Band inversion mechanism in topological insulators: A guideline for materials design

Zhiyong Zhu, Yingchun Cheng, and Udo Schwingenschlögl

Physical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia

(Received 22 November 2011; revised manuscript received 19 April 2012; published 1 June 2012)

Alteration of the topological order by band inversion is a key ingredient of a topologically nontrivial material. Using first-principles calculations for HgTe, PtScBi, and Bi2Se3, we argue that it is not accurate to ascribe the band inversion to the spin-orbit coupling. Instead, scalar relativistic effects and/or lattice distortions are found to be essential. Therefore, the search for topologically nontrivial materials should focus on band shifts due to these mechanisms rather than spin-orbit coupling. We propose an effective scheme to search for new topological insulators.

DOI: 10.1103/PhysRevB.85.235401 PACS number(s): 73.43.-f, 73.20.—r, 85.75.—d

I. INTRODUCTION

At the boundaries of topologically nontrivial states of matter charge can be transported without energy dissipation.1,2 The search for new such materials is driven not only by a fundamental scientific interest but also by applications in semiconductor spintronics. While the integer quantum Hall effect is relevant only at low temperature and high magnetic fields, a time-reversal invariant topological insulator can exist in the absence of magnetic fields and has the potential for room temperature application.4–7 The two-dimensional topological insulator, also called a quantum spin Hall insulator, was first predicted theoretically and then confirmed experimentally in HgTe/CdTe quantum wells.4,5 Over 50 three-dimensional topologically nontrivial materials have been predicted to date6–14 and some of them have been confirmed experimentally.15–21

A band inversion is needed to change the topological order.6 According to the type of band inversion, three-dimensional topologically nontrivial materials are classified as follows: (1) HgTe-like materials, including HgTe,22 α-Sn,6 several half-Heusler semiconductors,8–10 and several chalcopryte semiconductors,12,13 in which the s and p bands are inverted at the Γ point. Note that HgTe-like materials are essentially zero-gap semiconductors, instead of insulators, as long as the crystal stays cubic. They can be transformed into insulators by lattice distortions12 and the external potential in a quantum well structure.22 (2) Layered Bi2Se3-like materials, including Bi1−xSbx,6 Bi2Se3,7 Bi2Te3,7 Sb2Te3,7 and several Tl-based chalcogenides,14 where two p orbitals with opposite parity are inverted at the Γ point. In addition, three-dimensional topological insulators are predicted for cases with strong electron correlations.13,14 The band inversion is usually assumed to result from the band splitting at the Γ point caused by spin-orbit coupling (SOC)2,7,23 Consequently, the search for new topological insulators is focused on materials with strong SOC due to heavy elements.2 However, the band inversion can also be induced by lattice strain, without any SOC.24 The SOC only establishes excitation gaps around the Fermi level (EF) throughout the Brillouin zone. In fact, it has been shown that the band inversion in HgTe does not depend on the strength of the SOC.25 Therefore, further investigation of the mechanism leading to the band inversion is required.

In this paper, we use nonrelativistic (NR), scalar relativistic (SR), and fully relativistic (FR) first-principles calculations to study the band inversion mechanism in the prototypical topologically nontrivial compounds HgTe, PtScBi, and Bi2Se3 in comparison to their trivial counterparts CdTe, PtScSb, and Sb2Te3. The influence of SR effects, lattice distortion, and SOC on the energy shifts of the relevant bands is analyzed in detail. We find that SR effects and lattice distortion play a central role in band inversion, suggesting that search for topological insulators should focus on these two mechanisms, besides the band splitting due to the SOC. We propose an explicit procedure for establishing new materials and predict, as an example, a topologically nontrivial nature for strained InSb.

II. METHODOLOGY

Relativistic effects are enhanced in materials with heavy elements. Besides the SOC, SR effects of the mass velocity (−p^2/8m^3c^2) and Darwin (−h^2/4m^2c^2 · dV/dr · ∂/∂r) terms can also become prominent.26 Both SR and FR effects are treated accurately by full-potential linearized augmented plane wave calculations.27 We employ the WIEN2K software package.28 An energy threshold of ∼9 Ry to separate the valence and the core states is used in all the calculations, together with muffin-tin radii of 2.5 bohrs, a cutoff of R_m · K_{max} = 7, and l_{max} = 10. In addition, a 25 × 25 × 25 mesh is applied to the reciprocal space. We employ a combination of the modified Becke-Johnson exchange potential and the correlation potential of the local-density approximation (MBJLDA),29 as it predicts band gaps and the band order with an accuracy similar to more expensive hybrid functional and Gutzwiller calculations. The SOC is included by means of the second-variational method with the SR orbitals as the basis, where states up to 10 Ry above EF are taken into account in the basis expansion. Relativistic p_{1/2} corrections are applied to the 5p orbitals of the Hg, Pt, and Bi atoms.

III. RESULTS AND DISCUSSION

With the experimental lattice constants of HgTe (6.4603 Å) and CdTe (6.4820 Å),30 we obtain the FR band structures in Figs. 1(c) and 1(g), respectively. For both materials, the
near-\( E_g \) bands at the \( \Gamma \) point trace back to the \( s \)-type doublet \( \Gamma_6 \) state, the \( p \)-type quartet \( \Gamma_8 \) (\( J = 3/2 \)) state, and the SOC split-off doublet \( \Gamma_7 \) (\( J = 1/2 \)) state. The band structure of CdTe is similar to GaAs, with a \( \Gamma_6-\Gamma_8-\Gamma_7 \) band order from top to bottom. The energy gap \( E_g = E(\Gamma_0) - E(\Gamma_1) = 1.47 \) eV agrees almost exactly with the experimental value of 1.48 eV.\(^{31}\) This confirms the power of the MBJLDA approach to predict accurate band gaps and the band order. With respect to CdTe, the band order of HgTe is changed drastically to \( \Gamma_1-\Gamma_5-\Gamma_7 \), which agrees well with experiments.\(^{31}\) The quartet degeneracy of the \( \Gamma_8 \) state, protected by the cubic symmetry, ensures the zero-gap semiconducting nature of HgTe. The inversion between the \( s \)-type \( \Gamma_6 \) and \( p \)-type \( \Gamma_8 \) states and a negative \( E_g \) \( -0.34 \) eV, close to the experimental value of \(-0.29 \) eV (Ref. 31] signify the nontrivial band topology in HgTe.

SR band structures (i.e., SOC is excluded) of HgTe and CdTe are presented in Figs. 1(b) and 1(f), respectively. In contrast to the FR band structures, the \( s \)- and \( p \)-type bands at the \( \Gamma \) point yield singlet \( \Gamma_1 \) and triplet \( \Gamma_5 \) states, respectively. The energy gap \( E_g = E(\Gamma_0) - E(\Gamma_1) \) plays essentially the same role as in the FR case. According to Figs. 1(d) and 1(h) as well as Table I, the main consequence of the exclusion of SOC is that the downward shifted \( \Gamma_8 \) in the FR band structures, the \( s \)- and \( p \)-type bands at the \( \Gamma \) point yield singlet \( \Gamma_1 \) and triplet \( \Gamma_5 \) states. The energy gap \( E_g \) stays almost the same; only the label changes to \( \Gamma_1 \). As compared to the FR case, \( E_g \) thus increases to \(-0.08 \) and \( 1.72 \) eV in the SR calculations for HgTe and CdTe, respectively.

Most interestingly, the \( s \)- and \( p \)-type states in HgTe remain inverted at the \( \Gamma \) point when the SOC is not taken into account. Hence, \( \Gamma_8 \) is split essentially into a triplet \( \Gamma_5 \) state. The splitting of the \( p \)-like \( \Gamma_5 \) state is denoted as \( \Delta_{SOC} \) and \( E_g \) is the energy gap at the \( \Gamma \) point. All values are given in eV.

**TABLE I.** Band parameters of HgTe, CdTe, PtScBi, PtScSb, Bi\(_2\)Se\(_3\), strained Sb\(_2\)Se\(_3\) (\( s \)-Sb\(_2\)Se\(_3\)), Sb\(_2\)Se\(_3\), GaAs, InSb, and strained InSb (\( s \)-InSb). Energy shifts of the relevant bands (\( \Gamma_1 \) and \( \Gamma_3 \) for HgTe-like materials, \( \Gamma_1 \) and \( \Gamma_3 \) for Bi\(_2\)Se\(_3\)-like materials) caused by SR effects and SOC are denoted as \( \delta_{SR} \) and \( \delta_{SOC} \), respectively. Note that \( \delta_{SOC} \) of the \( \Gamma_5 \) state refers to the upward shift of the \( \Gamma_8 \) state. The splitting of the \( p \)-like \( \Gamma_5 \) state is denoted as \( \Delta_{SOC} \) and \( E_g \) is the energy gap at the \( \Gamma \) point. All values are given in eV.

<table>
<thead>
<tr>
<th></th>
<th>( \delta_{SR} )</th>
<th>( \delta_{SOC} )</th>
<th>( E_g )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \Gamma_1/\Gamma_1 )</td>
<td>( \Gamma_5/\Gamma_5 )</td>
<td>( \Delta_{SOC} )</td>
</tr>
<tr>
<td>HgTe</td>
<td>(-2.71 )</td>
<td>(-0.06 )</td>
<td>0.66</td>
</tr>
<tr>
<td>CdTe</td>
<td>(-1.27 )</td>
<td>(-0.10 )</td>
<td>0.73</td>
</tr>
<tr>
<td>PtScBi</td>
<td>(-3.90 )</td>
<td>(-0.28 )</td>
<td>0.59</td>
</tr>
<tr>
<td>PtScSb</td>
<td>(-2.08 )</td>
<td>(-0.18 )</td>
<td>0.53</td>
</tr>
<tr>
<td>Bi(_2)Se(_3)</td>
<td>(-0.57 )</td>
<td>(-1.52 )</td>
<td>-0.95</td>
</tr>
<tr>
<td>s-Sb(_2)Se(_3)</td>
<td>(-0.33 )</td>
<td>(-0.55 )</td>
<td>-0.19</td>
</tr>
<tr>
<td>Sb(_2)Se(_3)</td>
<td>(-0.32 )</td>
<td>(-0.52 )</td>
<td>-0.11</td>
</tr>
<tr>
<td>GaAs</td>
<td>(-0.69 )</td>
<td>(-0.05 )</td>
<td>0.29</td>
</tr>
<tr>
<td>InSb</td>
<td>(-1.48 )</td>
<td>(-0.11 )</td>
<td>0.64</td>
</tr>
<tr>
<td>s-InSb</td>
<td>(-1.43 )</td>
<td>(-0.09 )</td>
<td>0.61</td>
</tr>
</tbody>
</table>

FIG. 1. (Color online) Electronic band structure of HgTe obtained from (a) NR, (b) SR, and (c) FR calculations. The wave function symmetry at the \( \Gamma \) point is labeled. (d) Band energies at the \( \Gamma \) point from the NR, SR, and FR calculations. Percentages of orbital contributions at the \( \Gamma \) point from the SR calculation are given as superscripts (for values \( \geq 5\% \)). (e)–(h) Analogous to (a)–(d) but for CdTe.
consideration (SR calculation). This observation is at odds with the popular interpretation that the band inversion in HgTe with respect to CdTe is caused by the much stronger SOC in Hg than in Cd.\textsuperscript{2,23} The confusion is further promoted by comparing the SOC strength $\Delta_{\text{SOC}}$ in HgTe and CdTe. We calculate $\Delta_{\text{SOC}} = 0.66 \text{ eV}$ for HgTe, which is smaller than the CdTe value of 0.73 eV, even though the Hg atom is much heavier than the Cd atom. This insight agrees well with Ref. 25 and is ascribed to a lack of inversion symmetry in the zinc-blende structure. Furthermore, the comparable size of $\Delta_{\text{SOC}}$ in HgTe and CdTe is completely in line with the orbital characters of the $\Gamma_1$ and $\Gamma_5$ states. The replacement of Hg by Cd is not expected to influence $\Delta_{\text{SOC}}$ significantly, because $\Delta_{\text{SOC}}$ is related to the $\Gamma_5$ state with its dominating Te $p$ character [see Figs. 1(d) and 1(h)]. This fact suggests a less important role for the SOC and indicates that other mechanisms govern the band inversion in HgTe with respect to CdTe.

The secret can be unveiled by studying NR electronic band structures in which all relativistic effects are excluded. Results for HgTe and CdTe are displayed in Figs. 1(a) and 1(e), respectively. As compared to the SR case, $E_F$ is enhanced significantly to 2.89 eV for CdTe (see Table I). For HgTe it turns out to be 2.58 eV and the standard band order ($\Gamma_5$-$\Gamma_6$-$\Gamma_7$-$\Gamma_8$) is recovered. A topologically nontrivial nature is no longer maintained. Beyond, the change of $E_F$ related to SR effects amounts to 2.66 and 1.17 eV in HgTe and CdTe, respectively, while the SOC causes only changes of 0.26 and 0.25 eV. Hence, the effect of SOC is one order of magnitude smaller than SR effects, indicating that the different band topologies in HgTe and CdTe should be attributed to SR effects rather than to SOC. Our findings thus substantiate the point of view of Ref. 25.

While the change of $E_F$ by SOC can be explained mainly as a result of an upward shift of the $\Gamma_8$ state, the situation is more complicated for the SR effects. Hence, a brief discussion is required in this case to understand the band shifts.\textsuperscript{20} To begin with, the mass-velocity term gives a mass enhancement when the electron velocity approaches the velocity of light. The relevant SR electron orbitals therefore become more concentrated around the nuclei as compared to the NR orbitals, which is the well-known relativistic contraction. Consequently, electrons shift to regions where the potential is more attractive, i.e., the energies are lower. The downward band shift related to the mass-velocity term is most prominent for s electrons in heavy atoms, because s electrons with an angular momentum of zero approach the nuclei most closely and thus gain higher velocity in atoms with a larger nuclear charge. Second, the Darwin term results from a concentration of the wave function at the nuclear sites. According to the Heisenberg uncertainty principle, this strong localization leads to an increase of the electron momentum, i.e., of the kinetic energy. The downward mass-velocity shift is compensated only partially by the upward Darwin shift, giving rise to a net energy reduction.

Knowing the influence of SR effects and SOC on the band shifts, we can now understand the energy diagrams of HgTe and CdTe in Figs. 1(d) and 1(h), respectively. From NR over to FR the relativistic effects are turned on successively. In HgTe, the band shift related to SR effects is $\delta_{\text{SR}} = -2.71 \text{ eV}$ for the s-type $\Gamma_1$ state, whereas $\delta_{\text{SOC}}$ is almost zero. For the p-type $\Gamma_5$ state, we find $\delta_{\text{SR}} = -0.06 \text{ eV}$ and $\delta_{\text{SOC}} = 0.26 \text{ eV},$ which is more or less a mutual compensation. Due to the Hg $s$ and Te $p$ characters of the $\Gamma_1$ and $\Gamma_5$ states, respectively, it is clear that the strong SR effects on the Hg $s$ electrons rather than the SOC-induced band splitting related to the Te $p$ electrons gives rise to the band inversion in HgTe. Notice that this inversion affects only the $\Gamma_5$ point, while for other time-reversal invariant momenta the band is already far away from the Fermi level. Due to the small difference between the NR $E_F$ values of HgTe and CdTe, no band inversion is observed in CdTe just because of the weaker SR effects on the s electrons of the lighter Cd atom.

The same crucial role of SR effects on the s electrons for the band inversion is found in other HgTe-like materials, too. An example is the half-Heusler compound PtScBi.\textsuperscript{8} The ternary half-Heusler $CL_b$ $XYZ$ structure (with space group $F43m$) is obtained by adding a transition metal $Y$ sublattice to the zinc-blende $XZ$ sublattice.\textsuperscript{8} We have performed NR, SR, and FR calculations for PtScBi as well as its topologically trivial counterpart PtScSb. The lattice constants of PtScBi ($a = 0.66 \text{ Å}$) and PtScSb ($a = 0.63 \text{ Å}$) from Ref. 32 are used. The resulting energy diagrams are shown in Fig. 2 and relevant band parameters are listed in Table I. The NR $E_F$ values are close to identical for PtScBi ($3.22 \text{ eV}$) and PtScSb ($3.19 \text{ eV}$). In both materials, the s-type $\Gamma_1$ state has a mixed Bi/Sb s and Pt s character, while the character of the p-like $\Gamma_5$ state is predominantly Sc $d_{3z^2}$ for the $\Gamma_5$ state, $\delta_{\text{SR}}$ is more or less compensated by $\delta_{\text{SOC}}$. On the other hand, for the s-type $\Gamma_1$ state $\delta_{\text{SR}}$ is much higher in PtScBi than in PtScSb because of the heavier Bi atom. As a result, only PtScBi is topologically nontrivial with the s- and p-type bands inverted.

Different from the HgTe-like materials, Bi$_2$Se$_3$-like materials show mainly $p$-band characters around $E_F$.\textsuperscript{7} To examine the mechanism of the band inversion in this case, we choose the topological insulator Bi$_2$Se$_3$ and its trivial counterpart Sb$_2$Se$_3$. Again, we consider NR, SR, and FR configurations. Both Bi$_2$Se$_3$ and Sb$_2$Se$_3$ have a layered rhombohedral structure (with space group $R3m$), consisting of quintuple Se$_1$-Bi-Se$_2$-Bi-Se$_1$ slabs in the $z$ direction. The intraslab coupling is strong due to its covalent nature, while the inter slab coupling is mainly of van der Waals type and, consequently, much weaker. The used lattice parameters are $a = 4.138 \text{ Å}$, $c = 28.64 \text{ Å}$, $\mu = 0.399$, and $\nu = 0.206$ for Bi$_2$Se$_3$ (Ref. 33) and $a = 4.076 \text{ Å}$, $c = 29.83 \text{ Å}$, $\mu = 0.396$, and $\nu = 0.219$ for
Sb$_2$Se$_3$. The resulting band diagrams are shown in Fig. 3 and relevant band parameters are listed in Table I.

The topologically nontrivial nature of the Bi$_2$Se$_3$-like materials is reflected by an inverted band order with the odd parity state above the even parity state. The NR band structure of Bi$_2$Se$_3$ reveals an inverted $\Gamma_1$-$\Gamma_2$ order. From NR to SR, $\delta_{SR}$ of the $\Gamma_2$ state is much larger than that of the $\Gamma_1$ state, due to the Bi $s$ contributions to the former. Thus, we are back to the standard band order. Notice that $\delta_{SR}$ is much smaller for Bi$_2$Se$_3$ than for HgTe and PtScBi because of smaller Bi contributions to the $\Gamma_1$ and $\Gamma_2$ states. From SR to FR, $\delta_{SOC}$ of the $\Gamma_1$ state is large due to the Bi $p_z$ character, while $\delta_{SOC}$ of the $\Gamma_2$ state is close to zero as the SOC effects on the $p_z$ states of the Se1 and Se2 atoms cancel. Consequently, the $\Gamma_2$ and $\Gamma_4$ states become inverted again and turn Bi$_2$Se$_3$ into a topological insulator.

It has been assumed that the enhanced SOC introduced by Bi is the reason for the band inversion in Bi$_2$Se$_3$ as compared to Sb$_2$Se$_3$. However, we find that $\Gamma_1$ and $\Gamma_4$ remain inverted in hypothetical strained Sb$_2$Se$_3$ [see Fig. 3(b)] in which Bi is replaced by Sb in an otherwise unchanged Bi$_2$Se$_3$ structure. This indicates that the SOC is not the driving force of the band inversion in Bi$_2$Se$_3$. When the strain is released by a FR lattice optimization, a distinct elongation of the $c$ lattice constant is observed, while the $a$ lattice constant shrinks. For the optimized structure, Sb$_2$Se$_3$ is predicted to be a trivial insulator with the standard band order [see Fig. 3(c)]. We observe in each case (NR, SR, and FR) a rigid energy shift of the $\Gamma_1$ ($\Gamma_4$) state with respect to the $\Gamma_2$ ($\Gamma_4$) state, which traces back to the alterations in the chemical bonding in the optimized structure. Therefore, the change in topology from Bi$_2$Se$_3$ to Sb$_2$Se$_3$ results essentially from lattice distortions caused by the replacement of Bi by Sb rather than from changes in the SOC strength.

Accordingly, new topological insulators are likely to be found by considering band shifts caused by SR effects and lattice strain. A topological phase transition due to lattice strain, for instance, is found in layered GaS and GaSe, where the SOC is rather weak. Here, we propose a feasible path to new types of topological insulators. As an example, we derive such a topologically nontrivial nature for strained InSb. (1) We start from the prototypical semiconductor GaAs, which shows at $E_F$ band characters similar to HgTe. (2) We calculate the orbital contributions to the s-type $\Gamma_1$ and p-type $\Gamma_5$ states, obtaining a mixed Ga $s$ and As $s$ character for $\Gamma_1$ [see Fig. 4(a)]. (3) We therefore replace Ga and As by heavy elements, for example, In and Sb, in order to lower the s-type $\Gamma_5$ ($\Gamma_1$) state and, eventually, invert the s and p bands. In doing so we make sure that the resultant material maintains the original symmetry. For InSb, we find that the energy gap is indeed reduced significantly as a result of the much stronger SR effects on the In $s$ and Sb $s$ electrons [see Fig. 4(b)]. However, the...
and \( \Gamma_6 \) and \( \Gamma_8 \) states maintain the standard order. (4) Finally, the lattice is strained to weaken the hybridization and reduce the energy gap. For a 3\% increase of the lattice constant, the \( \Gamma_6 \) and \( \Gamma_8 \) states are inverted in strained InSb [see Fig. 4(c)], and a topologically nontrivial nature is established. Note that step (4) is dispensable if this goal is achieved already after step (3). According to the above discussion, atoms which contribute states at \( E_F \) should be heavy elements.

IV. CONCLUSION

In conclusion, we have performed first-principles calculations to study the band inversion mechanisms in the prototypical topologically nontrivial materials HgTe, PtScBi, and Bi\(_2\)Se\(_3\) as compared to their trivial counterparts CdTe, PtScSb, and Sb\(_2\)Se\(_3\). The band shifts caused by scalar relativistic effects, spin-orbit coupling, and lattice distortion have been analyzed in detail. We have argued that scalar relativistic effects and lattice distortion play an essential role for the band inversion, more than the spin-orbit coupling. Spin-orbit coupling, however, is required to open the gap and make a system topological.\(^{34}\) Scalar relativistic effects on the electrons of heavy elements result in the band inversion in HgTe-like materials, while lattice distortions drive the band inversion in Bi\(_2\)Se\(_3\)-like materials. We propose to focus the search for topologically nontrivial materials not on a strong spin-orbit coupling but on band shifts caused by scalar relativistic effects and strain. We have discussed strained InSb as an example to demonstrate the derived scheme to design new types of topological insulators.

\(^{1}\)udo.schwingenschlogl@kaust.edu.sa

\(^{2}\)M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. 82, 3045 (2010).


\(^{7}\)L. Fu and C. L. Kane, Phys. Rev. B 76, 045302 (2007).


\(^{9}\)S. Chadoev, X. Qi, J. Kuebler, G. H. Fecher, C. Felser, and S. C. Zhang, Nat. Mater. 9, 541 (2010).


