+\omega')\times G_{\beta\alpha}(\mathbf{r}',\mathbf{r};\omega')d\omega'. \tag{4}$$

The quasiparticle energies, finally, are determined from the quasiparticle equation

$$\sum_{\beta}h_{\alpha\beta}(\mathbf{r})\psi^{\beta}_{\mathbf{k}\mu}(\mathbf{r})+\sum_{\beta}\int\Sigma_{\alpha\beta}(\mathbf{r},\mathbf{r}';E_{\mathbf{k}\mu})\psi^{\beta}_{\mathbf{k}\mu}(\mathbf{r}')d^3r'=E_{\mathbf{k}\mu}\psi^{\alpha}_{\mathbf{k}\mu}(\mathbf{r}), \tag{5}$$

where $\psi^{\alpha}_{\mathbf{k}\mu}(\mathbf{r})$ is the $\alpha$ component of the quasiparticle spinor, and the single-particle operator $h_{\alpha\beta}(\mathbf{r})$ contains the relativistic kinetic-energy operator including the spin-orbit coupling term and the electrostatic potential created by the atomic nuclei and the electronic charge distribution. In the $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ and $G^{\mathrm{SOC}}W$ calculations, we solve Eq. (5) in the basis of the LDA single-particle states explicitly, while a perturbative solution is used for the $GW+\mathrm{SOC}$ approach, neglecting the off-diagonal elements of the self-energy.[9]

The spinor wave functions give rise to spin-off-diagonal elements in the Green function and the self-energy, and the coupling of the spin channels is mediated by the SOC. This allows for spin-flip processes and many-body renormalization of the spin-orbit interaction itself.

Equation (4) shows that $W$ does not depend explicitly on the spin, but it is affected by the SOC implicitly through the Green function. This leads to the question to which extent this implicit dependence affects the self-energy. Therefore, we consider a third approach in this work ($G^{\mathrm{SOC}}W$), in which the SOC is fully taken into account in the Green function [Eq. (2)], while the screened interaction is taken from a calculation without SOC. This approximation is a compromise between $GW+\mathrm{SOC}$ and $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$. It is computationally more expensive than the former but considerably cheaper than the latter, as the calculation of $W$ with full SOC is a particularly demanding step. A comparison with results from $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ will demonstrate how the SOC affects the screening. Furthermore, since the same $W$ is used, a comparison with results from $GW+\mathrm{SOC}$ allows us to assess the error introduced by adding the SOC *a posteriori* instead of already in the noninteracting reference system as in $G^{\mathrm{SOC}}W$. The $G^{\mathrm{SOC}}W$ approximation was found in Ref. 4 to be a very good approximation for mercury chalcogenides, which also present a band inversion due to spin-orbit interactions.

In this work we compare the three approaches to assess their validity for $Bi_2Se_3$ and $Bi_2Te_3$. In particular, we compare the $GW+\mathrm{SOC}$ results with those in Refs. 2 and 3. It should be noted that our calculations differ from the latter ones in two important aspects: First, we do not use the plane-wave pseudopotential method but an all-electron approach, which

treats core, valence, and conduction electrons on an equal footing. Second, instead of using a plasmon-pole model we take into account the full frequency dependence of the screening and evaluate the frequency convolution in $GW$ [Eq. (1)] explicitly with a contour-integration technique;[51,52] 18 frequencies are used to sample $W(\mathbf{r},\mathbf{r}';\omega)$ on the imaginary frequency axis, i.e., for $\omega = [0, i\infty[$ ($W$ is symmetric around $\omega = 0$), and an equidistant mesh with step size 0.01 Ha is employed for real $\omega$. The self-energy $\Sigma_{\alpha\beta}(\mathbf{r},\mathbf{r}';\omega)$ is evaluated on an equidistant mesh $\omega = [-0.2, 0.2]$ Ha relative to the Fermi energy with step size 0.01 Ha. For the solution of the quasiparticle equation [Eq. (5)] we employ a spline interpolation for all self-energy matrix elements. We note that no finite value is used for the infinitesimal $\eta$ in Eq. (2). Instead, we take the exact limit $\eta \to 0^+$, which leads to a mathematical $\delta$ function in the imaginary part $\mathrm{Im}P$ of Eq. (4), making the application of the tetrahedron method[53] straightforward. The real part $\mathrm{Re}P$ is then obtained by a Hilbert transformation.

## III. RESULTS

Figure 1 shows the LDA and $GW$ band structures of $Bi_2Se_3$ and $Bi_2Te_3$, as calculated from the standard LDA + SOC approach as well as the different flavors of $GW$ schemes that incorporate SOC discussed above. Along the high-symmetry directions, except around the Γ point, we can see that the three approaches to SOC give very similar results. We also see that the positions of the conduction-band minimum (CBM) and the valence-band maximum (VBM) within $GW$ change with respect to LDA. As has already been discussed in previous works of these materials,[3,7] we observe a flattening of the characteristic *camelback* shape of the valence band when $GW$ is applied. This might result in band gaps that change from indirect to direct, as in the case of $Bi_2Se_3$. The VBM of $Bi_2Se_3$ [Fig. 1(a)] changes from its LDA + SOC position in the Z-F direction to the Γ point, resulting in a direct $GW$ band gap, in agreement with Refs. 3 and 7. This happens in all three $GW$ approaches, but with $GW$ + SOC showing a significantly different dispersion at the Γ point than the other approaches (see below). For $Bi_2Te_3$ [Fig. 1(b)] we find the $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ (and $G^{\mathrm{SOC}}W$) CBM and VBM in the same directions as in the LDA + SOC calculation: Γ-Z and Z-F, respectively. This is in contrast to the results found by Yazyev *et al.*,[3] in which the CBM appears clearly at the Γ point, a result that is reproduced by our $GW$ + SOC approach.

Table I shows the corresponding LDA and $GW$ band gaps. We also indicate in square brackets whether the band gaps are direct or indirect and we compare them (in parentheses) with previous $GW$ results in the literature. The values of the $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ band gaps are in good agreement with experiments and previous results. The $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ band gap of $Bi_2Se_3$ is in better agreement with experiments than those from the other approaches studied in this work. As shown by Nechaev *et al.*,[7] different structural parameters used in the calculations can give rise to differences in the ($G^{\mathrm{SOC}}W^{\mathrm{SOC}}$) band gaps of up to 0.15 eV in $Bi_2Se_3$. The experimental structure used in the present work is very close to the LDA-relaxed structure of Ref. 7, where a band-gap value (0.19 eV) very close to ours was reported.

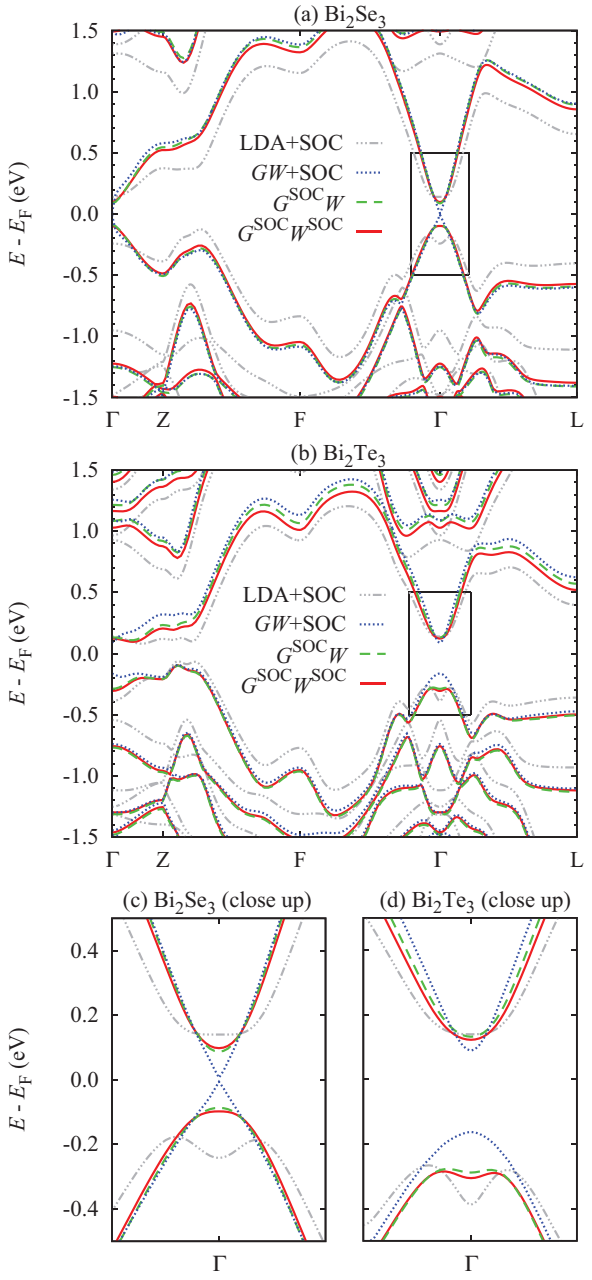


FIG. 1. (Color online) LDA + SOC, $GW$ + SOC, $G^{\mathrm{SOC}}W$, and $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ band structures of (a) $Bi_2Se_3$ and (b) $Bi_2Te_3$. The black rectangles show the range in which the magnifications in (c) and (d) are represented.

For $Bi_2Te_3$, the $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ indirect band gap is in good agreement with the experimental values. Our values are slightly bigger than the experimental ones, which might point to a location of the real VBM and CBM away from the high-symmetry **k** points examined here. The exact position of the VBM and CBM is still under debate. Kioupakis *et al.*[2] employed a Wannier interpolation to examine the $GW$ band structure in a two-dimensional sheet of the Brillouin zone. However, the authors used the $GW$ + SOC approach, which, as we will see, can lead to wrong dispersions of the bands. It has come to our attention that Nechaev and Chulkov[10] have recently performed similar calculations with the $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$

TABLE I. LDA + SOC, $GW$ + SOC, $G^{\mathrm{SOC}}W$, and $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ band gaps of $Bi_2Se_3$ and $Bi_2Te_3$ compared to experimental values (in eV). [D] and [I] indicate a direct and indirect band gap, respectively. In parentheses, we show the $GW$ values of previous works.

| | $Bi_2Se_3$ | $Bi_2Te_3$ |
|---|---|---|
| LDA + SOC | 0.27[I] | 0.14[I] |
| $GW$ + SOC | 0.01[D] (0.30[a]) | 0.18[I] (0.17[a,b]) |
| $G^{\mathrm{SOC}}W$ | 0.17[D] | 0.21[I] |
| $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ | 0.20[D] (0.19[c]; 0.34[d]) | 0.19[I] |
| Expt. | 0.2–0.33[D][e] | 0.13–0.17[I][f] |

[a]Reference 3, with experimental structure from Ref. 43.
[b]Reference 2, with experimental structure from Ref. 42.
[c]Reference 7, with LDA-relaxed structure.
[d]Reference 7, with experimental structure from Ref. 43.
[e]See, e.g., Refs. 7,15,32 and 33.
[f]References 34–37.

approach and without a Wannier-interpolation scheme and found different positions of VBM and CBM than Ref. 2. Based on their results, they infer an *optimal* band-gap value of 0.16 eV.

Figure 1 shows good agreement between the $G^{\mathrm{SOC}}W$ band structures (green dashed lines) and the $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ ones (red solid lines). They are for the most part situated between the $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ and $GW$ + SOC bands, in line with the formal theoretical level as well as the computational expense of the methods relative to each other. More importantly, the band-inversion region is described correctly within $G^{\mathrm{SOC}}W$ in contrast to $GW$ + SOC. The differences between the $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ and $G^{\mathrm{SOC}}W$ bands are not bigger than 40 meV, being much smaller for most of the states. Furthermore, Table I shows that the $G^{\mathrm{SOC}}W$ band gaps are in relatively close agreement with the $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ ones. This indicates that the inclusion of SOC induces only small changes in the screened interaction, while the fundamental changes appear in the Green function, confirming an earlier result by Sakuma *et al.*[4] Considering that the construction of $W$ with a full account of SOC is a very time-consuming step, this is an important finding for future $GW$ calculations of these materials.

The $GW$ + SOC calculations, shown as blue dotted lines in Fig. 1, were performed in the same way as in Refs. 2 and 3. When comparing these results to those of $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$, there are two main issues that have to be discussed.

First, the $GW$ + SOC band structure is, for the most part, in close agreement with the $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ one, but around the Γ point of $Bi_2Se_3$, where the band inversion takes place, the $GW$ + SOC band structure shows a peculiar "X" shape that looks like two metallic bands that cross right at the Fermi energy, in blatant disagreement with the $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ (and $G^{\mathrm{SOC}}W$) band structure. (The X shape should not be confused with the Dirac cone formed by the surface states of topological insulators since here we discuss the bulk band structure.) Having a closer look [Fig. 1(c)] at what seems to be a crossing, we find that it is actually composed of a valence and a conduction band, with a tiny—but still inverted—band gap of 10 meV (see Table I). We will discuss this behavior later, as it was not found by the previous $GW$ + SOC calculations.[3] In the case of $Bi_2Te_3$ [Fig. 1(d)], on the other hand, we do not see such a peculiar band dispersion, but it still looks qualitatively different from the $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ one. The highest valence band of $GW$ + SOC shows a local maximum at the Γ point, completely removing the pronounced *camelback* shape found in LDA + SOC. An orbital analysis shows, however, that the band inversion persists nevertheless. The $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ bands are in between, with a relatively flat but still identifiable *camelback* shape. Since the VBM is at a point in the Brillouin zone, which is described similarly by all $GW$ approaches [see Z-F direction in Fig. 1(b)], we find indirect band gaps with rather similar gap values (see Table I) in contrast to $Bi_2Se_3$.

Second, in addition to that of the highest valence band, the dispersion of the lowest conduction band of $Bi_2Te_3$ in the Γ-Z direction differs significantly between $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ and $GW$ + SOC. The highest occupied state at the Γ point is, as we have just seen, a local maximum in $GW$ + SOC instead of a local minimum. Inversely, the lowest unoccupied state is a minimum (actually the global one along the examined directions) instead of a local maximum. The CBM of $Bi_2Te_3$ is, thus, located at the Γ point within $GW$ + SOC, unlike what we find in the reference $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ approach, which places the CBM at a **k** point along the Γ-Z direction.

To clarify the origin of the X shape found at the Γ point of $Bi_2Se_3$ and, in particular, to explain why it was not found in Ref. 3, we present in Fig. 2 the band dispersions calculated with the second-variation approach with the following energies on the diagonal: (1) Kohn-Sham eigenvalues (LDA + SOC), (2) $GW$ quasiparticle energies ($GW$ + SOC), and (3) Kohn-Sham eigenvalues where the unoccupied states have been shifted rigidly upwards by 0.1, 0.2, 0.3,

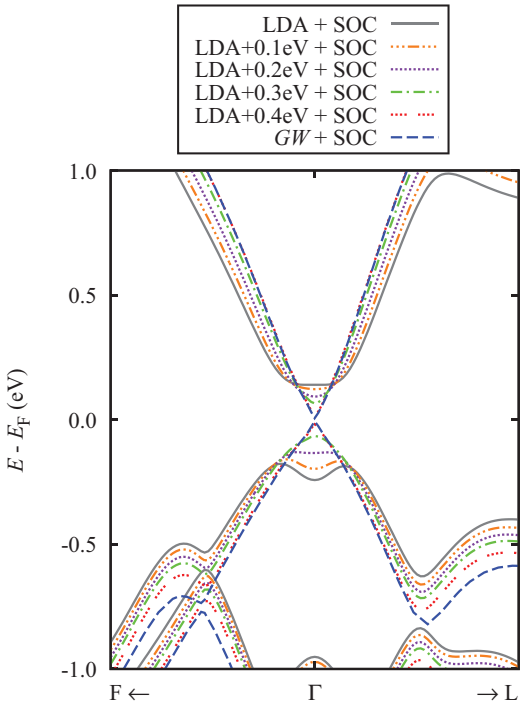


FIG. 2. (Color online) Band structure of $Bi_2Se_3$ around the Γ point calculated with the second-variation approach to SOC using as a starting point the following calculations without SOC: LDA, LDA with artificial openings of the band gap (labeled, e.g., LDA + 0.1 eV), and $GW$.

and 0.4 eV (e.g., LDA + 0.1 eV + SOC), thereby making a connection between LDA + SOC and $GW$ + SOC. As already mentioned above, LDA + SOC (gray solid curve) produces the well-known *camelback* shape, characteristic of band inversion, and $GW$ + SOC (blue dashed line) leads to the X-shaped bands described above. The band dispersions in between are realized by the artificial opening of the LDA band gap, where LDA + 0.4 eV matches approximately the $GW$ gap. We note that the corresponding bands are nearly identical only at the Γ point, while they are clearly different elsewhere, which shows that the $GW$ correction, even in the absence of SOC, cannot be described merely by a rigid upward shift of the unoccupied states.

Our LDA and $GW$ band gaps without SOC for $Bi_2Se_3$ are direct at the Γ point and have values of 0.08 and 0.41 eV, respectively. This $GW$ value is around 0.2 eV larger than that found in Ref. 3 (0.212 eV), certainly not a big difference in absolute terms, in particular for such different $GW$ calculations. However, the relative difference is rather large due to the small absolute band-gap value of $Bi_2Se_3$, which affects strongly the hybridizations arising from SOC and leading to the band inversion. There are several possible reasons for this difference between Ref. 3 and the present work: (1) the difference in the LDA band gaps (0.08 vs 0.151 eV), (2) different numbers of states used in the $GW$ calculations (500 vs 270), (3) different lattice structures (we found a $GW$ band gap of 0.35 eV with the structure of Ref. 43, which is the one used in Refs. 3 and 54, see also Ref. 7), (4) all-electron vs plane-wave basis, (5) full potential vs pseudopotential approximation, and (6) full frequency dependence of $W$ and contour integration vs plasmon-pole approximation. The effect that each of these points has on the quasiparticle correction is difficult to quantify individually without a thorough comparison of the two computer codes.

Figure 2 illustrates the origin of the X shape, found in this work and not in Ref. 3. As the inclusion of SOC produces the band inversion, the larger the band gap of the system without SOC, the smaller the inverted band gap found after the inclusion of SOC, until, around LDA + 0.5 eV in this case, the SOC is not strong enough to produce a band inversion anymore, and the material becomes a trivial semiconductor. Our $GW$ + SOC result is, thus, just at the border of band inversion, which manifests itself in the peculiar X-shaped bands.[55] In fact, using $k \cdot p$ perturbation theory in the form presented in Refs. 3 and 19 such an X shape can be found for suitably chosen parameters.

## IV. SUMMARY AND CONCLUSIONS

We have performed $GW$ calculations for the topological insulators $Bi_2Se_3$ and $Bi_2Te_3$ including spin-orbit coupling within three different approaches: (i) A full treatment of SOC in both $G$ and $W$ ($G^{\mathrm{SOC}}W^{\mathrm{SOC}}$), used here as a reference, and two approximations. (ii) A second-variation approach ($GW$ + SOC) that enables calculations with a considerably lower computational cost compared to (i) but fails in describing the important band-inversion region, which determines the topological properties of the materials. And (iii) a full treatment of SOC, like in (i), but only in the calculation of the Green function $G$ and not in $W$ ($G^{\mathrm{SOC}}W$). The latter has proven to be a useful approximation for the two materials, as it predicts results in close agreement to $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$, avoiding the time-consuming step of calculating the screened interaction $W$ including SOC.

The second-variation $GW$ + SOC approach has been found to yield an electronic band structure that is overall in reasonable agreement with the $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ one, but the important dispersion of the bands involved in the band inversion of $Bi_2Se_3$ and $Bi_2Te_3$ are not obtained correctly when compared to $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$. In particular, $Bi_2Se_3$ is on the brink of becoming a trivial semiconductor in $GW$ + SOC. We have seen that the $GW$ + SOC approach gives qualitatively different locations of the VBM and the CBM in these materials. For example, the CBM in the high-symmetry directions studied is at the Γ point in $Bi_2Te_3$, while it is at a **k** point away from the Γ point within $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$. A discussion extended to the mirror plane can be found in Ref. 10.

We have found that differences as small as 0.2 eV in the $GW$ starting point can lead to substantial qualitative changes in the $GW$ + SOC results (Fig. 2). This becomes especially delicate for compounds like $Bi_2Se_3$ for which the $GW$ band gap depends strongly on the structural parameters.

The qualitative changes between the LDA and $GW$ band structures in the vicinity of the Γ point, where the band inversion takes place, invites us to reinvestigate the surface states of these topological insulators within a $GW$ framework. At present, studying surfaces of these materials within the $GW$ approach with a full account of SOC (labeled here by $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$) is computationally too demanding. On the other hand, an approach on a higher theoretical level than LDA might help to shed light on some aspects of these compounds (a first attempt in that direction can be found in Ref. 3). Therefore, it is important to study ways of simplifications that can be applied in order to make $GW$ calculations of surface states feasible without a loss of accuracy. The analysis presented here can be useful in this respect.

## ACKNOWLEDGMENTS

We thank I. A. Nechaev for providing a copy of Ref. 10 prior to publication. We also thank G. Bihlmayer for helpful discussions and a critical reading of the manuscript. This work was supported by the Alexander von Humboldt Foundation through a postdoctoral fellowship, and by the Helmholtz Association through the Virtual Institute for Topological Insulators (VITI).

[1] L. Hedin, Phys. Rev. **139**, A796 (1965).
[2] E. Kioupakis, M. L. Tiago, and S. G. Louie, Phys. Rev. B **82**, 245203 (2010).
[3] O. V. Yazyev, E. Kioupakis, J. E. Moore, and S. G. Louie, Phys. Rev. B **85**, 161101(R) (2012).
[4] R. Sakuma, C. Friedrich, T. Miyake, S. Blügel, and F. Aryasetiawan, Phys. Rev. B **84**, 085144 (2011).
[5] J. Vidal, X. Zhang, L. Yu, J.-W. Luo, and A. Zunger, Phys. Rev. B **84**, 041109 (2011).

[6]A. Svane, N. E. Christensen, M. Cardona, A. N. Chantis, M. van Schilfgaarde, and T. Kotani, Phys. Rev. B **84**, 205205 (2011).
[7]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).
[8]Z. Zhu, Y. Cheng, and U. Schwingenschlögl, Phys. Rev. Lett. **110**, 077202 (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]C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).
[12]C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).
[13]S. Murakami, Phys. Rev. Lett. **97**, 236805 (2006).
[14]C.-X. Liu, X.-L. Qi, H.-J. Zhang, X. Dai, Z. Fang, and S.-C. Zhang, Phys. Rev. B **82**, 045122 (2010).
[15]S. K. Mishra, S. Satpathy, and O. Jepsen, J. Phys.: Condens. Matter **9**, 461 (1997).
[16]T. Thonhauser, T. J. Scheidemantel, J. O. Sofo, J. V. Badding, and G. D. Mahan, Phys. Rev. B **68**, 085201 (2003).
[17]P. Larson, Phys. Rev. B **74**, 205113 (2006).
[18]G. Wang and T. Cagin, Phys. Rev. B **76**, 075201 (2007).
[19]H. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, Nat. Phys. **5**, 438 (2009).
[20]O. V. Yazyev, J. E. Moore, and S. G. Louie, Phys. Rev. Lett. **105**, 266806 (2010).
[21]B. Y. Yavorsky, N. F. Hinsche, I. Mertig, and P. Zahn, Phys. Rev. B **84**, 165208 (2011).
[22]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).
[23]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).
[24]A. Herdt, L. Plucinski, G. Bihlmayer, G. Mussler, S. Döring, J. Krumrain, D. Grützmacher, S. Blügel, and C. M. Schneider, Phys. Rev. B **87**, 035127 (2013).
[25]S. A. Wolf, D. D. Awschalom, R. A. Buhrman, J. M. Daughton, S. von Molnar, M. L. Roukes, A. Y. Chtchelkanova, and D. M. Treger, Science **294**, 1488 (2001).
[26]L. Fu and C. L. Kane, Phys. Rev. Lett. **100**, 096407 (2008).
[27]P. J. Leek, J. M. Fink, A. Blais, R. Bianchetti, M. Göppl, J. M. Gambetta, D. I. Schuster, L. Frunzio, R. J. Schoelkopf, and A. Wallraff, Science **318**, 1889 (2007).
[28]D. Hsieh, Y. Xia, D. Qian, L. Wray, J. H. Dil, F. Meier, J. Osterwalder, L. Patthey, J. G. Checkelsky, N. P. Ong, A. V. Fedorov, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature (London) **460**, 1101 (2009).
[29]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).
[30]M. Bianchi, D. Guan, S. Bao, J. Mi, B. B. Iversen, P. D. King, and P. Hofmann, Nat. Commun. **1**, 128 (2010).
[31]Z. Liu, Y. Chen, J. Analytis, S. Mo, D. Lu, R. Moore, I. Fisher, Z. Hussain, and Z. Shen, Physica E **44**, 891 (2012).
[32]J. Black, E. M. Conwell, L. Seigle, and C. W. Spencer, J. Phys. Chem. Solids **2**, 240 (1957).
[33]E. Mooser and W. B. Pearson, Phys. Rev. **101**, 492 (1956).
[34]I. G. Austin, Proc. Phys. Soc. London **72**, 545 (1958).
[35]C. Li, A. L. Ruoff, and C. W. Spencer, J. Appl. Phys. **32**, 1733 (1961).
[36]R. Sehr and L. R. Testardi, J. Phys. Chem. Solids **23**, 1219 (1962).
[37]G. A. Thomas, D. H. Rapkine, R. B. V. Dover, L. F. Mattheiss, W. A. Sunder, L. F. Schneemeyer, and J. V. Waszczak, Phys. Rev. B **46**, 1553 (1992).
[38]A. A. Bayaz, A. Giani, A. Foucaran, F. Pascal-Delannoy, and A. Boyer, Thin Solid Films **441**, 1 (2003).
[39]www.flapw.de
[40]C. Friedrich, S. Blügel, and A. Schindlmayr, Phys. Rev. B **81**, 125102 (2010).
[41]J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[42]S. Nakajima, J. Phys. Chem. Solids **24**, 479 (1963).
[43]R. W. G. Wyckoff, *Crystal Structures 2* (John Wiley and Sons, New York, 1964).
[44]T. Kotani and M. van Schilfgaarde, Solid State Commun. **121**, 461 (2002).
[45]C. Friedrich, A. Schindlmayr, S. Blügel, and T. Kotani, Phys. Rev. B **74**, 045104 (2006).
[46]C. Friedrich, M. C. Müller, and S. Blügel, Phys. Rev. B **83**, 081101(R) (2011).
[47]C. Li, A. J. Freeman, H. J. F. Jansen, and C. L. Fu, Phys. Rev. B **42**, 5433 (1990).
[48]In the $GW$ + SOC approach, the effective Kohn-Sham potential of the reference system (used as a starting point for the quasiparticle correction) can be calculated with or without SOC. The calculations shown in this work are performed using potentials converged in the absence of SOC. We do so for a better comparison with Ref. 3, as that was the method used there.[54] We find that including the SOC in the effective potential changes the $GW$ + SOC energies only marginally. The maximal difference (30 meV) is seen for the conduction band in the Γ-Z direction.
[49]F. Aryasetiawan and S. Biermann, Phys. Rev. Lett. **100**, 116402 (2008).
[50]F. Aryasetiawan and S. Biermann, J. Phys.: Condens. Matter **21**, 064232 (2009).
[51]R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10159 (1988).
[52]F. Aryasetiawan, *Electronic Structure Calculations in Advances in Condensed Matter Science*, edited by V. I. Anisimov (Gordon and Breach, New York, 2000).
[53]J. Rath and A. J. Freeman, Phys. Rev. B **11**, 2109 (1975).
[54]O. V. Yazyev (private communication).
[55]In this sense, the X shape in Fig. 2 indeed bears a connection to the Dirac cone of the surface band structure, because the origin of both is a transition from an inverted to a noninverted band gap. Somewhere in between, the band gap must close giving rise to the X shape in Fig. 2 and to the Dirac cone on the surface, respectively.