PHYSICAL REVIEW B **91**, 125129 (2015)

Electronic phase transitions of bismuth under strain from relativistic self-consistent *GW* calculations

Irene Aguilera, Christoph Friedrich, and Stefan Blügel
Peter Grünberg Institute and Institute for Advanced Simulation, Forschungszentrum Jülich and JARA, 52425 Jülich, Germany
(Received 1 November 2014; revised manuscript received 18 February 2015; published 18 March 2015)

We present quasiparticle self-consistent *GW* (QS*GW*) calculations of semimetallic bulk Bi. We go beyond the conventional QS*GW* method by including the spin-orbit coupling throughout the self-consistency cycle. This approach improves the description of the electron and the hole pockets considerably with respect to standard density functional theory (DFT), leading to excellent agreement with experiment. We employ this relativistic QS*GW* approach to conduct a study of the semimetal-to-semiconductor and the trivial-to-topological transitions that Bi experiences under strain. DFT predicts that an unphysically large strain is needed for such transitions. We show, by means of the relativistic QS*GW* description of the electronic structure, that an in-plane tensile strain of only 0.3% and a compressive strain of 0.4% are sufficient to cause the semimetal-to-semiconductor and the trivial-to-topological phase transitions, respectively. Thus, the required strain moves into a regime that is likely to be realizable in experiment, which opens up the possibility to explore bulklike topological behavior of pure Bi.

DOI: 10.1103/PhysRevB.91.125129 PACS number(s): 71.55.Ak, 71.10.−w, 71.20.−b, 71.70.Ej

Bismuth exhibits a series of peculiarities that has made it the subject of experimental and theoretical interest for decades. The very low density of carriers with high carrier mobility and a very long mean free path make bismuth very interesting for electronic-transport studies. With the recent introduction of the classification of solids according to topology classes of the electronic structure [1], Bi has again moved into the limelight of current discussions. With a lattice structure that bears similarities to graphene and a very large spin-orbit interaction, it has all the potential to be a topological semimetal or semiconductor. From a theoretical point of view, the large spin-orbit coupling (SOC), the semimetallic character, and the very small local direct band gap at the L point of the Brillouin zone make *ab initio* calculations of the electronic structure of Bi very challenging.

As for its topological classification, some important aspects are not yet fully understood. It is widely accepted that bulk Bi is topologically trivial in contrast to Sb. This is in agreement with angle-resolved photoemission spectroscopy (ARPES) measurements of Bi [2–4] and $Bi_{1-x}Sb_x$ alloys [5–7], as well as transport studies, in which a semimetal-to-semiconductor (SMSC) [8] and a topological transition [9] are observed as the Sb concentration is varied from the topologically trivial Bi to the topological semimetal Sb, resulting in three-dimensional (3D) topological insulators for some of the alloys. However, controversies concerning the topological properties of bismuth persist. For example, it has recently been claimed [10] on the basis of ARPES measurements that bulk bismuth is topologically nontrivial, a claim that is in conflict with most previous findings including *ab initio* calculations based on density functional theory (DFT). The authors of Ref. [10] attributed these discrepancies to the systematic underestimation of the band gaps obtained in the local-density (LDA) or generalized gradient (GGA) approximations of DFT. Given the very small experimental value (11–15 meV [11–15]) for the direct band gap at the L point (E_g), whose "sign" (order of the states) controls the topological nature of Bi, it is very conceivable that DFT might incorrectly predict the sign of the band gap, and not just its magnitude. Our parity analysis within LDA is in agreement with results in the literature (see, e.g., Refs. [16–18]) and confirms the trivial character of Bi. But, surprisingly, LDA presents an unusual *overestimation* of the band gap at L (86 meV; see Table I), which is an effect usually characteristic of inverted gaps, as with those in topological insulators [19,20]. The gap at L is indeed inverted, as can be seen by varying the SOC strength from zero to 100%: the two states at L exchange order for a certain SOC strength. But the gap at T is inverted as well. There is, hence, an even number of inverted gaps at the time-reversal invariant momenta, giving rise to a topologically trivial value of the Z_2 invariant [21,22]. The GGA approximation hardly improves the LDA results (Table I). The poor description of the band gaps casts doubt on the LDA and GGA studies of the trivial-to-topological (TT) [23,24] or the SMSC [23,25] transitions of Bi, as well as the TT transition in $Bi_{1-x}Sb_x$ alloys, as these studies critically depend on the sign and the value of the band gap at L.

The cause of the band-gap problem of DFT is well known. Strictly, the single-particle energies cannot be interpreted as quasiparticle energies of the interacting electron system. The *GW* approximation [26] for the electronic self-energy, applied as a one-shot correction to DFT or, more rarely, in the quasiparticle self-consistent *GW* (QS*GW*) approach [27], remedies the aforementioned problem and is often used to correct the band gaps obtained from LDA or GGA. In most cases, it increases the band gap, leading to a much better agreement with experiments [28]. However, we show here (cf. Table I) that in the case of bismuth, the *GW* quasiparticle correction is negative, i.e., it correctly *reduces* the band gap to 32 meV in one-shot *GW* (which corresponds to the first iteration of QS*GW*) and to 13 meV within the self-consistent QS*GW* scheme. The latter value lies in the range of experimental values. For these calculations, we have devised and implemented a method that combines the SOC with the many-body renormalization in a consistent way (see below). The sign of the band gap is not changed by the quasiparticle correction, i.e., bismuth remains trivial, in accordance with the

TABLE I. Values of the direct band gap at L (E_g) and the indirect band gap (E_0) [see Fig. 2(b)] for bulk bismuth (in meV) calculated within LDA, GGA, GW, and QSGW, compared with experimental results.

	LDA	GGA	GW	QSGW	Expt.
Direct gap at L (E_g)	86	74	32	13	11 to 15[a]
Indirect gap T-L (E_0)	−105	−72	−66	−33	−32 to −39[b]

[a]11.0 [11], 13.6 [12], 15.0 [13,14], 15.3 [15].
[b]−32.0 [29], −36.0 [13], −38.0 [30], −38.2 [13], −38.5 [15].

vast majority of experimental and theoretical studies. However, the phase boundary at which bismuth becomes topologically nontrivial is considerably closer in QSGW than in LDA. We will show that only a relatively modest in-plane tensile strain of 0.3% and compressive strain of 0.4% are sufficient to cause the SMSC and TT phase transition, respectively, which opens up the possibility to explore the 3D topological properties of *pure* Bi.

Bulk bismuth crystallizes in the $A7$ rhombohedral structure with $R\bar{3}m$ space group and two atoms per unit cell. The cell and the atomic positions are determined by the internal parameter u, the rhombohedral lattice parameter $a_{\rm rho}$, and the rhombohedral angle $\alpha_{\rm rho}$ [see Fig. 1(a)]. The layered structure consists of Bi bilayers separated by a van der Waals gap. For the undistorted reference lattice, we employ experimental parameters from Ref. [31]: $a_{\rm rho} = 4.7458$ Å, $\alpha_{\rm rho} = 57.230°$, $u = 0.23389$. Equivalently, the structure can be described as hexagonal with in-plane and out-of-plane lattice constants $a_0 = 4.5460$ Å and $c_0 = 11.862$ Å ($c_0/a_0 = 2.6093$). These lattice parameters were measured at 298 K. The gray vertical areas in Fig. 3 show the range of experimental values of c/a for temperatures between 4 and 298 K [31]. All of our calculations are carried out with the DFT code FLEUR [32] and the GW code SPEX [33], realizations in the all-electron full-potential linearized augmented-plane-wave (FLAPW) formalism. The convergence parameters of the DFT, GW, and QSGW calculations are given in the Appendix. To include relativistic effects in the DFT calculations, we employ the scalar-relativistic [34] approximation inside the muffin-tin spheres, and the SOC is included in a self-consistent manner [35].

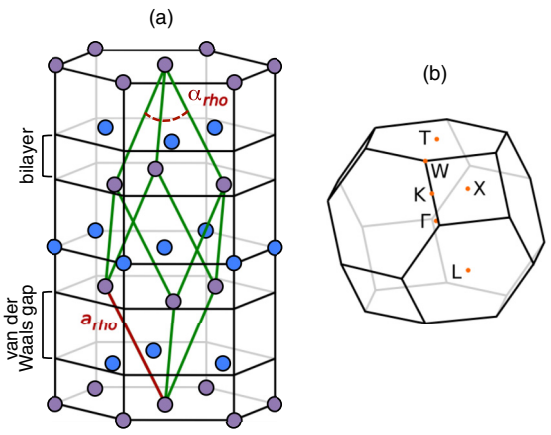

FIG. 1. (Color online) (a) Rhombohedral (green) unit cell and hexagonal crystal structure of bulk Bi. The blue and purple atoms distinguish the atoms in the two layers of each bilayer. (b) Bulk Brillouin zone of Bi expressed in the rhombohedral unit cell.

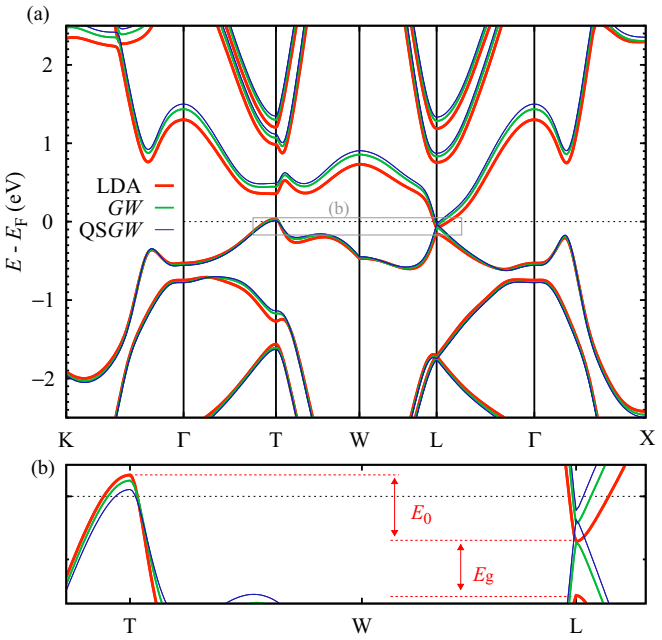

FIG. 2. (Color online) LDA, GW, and QSGW band structures of bulk Bi. The inset in (b) shows the electron and hole pockets together with the E_g and E_0 gaps corresponding to the LDA bands.

Most of the one-shot GW calculations that include SOC and all of the QSGW ones published so far employ the SOC as an *a posteriori* correction after the quasiparticle correction to DFT has been carried out without SOC. This is, of course, an approximation. For example, many-body renormalization effects of spin-orbit split bands are then not taken into account (see Ref. [36] for a detailed discussion). In our relativistic calculations, the SOC is already incorporated in the noninteracting system that serves as the starting point for the quasiparticle calculation [36,37]. The one-shot GW self-energy thus acquires terms that couple the two spin channels, enabling a many-body renormalization of the SOC itself. In this work, we propose to extend this principle to the self-consistent QSGW calculations, so that the self-energy contains spin off-diagonal blocks that it inherits from the SOC throughout the whole self-consistent process. (In analogy to the notation $G^{\rm SOC}W^{\rm SOC}$ of Ref. [36], one could denote this approach by QS$G^{\rm SOC}W^{\rm SOC}$, but for simplicity we will simply write GW and QSGW in the following.) Our results show that this treatment is crucial for Bi. The *a posteriori* correction mentioned above, on the other hand, leads to unphysical results for Bi (see Appendix), which then appears as a topological insulator with large band gaps (83 and 259 meV in one-shot GW and QSGW, respectively) instead of a trivial semimetal.

In addition to the value of the band gap at the L point, the properties of Bi are also determined by the states at the symmetry point T. In particular, the semimetallic character of bismuth is caused by the overlap between the highest valence band at T and the lowest conduction band at L [see Fig. 2(b)].

This creates three very small electron pockets (at the three L points in the Brillouin zone) and one hole pocket at T. Several attempts to describe these pockets theoretically have been published in recent years. For a review and comparison of them see, e.g., Table II in Ref. [17]. The approaches used include LDA and GGA [17,38,39], tight-binding models [40], and empirical pseudopotential calculations [16], but no GW results have been published so far. The results from the models and DFT calculations have not been satisfactory: the absolute value of the overlap or, in other words, the indirect band gap E_0 came out too large in general compared to experiment.

Our values of E_0 are shown in Table I. Both the LDA and GGA overestimate indeed this indirect band gap, in line with previous publications. While the one-shot GW approach does correct E_0 in the right direction, it uses LDA as a starting point and, thus, inherits some of this overestimation. The self-consistent procedure of QSGW, on the other hand, makes the result independent of the starting point and moves E_0 inside the range of experimental values.

Figure 2 shows the band structure of bulk bismuth obtained with LDA, GW, and QSGW along the path connecting the high-symmetry **k** points represented in Fig. 1(b). Whereas the overall shape of the band structure remains basically unaffected by the many-body corrections, there are important changes in the details. In particular, while in most parts the valence and conduction bands are shifted apart by the many-body corrections as commonly seen in GW calculations, the changes of the dispersion in the vicinity of the Fermi energy are such that the absolute values of E_g and E_0 get smaller and reach values of 13 and −33 meV, respectively. As has already been discussed before [19,20,36,37], this is the result of a delicate interplay between the SOC and the many-body renormalization described by the GW approximation. We conclude that the present relativistic QSGW is the method of choice to study the band structure of Bi and its topological character as well as the TT and SMSC transitions that Bi experiences under strain.

A TT transition was indeed recently observed experimentally for strained Bi(111) ultrathin films [23], and a similar transition was found for strained Bi nanowires elongated about 2% [41]. In the case of thin films, a strain can be introduced naturally by a lattice mismatch between film and substrate. To simulate the strain in our calculations, we apply a volume-conserving distortion of the structure by varying the c/a ratio, and keeping the internal parameter u constant. A compression in the c direction is accompanied by a corresponding increase of a, and vice versa. This is equivalent to varying the rhombohedral angle α_{rho}; see Fig. 1(a). As the c/a ratio increases, the nearest-neighbor distance reduces and the van der Waals gap expands. Both effects lead to a strengthening of the in-plane interactions, while the interbilayer coupling weakens. The altered crystal field reduces the band gap at L until it reaches a zero value at a critical c/a ratio and becomes negative. This undoes the band inversion and causes a sign change in the calculation of the Z_2 topological invariants [21,22], which, for systems with inversion symmetry, derive from the parities of the occupied states at the time-reversal invariant momenta. The parity of a wave function remains unchanged upon applying the GW quasiparticle correction. The Z_2 topological invariants can, thus, only change if there is an interchange of valence and conduction states with respect to LDA. A topological-to-trivial transition in the topological insulator Bi_2Se_3 as a result of strain has also been discussed in the literature. It has been argued [42,43] that tensile out-of-plane strain reduces the spin-orbit strength, leading to a trivial phase. We do not find such an effect for bulk Bi: no substantial increase or decrease in the spin-orbit strength is observed upon application of strain.

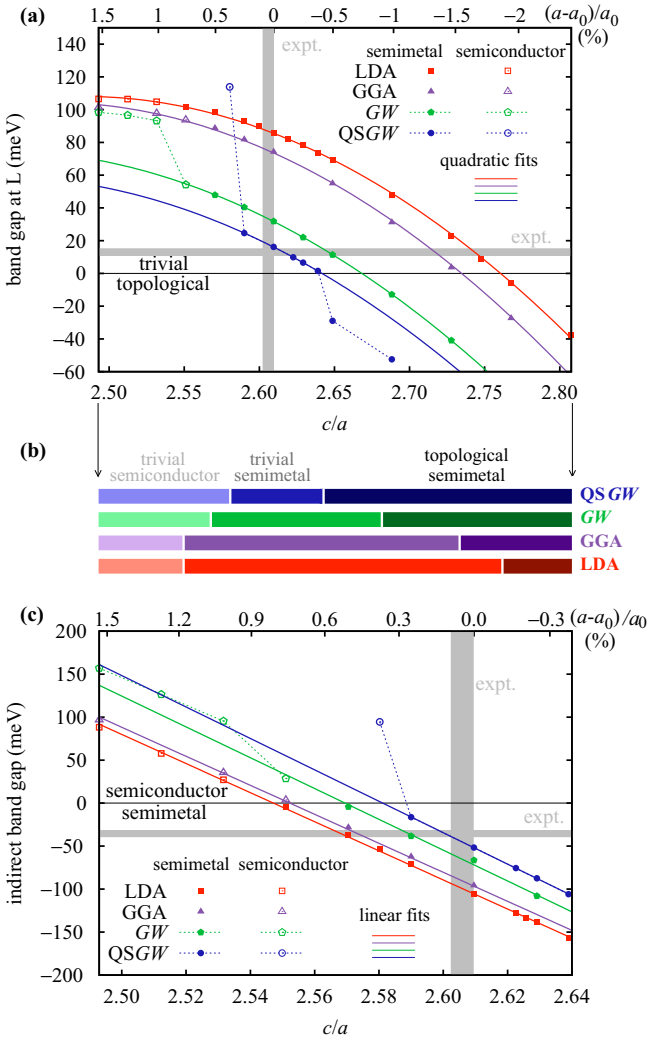

FIG. 3. (Color online) (a) LDA, GGA, GW, and QSGW direct band gap E_g of bismuth at L for different strains (varying the ratio c/a and keeping the volume constant). (b) Three electronic phases are found for Bi as a function of the ratio c/a within all the approaches. (c) LDA, GGA, GW, and QSGW indirect band gap E_0 of bismuth for different values of c/a. On the upper x axis of (a) and (c), we show the percentage of in-plane strain. The range of experimental values of c/a [31], the experimental gap at L [11–15], as well as the experimental indirect gap [13,15,29,30] are represented as gray shaded areas in (a) and (c). Small differences to the QSGW values in Table I are due to the use of a reduced **k**-point set; see Appendix.

Figure 3(a) shows the direct band gap as a function of the c/a ratio with the range of equilibrium experimental values of bulk Bi represented as gray shaded areas. Positive values of the gap denote a trivial phase, whereas negative

values correspond to a topologically nontrivial case. In addition to the semimetallic state (solid symbols) in Fig. 3(a), we also observe a transition to a semiconducting state (open symbols) for tensile in-plane strain. This SMSC transition can also be seen in Fig. 3(c), showing the variation of the indirect band gap, which changes sign and becomes positive at the point of transition.

At the SMSC transition, the GW and QSGW curves exhibit abnormal behavior; there is a jump (dashed lines) where we expect the curve to be continuous. An analogous steplike behavior can be seen in the QSGW curve at the TT transition. As shown in the Appendix, these anomalies are an artifact of the **k**-point convergence: The steps get systematically smaller when one improves the **k**-point sampling, and convergence is achieved much more readily in the trivial semimetallic phase than in the other two phases, so that we can take the points between the transitions in Fig. 3 as converged, and the fitted curves (solid lines) show the expected behavior in the sampling limit. We have employed a quadratic function for the direct gap and a linear function for the indirect gap as a function of c/a, which for QSGW give $E_g = -1226.28(c/a)^2 + 5939.05(c/a) - 7131.51$ and $E_0 = -1827.57(c/a) + 4717.26$ in meV.

Figure 3(b) shows the three electronic phases that we can distinguish in bulk Bi under strain: a trivial semiconducting phase, a trivial semimetallic phase, and a topological semimetallic phase. The overlap between the electron and hole pockets, or the indirect gap, is caused by the distortion of the cell with respect to a cubic structure. It seems, therefore, that bismuth cannot be in a topological insulating phase because the stress necessary to make it topological distorts the cell rhombohedrally, distancing it even further from the cubic symmetry so that Bi becomes more and more metallic. Figure 3 demonstrates that whereas the two transitions are qualitatively described correctly by the four approaches, the LDA and GGA predict a large critical strain that would be very difficult to achieve experimentally. In particular, the TT transition is predicted to occur at a critical in-plane compressive strain of 1.9% and 1.6%, respectively (see Table II). On the contrary, the strain needed to cause the transition is significantly smaller in GW (0.7%) and, in particular, in QSGW (0.4%). In other words, taking into account structural relaxation, significantly thicker samples could be grown under the relatively modest critical strain that is predicted by QSGW than under that predicted by DFT.

In conclusion, we propose a QSGW scheme in which the SOC is included throughout the whole self-consistent process. Our results show that such relativistic QSGW calculations are the right *ab initio* approach to determine the electronic structure and the topological classification of bismuth. They predict its two most significant and delicate electronic properties in quantitative agreement with experiments: the direct gap at L, which determines whether Bi is topological or trivial, and the indirect gap between L and T, which describes the overlap between the electron and hole pockets and thus determines the semimetallic character of Bi. In response to a recent controversy [10], we conclude the following from our calculations: bulk Bi is a trivial semimetal.

We showed that Bi can undergo electronic phase transitions under rather small lattice strain: a trivial SMSC transition under 0.3% in-plane tensile strain and a TT semimetal transition under 0.4% compressive strain. In contrast to LDA and GGA, the critical strains predicted by QSGW are attainable by experiments. The presented results motivate additional experimental efforts to prepare Bi as a topological insulator by opening band gaps in slightly strained Bi. For example, the small strain needed to cause the TT transition concluded by this work, together with the SMSC transition driven by quantum-size effects in relatively thick films [44], reveals the potential to observe a topological insulating behavior in bulklike films of pure Bi. In fact, first experimental evidences of 3D-like topological thin films of Bi are currently under discussion [45].

TABLE II. Values of the lattice parameters a and c and the ratio a/c for which (a) the TT and (b) the SMSC transitions occur for the different theoretical approaches. The values in brackets indicate the compressive (−) or tensile (+) strain in percentage. The difference in the total energies (ΔE_{tot}) between the structure for which the transition occurs and the unperturbed reference lattice has been calculated within LDA. Note that we used $a_0 = 4.5460$ Å, $c_0 = 11.862$ Å, and $c_0/a_0 = 2.6093$ [31] for the undistorted lattice.

	a_t (Å)	c_t (Å)	c_t/a_t	ΔE_{tot} (meV)
(a) TT:				
LDA	4.461(−1.9%)	12.315	2.761(+5.8%)	15.8
GGA	4.475(−1.6%)	12.238	2.735(+4.8%)	11.7
GW	4.513(−0.7%)	12.036	2.667(+2.2%)	4.1
QSGW	4.527(−0.4%)	11.960	2.642(+1.2%)	1.9
(b) SMSC:				
LDA	4.579(+0.7%)	11.690	2.553(−2.2%)	0.5
GGA	4.580(+0.7%)	11.685	2.551(−2.2%)	0.6
GW	4.569(+0.5%)	11.744	2.570(−1.5%)	0.2
QSGW	4.562(+0.3%)	11.776	2.581(−1.1%)	0.1

The present results demonstrate that LDA and GGA lack the quantitative accuracy to predict the critical strain for which the electronic phase transitions occur. We speculate that standard DFT also yields wrong predictions for the critical concentration at which Bi experiences a TT transition upon alloying with other elements (such as in $Bi_{1-x}Sb_x$ alloys [5,7–9,18]) or for the critical thickness at which a SMSC transition caused by quantum-size effects takes place. Bi is one example, but we believe it is much more general. Other examples could include the very large strains necessary to induce topological phase transitions in Bi_2Se_3 [42,43] and in $TlBiS_2$ and $TlSbS_2$ [46]. It would thus be desirable to reinvestigate the critical points for these transitions within the relativistic QSGW method proposed in this work.

We thank Xiaofeng Jin for the very interesting discussions that motivated this work, Gregor Mussler for his comments on the experimental growth of Bi, and Gustav Bihlmayer for valuable 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).

APPENDIX

The calculations are carried out with the DFT code FLEUR [32] and the GW code SPEX [33], which are based on the all-electron FLAPW formalism. For the valence electrons, space is partitioned into spherical regions around the atoms (muffin-tin spheres) and interstitial region between the spheres. We use an angular momentum cutoff $l_{\max} = 10$ in the muffin-tin spheres and a plane-wave cutoff of 4.5 bohr^{-1} in the interstitial region. For the DFT calculations, we employ either the Perdew-Zunger [47] parametrization of the LDA exchange-correlation functional or the Perdew-Burke-Ernzerhof [48] parametrization of the GGA. The one-shot GW calculations are always performed using the LDA mean-field system as a starting point.

Due to the very small band gap of bulk Bi (11 to 15 meV for the direct one [11–13,15] and −32 to −39 for the indirect one [13,15,29,30]), all computational parameters entering the GW and QSGW calculations are chosen after a systematic study in order to obtain thoroughly converged results. An angular momentum cutoff of $l = 5$ and a linear momentum cutoff of 2.9 bohr^{-1} are employed to construct the mixed product basis [33,49], used to represent the dielectric matrix and the screened interaction.

The one-shot GW and self-consistent QSGW results for the undistorted structure are obtained with a $6 \times 6 \times 6$ **k**-point mesh sampling the Brillouin zone (Table I and Fig. 2). Fortunately, we found that the QSGW results converge faster with respect to the number of **k** points than the results of one-shot GW so that we could afford to reduce the **k**-point mesh to $4 \times 4 \times 4$ for the QSGW calculations of bismuth under strain. For more details about the **k**-point convergence, see below.

To compute the Green function and the polarization function, 500 bands are used. This corresponds to approximately 130 eV above the Fermi energy. The complete fifth shell of Bi is treated as valence states by the use of local orbitals. In this way, the contribution of the $5s5p5d$ states to the Green function and the electronic screening is fully taken into account. Despite their low energetic position, inclusion of the $5s$ and $5p$ states effects a change of 5 meV in the band gap at L. To describe high-lying states accurately and to avoid linearization errors [50–52], we complement the basis for the valence electrons for each atom by two local orbitals per angular momentum up to $l = 3$ with energy parameters far up in the unoccupied states. For the interpolation of the band structures in Fig. 2, maximally localized Wannier functions obtained by the WANNIER90 library [53] and a $6 \times 6 \times 6$ **k**-point mesh were employed.

We employ the Dirac equation for the core electrons (up to and including the fourth shell) so that relativistic effects are fully accounted for. The core states are thus represented by four-component spinors. This representation is retained for the evaluation of the core-valence contribution to the exchange self-energy. (Using averaged core states that are represented in terms of nonrelativistic lm spherical harmonics instead of the jm_j spinors introduces an error of 4 and 7 meV in the direct band calculated with one-shot GW and QSGW, respectively.)

As a semimetal, bismuth exhibits metallic screening, which is described technically by the so-called Drude term in the screened interaction W. This term stems from virtual intraband transitions across the Fermi surface. It can be formulated in a functional form being proportional to the square of the plasma frequency, which in turn is evaluated by an integration over the Fermi surface. The Drude term can be treated analytically [33] and, as long as the Fermi surface is sufficiently big, it normally does not pose any numerical problem. However, bismuth has a very small Fermi surface due to the tiny electron and hole pockets, which eventually leads to a very sharp "Drude peak" in the GW self-energy, impeding a straightforward numerical solution of the nonlinear quasiparticle equation. One could also say that while the Drude term is actually treated correctly, it gets too much weight because of the finite **k**-point mesh. Therefore, we neglect the Drude term in our calculations and, instead, simply scale the head element of $W(\mathbf{k},\omega)$ in the limit $\mathbf{k} \to \mathbf{0}$ to enforce metallic screening. This hardly changes W, and we have found that it leads to a very favorable **k**-point convergence; see below. (In the limit of dense **k**-point sets, the treatment of the Drude term, which affects only a single **k** point, i.e., the Γ point, becomes immaterial.)

In Fig. 4, we show the direct band gap at the L point for each iteration in the QSGW self-consistency procedure. In particular, we distinguish between calculations (i) without SOC, (ii) with spin-orbit coupling added *a posteriori* to a result obtained without SOC (QSGW+SOC), which is a correction scheme used in most GW calculations so far, and (iii) the relativistic approach used in the present work in which spin-orbit coupling is included in the calculation of both G and W throughout the whole self-consistent procedure. (In an earlier publication [36], we have used the notation $G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ for the corresponding one-shot calculation. By analogy, one could use the notation QS$G^{\mathrm{SOC}}W^{\mathrm{SOC}}$ for the present self-consistent approach.) The *a posteriori* SOC correction leads to unphysical results for Bi, which appears as a topological insulator instead of a trivial semimetal. We attribute the lack of published GW results of bulk Bi to the crucial and delicate treatment of SOC, but also to the difficulties encountered with the semimetallicity of the system exhibiting very tiny electron and hole pockets (leading to a very sharp Drude peak in the

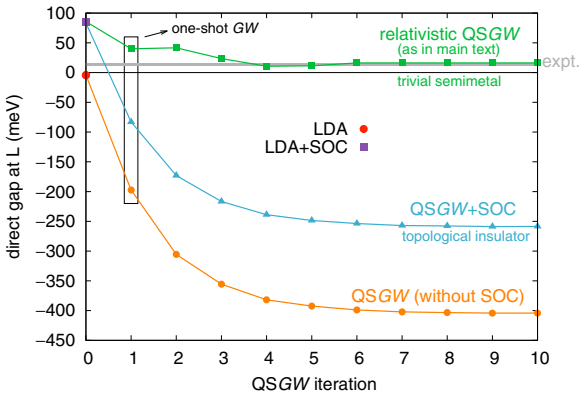

FIG. 4. (Color online) Convergence of the QSGW direct band gap at the L point: without SOC (circles), with SOC added *a posteriori* (triangles), and with the relativistic approach used in the main text (squares). Iteration 0, 1, and 10 correspond to the LDA, one-shot GW, and QSGW values, respectively.

self-energy, as discussed above) and to the smallness of the gaps requiring the calculations to be converged to within a few meV.

We have also performed relativistic QSGW_0 calculations, i.e., we keep the W at the LDA level while G is updated self-consistently. The converged QSGW_0 result for the band gap at L differs only by 0.3 meV from the QSGW one. This indicates that the screening obtained with LDA is a good approximation and that self-consistency induces changes mainly in G. Finally, to quantify the effects of dynamics, we have also performed self-consistent calculations within the static Coulomb-hole screened-exchange (COHSEX) [26] approximation to the GW self-energy (including SOC). In this approximation, the changes are more pronounced. With respect to QSGW, the direct gap at L is reduced by 12 meV, nearly making the system topological, and the indirect band gap shrinks to −4 meV, nearly making the system semiconducting.

To explain the discontinuities in Figs. 3(a) and 3(c) and to prove their origin as a **k**-point convergence issue, we have performed GW and QSGW calculations with increasing number of **k** points (Fig. 5). To allow for these calculations, we had to reduce some of the parameters in the calculation such as the plane-wave cutoff in the interstitial region for the LDA calculations (now 4.0 bohr^{-1}), and both the angular momentum cutoff ($l = 4$) and the linear momentum cutoff (2.6 bohr^{-1}) to construct the mixed product basis for the GW calculations. We also reduced the number of bands for the GW calculations to 100 and only 5d local orbitals were used in the calculations. This obviously reduces the accuracy of the band gaps obtained—and therefore they differ from those of the main text— but it is not our intention in this appendix to analyze the quantitative results, only the qualitative behavior due to the **k**-point convergence.

From Fig. 5, we can conclude the following about the **k**-point convergence of one-shot and self-consistent GW results: (1) The calculations of the trivial semimetallic phases can be considered converged with a $6 \times 6 \times 6$ **k**-point set for the one-shot GW calculations and $4 \times 4 \times 4$ for the QSGW ones. (2) The semiconducting and topological phases require much larger **k**-point meshes, but (3) the corresponding results converge systematically towards an extrapolated quadratic (linear) fit to the converged trivial semimetallic phases for the direct (indirect) band gap. We note that in the case of one-shot GW calculations, the discontinuity in the curves due to the **k**-point convergence of the topological phases does not appear at the TT transition of GW but at that of LDA because of the use of the LDA mean-field system as the starting point for the one-shot GW quasiparticle correction.

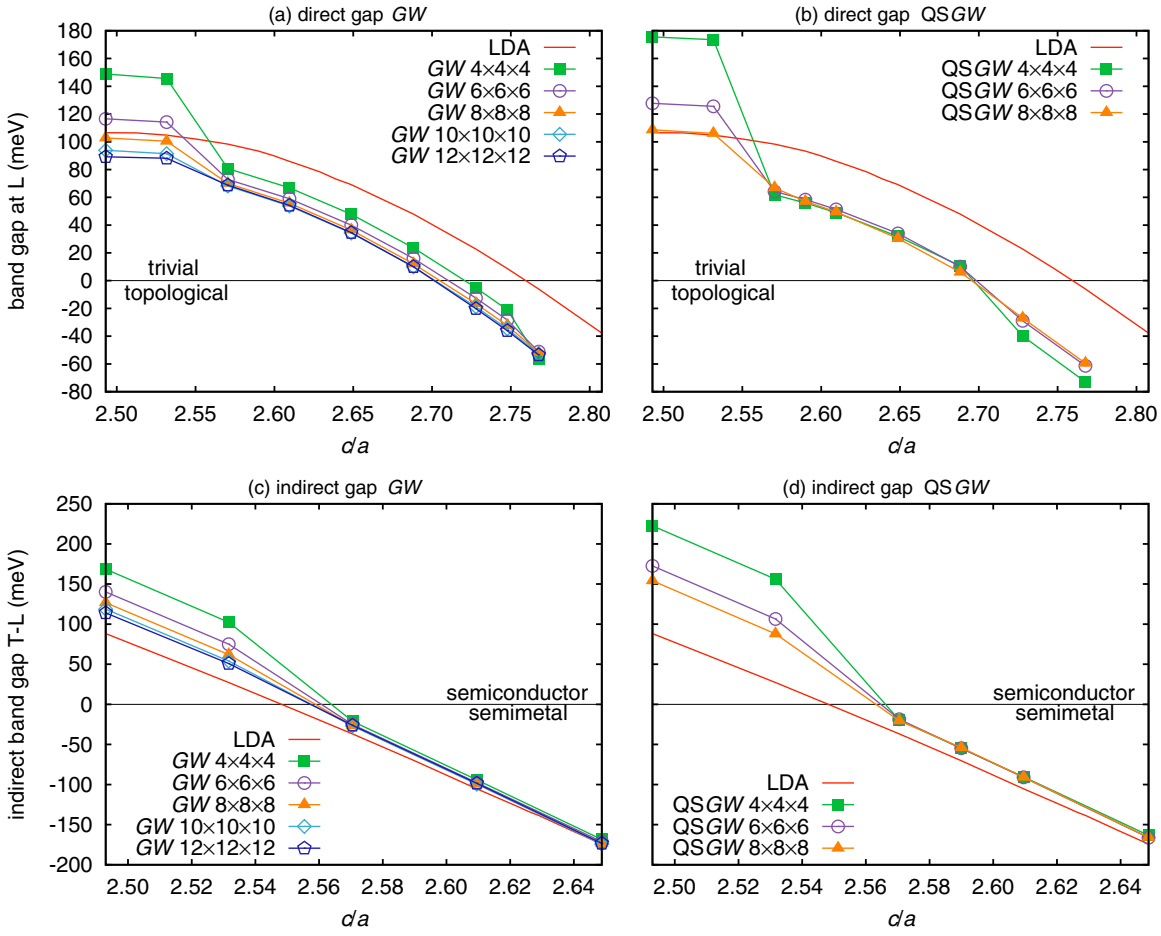

FIG. 5. (Color online) LDA, GW, and QSGW values of the direct and indirect band gaps of bulk Bi as a function of the ratio c/a for different **k**-point sets, but for a set of reduced convergence parameters; see text.

[1] J. E. Moore, Nature (London) **464**, 194 (2010).
[2] C. R. Ast and H. Höchst, Phys. Rev. B **67**, 113102 (2003).
[3] Y. M. Koroteev, G. Bihlmayer, J. E. Gayone, E. V. Chulkov, S. Blügel, P. M. Echenique, and P. Hofmann, Phys. Rev. Lett. **93**, 046403 (2004).
[4] P. Hofmann, Prog. Surf. Sci. **81**, 191 (2006).
[5] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature (London) **452**, 970 (2008).
[6] P. Roushan, J. Seo, C. V. Parker, Y. S. Hor, D. Hsieh, D. Qian, A. Richardella, M. Z. Hasan, R. J. Cava, and A. Yazdani, Nature (London) **460**, 1106 (2009).
[7] D. Hsieh, Y. Xia, L. Wray, D. Qian, A. Pal, J. H. Dil, J. Osterwalder, F. Meier, G. Bihlmayer, C. L. Kane, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Science **323**, 919 (2009).
[8] E. Tichovolsky and J. Mavroides, Solid State Commun. **7**, 927 (1969).
[9] N. A. Red'ko and N. A. Rodionov, Pis'ma Zh. Eksp. Teor. Fiz. **42**, 246 (1985) [J. Exp. Theor. Phys. Lett. **42**, 303 (1986)].
[10] Y. Ohtsubo, L. Perfetti, M. O. Goerbig, P. L. Fèvre, F. Bertran, and A. Taleb-Ibrahimi, New J. Phys. **15**, 033041 (2013).
[11] M. Maltz and M. S. Dresselhaus, Phys. Rev. B **2**, 2877 (1970).
[12] M. P. Vecchi and M. S. Dresselhaus, Phys. Rev. B **10**, 771 (1974).
[13] R. T. Isaacson and G. A. Williams, Phys. Rev. **185**, 682 (1969).
[14] R. N. Brown, J. G. Mavroides, and B. Lax, Phys. Rev. **129**, 2055 (1963).
[15] G. E. Smith, G. A. Baraff, and J. M. Rowell, Phys. Rev. **135**, A1118 (1964).
[16] S. Golin, Phys. Rev. **166**, 643 (1968).
[17] I. Timrov, T. Kampfrath, J. Faure, N. Vast, C. R. Ast, C. Frischkorn, M. Wolf, P. Gava, and L. Perfetti, Phys. Rev. B **85**, 155139 (2012).
[18] J. C. Y. Teo, L. Fu, and C. L. Kane, Phys. Rev. B **78**, 045426 (2008).
[19] O. V. Yazyev, E. Kioupakis, J. E. Moore, and S. G. Louie, Phys. Rev. B **85**, 161101(R) (2012).
[20] I. Aguilera, C. Friedrich, G. Bihlmayer, and S. Blügel, Phys. Rev. B **88**, 045206 (2013).
[21] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).
[22] L. Fu, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **98**, 106803 (2007).
[23] T. Hirahara, N. Fukui, T. Shirasawa, M. Yamada, M. Aitani, H. Miyazaki, M. Matsunami, S. Kimura, T. Takahashi, S. Hasegawa, and K. Kobayashi, Phys. Rev. Lett. **109**, 227401 (2012).
[24] Z. Liu, C.-X. Liu, Y.-S. Wu, W.-H. Duan, F. Liu, and J. Wu, Phys. Rev. Lett. **107**, 136805 (2011).
[25] G. Gutierrez, E. Menendez-Proupin, and A. K. Singh, J. Appl. Phys. **99**, 103504 (2006).
[26] L. Hedin, Phys. Rev. **139**, A796 (1965).
[27] S. V. Faleev, M. van Schilfgaarde, and T. Kotani, Phys. Rev. Lett. **93**, 126406 (2004).
[28] M. van Schilfgaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).
[29] V. S. Edel'man, J. Exp. Theor. Phys. **41**, 125 (1975).
[30] R. J. Dinger and A. W. Lawson, Phys. Rev. B **7**, 5215 (1973).
[31] D. Schiferl and C. S. Barrett, J. Appl. Crystallogr. **2**, 30 (1969).
[32] www.flapw.de
[33] C. Friedrich, S. Blügel, and A. Schindlmayr, Phys. Rev. B **81**, 125102 (2010).
[34] D. D. Koelling and B. N. Harmon, J. Phys. C **10**, 3107 (1977).
[35] C. Li, A. J. Freeman, H. J. F. Jansen, and C. L. Fu, Phys. Rev. B **42**, 5433 (1990).
[36] I. Aguilera, C. Friedrich, and S. Blügel, Phys. Rev. B **88**, 165136 (2013).
[37] R. Sakuma, C. Friedrich, T. Miyake, S. Blügel, and F. Aryasetiawan, Phys. Rev. B **84**, 085144 (2011).
[38] X. Gonze, J.-P. Michenaud, and J.-P. Vigneron, Phys. Scr. **37**, 785 (1988).
[39] A. B. Shick, J. B. Ketterson, D. L. Novikov, and A. J. Freeman, Phys. Rev. B **60**, 15484 (1999).
[40] Y. Liu and R. E. Allen, Phys. Rev. B **52**, 1566 (1995).
[41] E. Condrea, A. Gilewski, and A. Nicorici, J. Phys.: Condens. Matter **25**, 205303 (2013).
[42] S. M. Young, S. Chowdhury, E. J. Walter, E. J. Mele, C. L. Kane, and A. M. Rappe, Phys. Rev. B **84**, 085106 (2011).
[43] W. Liu, X. Peng, C. Tang, L. Sun, K. Zhang, and J. Zhong, Phys. Rev. B **84**, 245105 (2011).
[44] S. Xiao, D. Wei, and X. Jin, Phys. Rev. Lett. **109**, 166805 (2012).
[45] K. Zhu, L. Wu, X. Gong, S. Xiao, S. Li, X. Jin, M. Yao, D. Qian, M. Wu, J. Feng, Q. Niu, F. de Juan, and D.-H. Lee, arXiv:1403.0066.
[46] Q. Zhang, Y. Cheng, and U. Schwingenschlögl, Scientific Reports **5**, 8379 (2015).
[47] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[48] J. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **78**, 1396(E) (1997).
[49] T. Kotani and M. van Schilfgaarde, Solid State Commun. **121**, 461 (2002).
[50] C. Friedrich, A. Schindlmayr, S. Blügel, and T. Kotani, Phys. Rev. B **74**, 045104 (2006).
[51] C. Friedrich, M. C. Müller, and S. Blügel, Phys. Rev. B **83**, 081101(R) (2011).
[52] G. Michalicek, M. Betzinger, C. Friedrich, and S. Blügel, Comput. Phys. Commun. **184**, 2670 (2013).
[53] A. A. Mostofi, J. R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, Comput. Phys. Commun. **178**, 685 (2008).