A closer look at how symmetry constraints and the spin–orbit coupling shape the electronic structure of Bi(111)

Marisol Alcántara Ortigoza1,* and Talat S Rahman2

1 Department of Physics, Tuskegee University, Tuskegee Institute, Tuskegee, AL 36088, United States of America
2 Department of Physics, University of Central Florida, Orlando, FL 32816, United States of America

E-mail: Marisol.AlcantaraOrtigoza@ucf.edu and malcantaraortigoza@tuskegee.edu

Received 5 July 2023, revised 13 September 2023
Accepted for publication 19 September 2023
Published 28 September 2023

Abstract

Fully relativistic density-functional-theory calculations of Bi(111) thin films are analyzed to revisit their two metallic surface-states branches. We first contrast these metallic branches with surface states arising at gaps in the valence band opened by the spin–orbit coupling (SOC). We find that the two metallic branches along \( \Gamma M \) do not overlap with the bulk band at the zone boundary, \( M \). We show that the spin texture observed in such states cannot be traced to the lifting of Kramers’ degeneracy. Instead, we track them to the \( m_j = \pm \frac{1}{2} - m_j = \pm \frac{3}{2} \) SOC splitting, the potential anisotropy for in-plane and out-of-plane states, and the coupling between the opposite surfaces of a slab occurring near \( M \), which is driven by a spatial redistribution of the four metallic states composing the two metallic branches. Each of these branches appears to be non-degenerate at the tested surface, yet each is degenerate with another state of opposite spin at the other surface. Nevertheless, the four metallic states bear some contribution on both surfaces of the film because of their spatial redistribution near \( M \). The overlapping among these states near \( M \), afforded by their spatial redistribution on both surfaces, causes a hybridization that perpetuates the splitting between the two branches, makes the film’s electronic structure thickness dependent near \( M \), extinguishes the magnetic moment of the metallic states avoiding the magnetic-moment discontinuity at \( M \), and denies the need or expectancy of the metallic branches becoming degenerate at \( M \). We propose that the opposite spin polarization observed for the two metallic branches occurs because the surface atoms retain their covalent bonds and thus cannot afford magnetic polarization. We show that the Rashba-splitting of the metallic states for inversion-asymmetric films does not have a fixed magnitude but can be tuned by changing the perturbation breaking inversion symmetry.

Keywords: Bi(111), spin orbit coupling, surface states, topological semimetals, Rashba splitting, Kramers’ degeneracy, density functional theory

* Author to whom any correspondence should be addressed.
1. Introduction

The singular properties and applications of Bismuth sprouting since 1911 make it one of the most remarkable and enlightening elements of the periodic table. It is the heaviest non-radioactive element. Bi allowed an ‘early’ discovery of the Seebeck, the de Haas–van Alphen, the Shubnikov–de Haas, and the Nernst effects, all of which are inherently present in all metals but were more challenging to observe. Bismuth was in fact the first metal whose Fermi surface was experimentally identified [1] and provided the basis to determine that of any metal. Bi also set the grounds for the still ongoing rational search and optimization of thermoelectric materials. Furthermore, Bi and Bi$_{1-x}$Sb$_x$ ($x < 0.06$) were among the first materials for which the excitonic-insulator phase was considered and searched [2–4]. Bulk Bi was also the first non-superconducting material (at least in the ordered phase) that was found to be a superconductor for structurally bulk-like nanoparticles [5]. Furthermore, and unoincidentally perhaps [5, 6], Bi bicrystals and nanoparticles display a superconducting phase ($T_c = 5$ K) not observed in ordered bulk Bi down to 50 mK [5, 7–9]. Not to mention here the role of Bi and its alloys that has led to the discovery of topological insulators, most of which are Bi-based (see [10–12] and references therein). Yet today, Bi remains a prime subject of intense investigation just as more than 100 years ago.

Unsurprisingly, the interest in Bi has revived at various research fronts. In particular the Bi(111) surface has been of great interest. For example, there is the question whether or not the Bi(111) surface is a better conductor than bulk Bi [13–15]. More importantly, the interpretation of angle-resolved photoemission spectroscopy (ARPES) data [11] has led to the conclusion that Bi is a non-trivial semimetal since the metallic surface states of Bi(111) remain split at the zone boundary of the surface Brillouin zone (SBZ), the $M$-point. Recent tight binding calculations, however, indicate that the evidence of a non-trivial surface-bulk connectivity of the metallic states of Bi(111) may be deceptive because both sides of the film are coupled, thus the available observations are inconclusive about Bi being non-trivial [16]. Nevertheless, the two metallic surface states would characterize Bi as trivial or non-trivial semimetal depending on whether they comprise a Kramers’ pair that split away from $\Gamma$ and $M$. The origin and properties of these states, particularly at $M$, have been the subject of much discussion given the inherent complexities and richness in geometric and electronic structure arising from the interplay between inversion symmetry or lack thereof, spin–orbit coupling (SOC), and long-range interactions. The above continues to be the focus of the present work.

Before providing a summary of the current state of knowledge on the geometric and electronic structure of Bi(111), we remind the reader of some of the same for bulk Bi. As is known, the electronic properties of Bi are tightly related to its crystal structure, the so-called A7 or $\alpha$As structure, which has the space group $R3m$. The primitive cell of the A7 structure is a rhombohedron that contains two atoms. The A7 structure largely resembles a simple cubic structure with a slight distortion along the body diagonal: the so-called Peierls or A7 distortion. In bulk Bi, each atom has three equidistant nearest-neighbor atoms and three equidistant next-nearest neighbors, slightly further away (see figure 1). This structure thus renders A7-materials as composed of puckered bilayers (BLs) stacked perpendicular to the c-axis or trigonal axis (see figure 2(a)). The three nearest neighbors of a Bi atom lie within the same bilayer (BL), whereas its three next-nearest neighbors are located in the adjacent BL. As a result, every atom in an A7-material has three equivalent ‘short’ and ‘strong’ covalent-like bonds and three equivalent ‘long’ and ‘weak’ metallic-like bonds. A detailed description of the above rhombohedral and hexagonal representations can be found elsewhere [17]. Group–V elements arrange themselves in the A7 structure because the A7-distortion creates a band gap almost everywhere in the Brillouin zone (BZ) that stabilizes the structure [17, 18]. We shall refer to this partial gap as the Peierls-distortion or the A7-distortion gap.

The covalent intra-bilayer bonds (within each bilayer) are much stronger than the metallic-like inter-bilayer bonds, for which Bi crystals easily cleave along the (111) planes and do not reconstruct [19–21], giving rise to the rhombohedral Bi(111) surface (see figure 2(b))—also known as the pseudocubic–(111) or the hexagonal–(0001) surface (In the following, however, we confine ourselves to the rhombohedral notation). Cleavage of Bi perpendicular to the trigonal direction, leading to the formation of the Bi(111) surface, breaks only the long and ‘weak’ metallic bonds, thus yielding a surface (1) terminated in a bilayer and with no true dangling bonds and (2) quite inert to oxygen adsorption—unlike the other two low-index surfaces of Bi [19].

In order to understand the electronic structure of Bi(111), one must distinguish between the A7-distortion (Peierls-distortion) gap and the spin–orbit gaps. The A7-distortion gap is found around the Fermi level ($E_F$) and separates the conduction and valence bands. Its existence (though not its magnitude) is independent of SOC, inasmuch as it is the A7-distortion gap that makes all A7 materials semimetals. In addition to the A7-distortion gap around $E_F$, ARPES has shown that Bi(111) also has gaps that lie inside the valence band [22]. These are caused solely by the SOC, as predicted by very early DFT calculations of bulk Bi [23]. A Green-function formalism in the tight-binding representation [24] explicitly including the SOC in the Hamiltonian, showed that the energy and wavevector range of the calculated spin–orbit pseudogaps were consistent with the ARPES observations (except for a surface state observed at 3 eV) and that these states have indeed a $p$-character [22]. Although the effect of the SOC on the electronic structure of bulk Bi has long been known from ab initio calculations [25], until now, to our knowledge, the projected band structure of bulk Bi has not been calculated from first principles to corroborate the existence and position of such gaps by comparing with the ARPES experiments in [22]. The projected band-structure of bulk Bi on the SBZ of Bi(111) has so far only been obtained from tight-binding calculations. In fact, even ab initio studies of Bi(111) have used the tight-binding band structure of the bulk to draw conclusions.
about that of Bi(111) and about the effects of the SOC on it [6, 26].

ARPS experiments that followed the above found that Bi(111) has an enhanced density of states at the Fermi energy \( (E_F) \) with respect to bulk Bi [27–29] and displays two metallic branches of surface states along the \( \Gamma \alpha \) line (see figure 2(c)) with energies smaller than 200 meV below \( E_F \) that are gapped at \( M \) [13]. Tight-binding-approximation calculations of the bulk-projected bands [26, 30] and the energy range within which the metallic surface states are found at \( M \) [13] have been the basis for suggesting that these states overlap with the bulk valence band well before reaching the \( M \)-point. If so, then the ARPS experiment in [13] could not have allowed the metallic surface states to be distinguished from bulk states around the zone edge (\( M \)) [6]. Nevertheless, two independent ARPS measurements around the \( M \)-point for the Bi(111) surface of single crystals have reaffirmed the existence of a gap between the two metallic surface states at \( M \) of \( \sim 25 \) meV [11] or \( \sim 37 \) meV [31].

Further attention has been drawn to the calculations of the two surface-state metallic branches of the Bi(111) surface of single crystals and thin films. One aim of the most recent investigations has been to understand the origin of the energy splitting between them [6, 26, 32–35]. However, before going any further with the details, it may be helpful to clarify the terminology to which we shall adhere. Lifting of the Kramers’ degeneracy because of the lack of inversion-symmetry centers and the ubiquitous SOC is known in the literature (including the works on Bi(111) [36]) as Rashba splitting [36–45]. In order to isolate the effect of time reversal symmetry, inversion symmetry, and SOC, we can start by considering systems whose Hamiltonian is conservative and whose geometric structure has an inversion center. Such systems have at least two symmetries. First is time-reversal symmetry, which guarantees that for any electronic state defined by \( \epsilon(k, \uparrow) \), there is a degenerate state with \( \epsilon(-k, \downarrow) = \epsilon(k, \uparrow) \). The second is inversion symmetry, which implies that for each state defined by \( \epsilon(k, \uparrow) \), there is a degenerate state with \( \epsilon(-k, \downarrow) = \epsilon(k, \uparrow) \). Therefore, if a system satisfies both symmetries, all its electronic states are at least doubly degenerate. Specifically, by combining the above two conditions, one gets that \( \epsilon(k, \uparrow)^2 = \epsilon(k, \downarrow)^2 \) for each \( k \)-vector. The latter is known as the Kramers’ degeneracy [6, 46] and it is also often referred as the ‘spin degeneracy’ because the two states have same \( k \)-vector but opposite spin. However, one must be careful with the latter designation because it assumes a framework in which the SOC is neglected. In other words, in reality the spin and the orbital angular momenta are not conserved because of the SOC, and thus their eigenvalues are no longer good quantum numbers. Thus, while it is acceptable to think of spin states for materials for which the SOC is small, it is not the case for Bi. Instead, we must refer to them as the magnetic-quantum-number states. Therefore, the Rashba-type splitting breaks the magnetic quantum number degeneracy \( \pm |m_j| \) rather than the ‘spin degeneracy.’ That being said, the Rashba-type splitting is the lifting of the Kramers’ degeneracy caused by the combination of two factors: [47] (1) an external and perhaps tunable perturbation that breaks the inversion symmetry and leaves the system without an inversion center [48] and (2) the inherent presence of the SOC. Clearly, since the Rashba-splitting exists in part because of the SOC, the effect of inversion asymmetry should be most noticeable for the electronic states of heavy elements such as Bi.

Figure 3 is a qualitative diagram of the effect of the SOC and inversion asymmetry on \( p \)-levels [49]. Figure 3(a) consider the case when the SOC is neglected and therefore all states must be at least doubly degenerate at any \( k \)-point and all six states must be degenerate at \( \Gamma \). In figure 3(b) the SOC is taken into account and leads to a splitting of the \( p_j = \frac{3}{2} \) and \( p_j = \frac{1}{2} \) levels even at \( \Gamma \), but leaves the \( \pm |m_j| \) degeneracy for systems with an inversion center. The latter case, in particular, commonly characterizes the splitting of the bands of semiconductors [46, 50]. On the other hand, in a lattice without any inversion center, the Kramers’ or \( m_j \) degeneracy \( \pm |m_j| \) is lifted at all \( k \)-points (figure 3(c)), except at high symmetry points of the SBZ (see below). Figure 3(c) thus illustrates the Rashba or so-called ‘spin’ splitting.

Why should all states be degenerate at \( M \) regardless of the presence or lack of an inversion symmetry center? In addition
Figure 2. (a) Side view of a Bi slab underlying the Bi(111) surface. A, A', B, B', C, and C' represent the planes defined by each layer, which are perpendicular to the trigonal axis. The sticks connecting the atoms indicate bonds between first NNs (between atoms on planes A' and B, C' and A, or B' and C), in order to highlight the puckered bilayer-like structure of Bi(111). Lines $l_1$, $l_2$ and $l_3$ and the white dots show the large separation between an atom and the atom directly below (or above) along a direction parallel to the normal to Bi(111). (b) Top view of the Bi(111) surface. The experimental and calculated values of $a_h$ are given in table 1. The first light grey (yellow), grey (pink) and dark grey (blue) balls represent the first, second and third layers of Bi(111). (c) Surface Brillouin zone of Bi(111) indicating the high-symmetry points, including the six equivalent $M$ points around $\Gamma$, and the irreducible Brillouin zone (shaded).
Figure 3. Schematic representation of $p$-levels (a) without the influence of the SOC or inversion symmetry (there is a splitting because of possible anisotropy of the crystal; $p_x, p_y, p_z$); (b) with the influence of the SOC and inversion symmetry (SOC splits the states at $\Gamma - j = 1/2$ and $j = 3/2$; and further splits the $j = 3/2$ states into $m_j = \pm 1/2$ and $m_j = \pm 3/2$); (c) with the influence of the SOC and without inversion symmetry (The inversion symmetry together with SOC splits the Kramers’ pairs $m_j = \pm 1/2$ and $m_j = \pm 3/2$ into $m_j = +1/2, m_j = -1/2, m_j = +3/2, m_j = -3/2$: Rashba splitting) [49, 50].

As the latter involves lack of inversion symmetry in the system [47, 48]. These conclusions have been based on calculations performed by adsorbing atomic hydrogen on one surface of a Bi(111) film of 11 BL [12, 26, 35]. Adsorption of H on the ‘bottom’ surface was meant to avoid the interaction between the two surfaces. However since the surfaces are coupled, it is expected that adsorption of hydrogen brings in changes in the electronic structure that are not representative of Bi(111).

In regard to recent ARPES experiments made on the Bi(111) surface of single crystals, the metallic surface states have been clearly detected all the way to $M$ [11, 31]. As mentioned before, it is still argued that the two metallic branches cannot be and have not been measured because they are bound to disappear in the projected bulk band structure before reaching the $M$ point. The arguments for this assumption are that (1) tight-binding calculations predict a narrower gap between the bulk bands of Bi at $M$ (around the Fermi level) and (2) ARPES measurements indicate that the metallic surface states are weakly detectable around $M$ and thus assumed to be highly delocalized. That the metallic branches merge with the bulk bands has received credence because in this way the degeneracy at the $M$ point may seem irrelevant.

ARPES investigations of the two metallic surface-state branches on thin films have been performed using Si(111) as a substrate, which is expected to break the inversion symmetry [33]. Spin-resolved ARPES (SARPES) measurements on thin-films have also shown that the two metallic branches display opposite magnetic-quantum-number polarization [34, 36]. These observations have been considered to support the view that the Rashba-effect lifts Kramers’ degeneracy by giving rise to two branches of opposite magnetic quantum number. Nevertheless, these ARPES and SARPES measurements also displayed the splitting at $M$, even though there is no bulk band for thin films. The authors—aided by band-structure calculations of free-standing Bi(111) films—concluded that the weak interaction between the Bi(111) surface and the Si(111) substrate does not lift the spin degeneracy and the states behave as if the film preserved space-inversion symmetry [26, 33–35], as usual.
symmetry because of the gap at \( M \) [33]. Importantly, the splitting of the spin-textured states at \( M \)—despite the Si substrate breaks inversion symmetry—suggested a violation to translational symmetry because it was assumed that the states are Kramers’ pairs, and as such they must be degenerate at \( M \) [34]. The unexpected splitting at \( M \) was explained again by the overlapping of the metallic states with the bulk band (yet, there is no bulk band for thin films). The supposed overlapping with the bulk band was also linked to the observed evanescence of the spin polarization near \( M \). The metallic states were also assumed to reflect specularly at the film/substrate interface, and create quantum-well states, which in turn yield energetically separated bands at the \( M \)-point [34]. Then, it was proposed that the states become non-spin-polarized near \( M \) simply because only in this region the splitting has no relation to the Rashba effect, i.e. that the states somehow become ‘electronically inversion symmetric’ only near \( M \) [36]. As such, the splitting of the metallic bands is now believed to arise from a combination of an inversion-asymmetry-driven Rashba-splitting (away from \( M \)) and a non-inversion-asymmetry-driven parity-effect splitting (near \( M \)) [36]. However, this is not possible because the inversion symmetry or lack of it cannot be contingent to the region of the SBZ that is being tested.

Regarding slab calculations, the seeming violation to the translational symmetry was attached to the fact that the clean and relaxed Bi(111) slabs (a) do have inversion symmetry [33] or (b) are sufficiently thin so that the two surfaces interact [26, 35], for which they do not represent an ideal semi-infinite surface [35], notwithstanding the splitting at all other intermediate points along the \( \Gamma M \) line has always been attributed to the lack of inversion symmetry (Rashba splitting) [26, 33, 35] and fulfillment of the translation symmetry cannot be contingent to inversion asymmetry. It was found that by increasing the separation between the upper and lower surfaces (i.e. increasing the film thickness), the splitting between the two branches at \( M \), \( \Delta E \), progressively decreases. It was estimated that \( \Delta E \) decreases as \( \frac{1}{N} \), which indicates that the gap \( \Delta E \) decreases by a factor 2 if the number of layers is doubled [35], suggesting that the two bands are degenerate only in the limit of an infinite number of layers and the converged band structure can be attained only for semi-infinite slabs. Thus, the electronic structure of Bi(111) obtained by adsorbing atomic hydrogen on one side of a film of only 11 BL [12, 26, 35], which merges the two metallic states at \( M \), is considered to be representative of that of Bi(111) of single crystals.

Calculations for clean Bi(111) slabs, however, have shown that the splitting between the two metallic bands exists even in the presence of inversion symmetry and that each of the two branches are ‘spin-degenerate.’ It has also been acknowledged that unsupported Bi(111) films, or films supported by weakly-interacting substrates such as Si(111), behave as if they preserved inversion symmetry and for that reason the degeneracy of the surface states is not lifted [34, 35]. Also, according to all available DFT calculations, the two metallic branches of Bi(111) become weakly localized at the surface near \( M \). DFT calculations have in addition reported that the two degenerate states composing either branch are each localized on opposite sides of a film [35]. This is actually not surprising; the time-reversal and the translational symmetries do not demand that the two degenerate states must be localized in a particular spatial region (thus, in practice, depending on the situation, one could use the entanglement of the states of the two sides of the slab or completely ignore the states of one side). Unfortunately, despite these recognitions, some confusion has remained in the literature by labeling this splitting as a Rashba-effect [36–45, 55], which implies lack of inversion symmetry and which is not the case at hand [33–35]. As we shall see, assuming that the measured splitting is of the Rashba-type brings out several inconsistencies about the origin of the two metallic surface-state branches.

First, as mentioned earlier in the discussion of figure 3, lifting of the Kramers’ degeneracy is always attached to the lack of an inversion-symmetry center of the system plus the inherent SOC. However, DFT calculations of slabs with inversion symmetry show the splitting. Second, the Rashba splitting of the Kramers’ degenerate states may not be seen experimentally unless both surfaces of the slab are probed simultaneously since each surface is host to one of the Kramers’ degenerate states. Understandably, all measurements are on a single side of the sample. Third, Rashba-split states must become degenerate at the zone boundary (the \( M \)-point); yet so far all measurements have detected a gap between the two metallic branches at \( M \) [11, 13, 31], as have the various clean slab calculations [12, 33, 35, 36]. Fourth, it has been argued that inversion-symmetric slab calculations do not represent an ideal semi-infinite slab (inversion-asymmetric by definition) [35], and yet the splitting at all other intermediate points along the \( \Gamma M \) line has always been attributed to the lack of inversion symmetry at the surface (Rashba splitting) [26, 33, 35]. Fifth, if an expected degeneracy at \( M \) is apparently not satisfied, then a symmetry is violated (translational symmetry) and there is no reason for that. Sixth, that the two bands become degenerate only in the limit of an infinite number of layers and the converged band structure can be attained only for semi-infinite slabs implies that the translational symmetry is fulfilled only in the limit of infinitely thick slabs. However, that is unphysical because the translational symmetry is only dependent on the in-plane periodicity and should not depend on the slab thickness. The degeneracy at \( M \) is required by the above symmetries regardless of whether or not the state becomes a bulk state and/or it overlaps with the bulk band. This is because both time reversal symmetry and translational symmetry apply to surface states and bulk states, as they do to states localized at the second layer, third layer, and so on. Seventh, it has been accepted that each metallic branch corresponds to two degenerate states [33], that automatically satisfy the degeneracy at \( M \) so there is no violation of translational symmetry.
branches along the $\Gamma M$ direction? (2) Do they overlap with the bulk band, otherwise, why are they hardly detectable around the $M$ point? (3) Can fully relativistic calculations reproduce the surface states arising purely from the SOC? (4) What exactly is the role of the time-reversal, inversion, and translation symmetries in shaping the band structure of Bi(111)? (5) Is the gap at $M$ native to Bi(111) surface (as several experiments indicate [11, 13, 31]) or a long-range effect only noticeable for thin films? (6) How is the splitting or degeneracy at $M$ relevant to the origin of the two metallic bands measured at $M$? (7) Can the coupling between the two surfaces of the slab cause the violation of a symmetry in nature? (8) What is the origin of the magnetic moment (spin texture) of each branch, (8) why the magnetic moment of each single band on a given side of the slab is double-valued and discontinuous at the $M$-point (as shown in figure 4(b) of [34]), (10) How does the Rashba splitting look like in Bi(111)? Can we tune it?

The rest of the manuscript is organized as follows: section 2 describes the computational details of our investigation. Section 3 summarizes our results concerning the geometric structure of bulk Bi and Bi(111) films for various thicknesses. Section 4 contains our results for the band structure of Bi(111) for various thicknesses and examines the various aspects related to the electronic structure of Bi(111) that we have discussed above. Section 5 describes the Rashba splitting of the two metallic branches of Bi(111) achieved by three different perturbations and analyzes how the magnitude of the perturbation correlates with that of the Rashba splitting and spatial localization of the surface states. Finally, section 6 summarizes our main results and the conclusions drawn from them.

2. Computational details

Our investigation is based on fully relativistic periodic-supercell DFT calculations of the electronic structure of bulk Bi and Bi(111). Our calculations are performed within the local density approximation (LDA). For the exchange and correlation terms, we use the Ceperley and Alder [56] functional. The generalized gradient approximation with the Perdew–Burke–Ernzerhof [57] functionals was also applied to obtain the structural properties of bulk Bi and used for comparison with the LDA functional.

The electron-ion interaction is treated within the pseudopotential approach [58]. We use projector-augmented-waves pseudo-potentials [59]. As valence electrons, we have taken into account only the 6$s^2$ and 6$p^3$ electrons of Bi. The Kohn–Sham orbitals are expanded in plane waves (PWs) as implemented in the Vienna ab initio simulation package [60]. The SOC is implemented by the developers as described in [61]. Notice that in this scheme, the inversion symmetry is not assumed within the calculation even if the structure has inversion symmetry so that the Kramers' degeneracy is not assumed but derived. Moreover, all $k$-points in the grid are used in the calculation.

We have used the rhombohedral unit cell of Bi (two atoms per unit cell) to describe the bulk and the hexagonal one [6] (six atoms per unit cell) to describe the Bi(111) surface. The Bi(111) surface was modeled with a slab having one atom per layer ($1 \times 1$ in-plane periodicity) and up to 39 bilayers. In the calculations involving the surface, a vacuum layer of 14 Å separate the periodic images of the slab to avoid interaction between them.

For integrations over the BZ, that of the bulk (in the hexagonal representation) and surface were sampled by 2592 and 432 irreducible $k$-points, respectively, which are selected while ignoring inversion symmetry. The integrations use a Gaussian broadening with a smearing parameter of 0.1 eV. The Kohn–Sham orbitals were expanded in a PW basis set with a maximum kinetic energy of 163 eV. The kinetic energy for the augmentation of charges is cut off at 265.4 eV. The positions of all atoms in the slab were optimized until the forces on each atom and each direction was smaller than $6 \times 10^{-3}$ eV Å$^{-1}$. For this purpose, the conjugate-gradient algorithm has been applied. The total energy for each ionic configuration sampled within the relaxation process was converged up to $2 \times 10^{-5}$ eV.

Preliminary calculations of the lattice parameters, volume and cohesive energy of bulk Bi show that (1) semicore $d$-electrons of Bi influence these properties negligibly; (2) the generalized gradient approximation using the Perdew–Burke–Ernzerhof [57] functional does not outperform LDA since significantly overestimates the length of the basis vectors and thus nearest-neighbor distances and the supercell volume; (3) denser $k$-point grids to sample the bulk BZ do not yield variations in the electronic properties of bulk Bi; (4) doubling the energy cutoff for plane waves does not affect the electronic properties of bulk Bi.

For the calculations concerning the Bi(111) surface, we consider slabs of various thicknesses, with inversion symmetry and without it. The inversion symmetry of a relaxed Bi(111) slab is broken by (1) placing hydrogen at 2 Å of Bi, as in [26], and at 1.86 Å, which is the distance resulting after relaxation; (2) an inter-bilayer perturbation—achieved by increasing by 0.3 Å the distance between the first two bilayers of one side of the slab; and (3) an intra-bilayer perturbation—achieved by increasing by 0.1 Å the distance between the two layers forming the first bilayer of one side of the slab. The XCrySDen software [62] is used to obtain a schematic representation of the structure of the bulk Bi and Bi(111).

3. Geometric structure of bulk Bi and clean Bi(111)

Results of our DFT calculations for the parameters characterizing the rhombohedral unit cell of bulk Bi and the comparison with previous results as well as with experiments are shown in table 1. The slight differences in values obtained by DFT can be attributed to the difference in the exchange-correlation functional that is used. Overall, there is good agreement among the techniques used to obtain the parameters.

The Bi(111) surface is a triangular lattice (see figure 2(b)) perpendicular to the trigonal axis. There are four parameters in
Table 1. Structural rhombohedral and hexagonal parameters describing bulk Bi (see text and figure 1); volume of the rhombohedral supercell (V), first and second nearest-neighbor distances (d_{NN} and d_{2NN}); the two bulk-like vertical inter-layer separations along the trigonal axis (d_{ij} and d_{jk}). The results of the present work are compared with measurements at liquid helium and room temperature (T), as well as with previous calculations.

| Calculations | This work | [65] | Experiments | [66] | [67] | [68] | [69] | [70] | [71] |
|--------------|-----------|------|-------------|------|------|------|------|------|------|
|              | GGA       | LDA  | LDA         | T = 4.2 K | T = 4.2 K | T = 300 K |
| \(a_{0}\) (Å) | 4.85      | 4.68 | 4.69        | 4.75  | 4.75 | —      | 4.73 | —    | 4.71–4.77 |
| \(c_{0}\) (Å) | 57.30     | 58.05| 57.57       | 57.27 | 57.24| —      | 57.31| —    | 57.20–58.12 |
| \(d_{ij}/d_{2j}\) | 0.233    | 0.237| 0.234       | 0.233 | 0.237| 0.234 | 0.234| 0.235| 0.237 |
| \(\Delta d_{ij}/d_{2j}\) | 0.87     | 0.90 | 0.88        | 0.87  | 0.90 | 0.88  | 0.88 | 0.89 | 0.90 |
| \(V(\text{Å}^3)\) | 75.82     | 69.24| 69.06       | 71.02 | 70.78| —      | 70.05| —    | 70.55–73.30 |
| \(d_{ij}\) (Å) | 4.65      | 4.54 | 4.52        | 4.55  | 4.55 | —      | 4.53 | —    | 4.57–4.63 |
| \(d_{jk}\) (Å) | 12.12     | 11.63| 11.71       | 11.87 | 11.86| 11.80  | 11.81| 11.84| 11.69–11.84 |
| \(c_{ijk}\) (Å) | 2.61      | 2.56 | 2.59        | 2.61  | 2.61 | —      | 2.60 | —    | 2.56 |
| \(d_{ij}^{NN}\) (Å) | 3.13      | 3.08 | 3.05        | 3.06  | 3.11 | —      | 3.06 | —    | 3.11–3.15 |
| \(d_{NN}\) (Å) | 3.63      | 3.46 | 3.50        | 3.55  | 3.48 | —      | 3.51 | —    | 3.47–3.52 |
| \(d_{jk}\) (Å) | 1.608     | 1.615| 1.58        | 1.58  | 1.67 | 1.59  | 1.59 | 1.62 | 1.66 |
| \(d_{jk}^{NN}\) (Å) | 2.432     | 2.262| 2.33        | 2.38  | 2.28 | 2.34  | 2.35 | 2.33 | 2.28 |

Table 2. Converged changes (%) in the interlayer spacing, \(\Delta d_{ij,i+1}\), of Bi(111) films with respect to those in bulk Bi obtained in this work, previous calculations and low-energy electron diffraction experiments (LEED). The experimental values at \(T = 0\) K are values obtained from an extrapolation and linear fit based on measurements obtained at \(T = 140\) K [32].

| Calculations | 10–39 BLs | 6–12 BLs | 7 BLs | 6 BLs | 6 BLs | LEED Experiment [32] |
|--------------|-----------|----------|-------|-------|-------|----------------------|
|              | LDA (This work) | LDA [17] | LDA [32] | LDA [35] | GGA [72] | T = 140 K | T = 0 K |
| \(\Delta d_{ij}\) (%) | −1.57 | −1.62–1.56 | +0.6 | +0.9 | −0.83 | +0.5 ± 1.1 | +1.2 ± 2.3 |
| \(\Delta d_{2ij}\) (%) | +1.31 | 1.37/1.32 | +6.2 | +6.5 | +3.13 | +1.9 ± 0.8 | +2.6 ± 1.7 |
| \(\Delta d_{kk}\) (%) | −0.87 | −0.87/–0.87 | — | — | — | 0.0 ± 1.1 | — |
| \(\Delta d_{kk}\) (%) | +0.13 | 0.13/0.13 | — | — | — | — | — |
| \(\Delta d_{kk}\) (%) | −0.31 | −0.25/–0.31 | — | — | — | — | — |
| \(\Delta d_{kk}\) (%) | 0.00 | — | — | — | — | — | — |

A7-materials characterizing the bilayered structure that facilitate the analysis of the properties of bulk Bi and, in particular, of Bi(111). These parameters are the distance between first nearest neighbors \(d_{NN}\), the distance between second nearest neighbors \(d_{2NN}\), the intra-bilayer distance \(d_{ij}\), and the inter-bilayer distance \(d_{jk}\). The bilayered structure of Bi underlying the Bi(111) surface is shown in figure 2(a). As shown by the color code and the labels, the bilayered structure repeats itself only every six planes. The vertical spacing between the layers forming a bilayer, the \(\text{intra-bilayer distance}\), is denoted by \(d_{ij} = d_{12}, d_{23}\), etc and is related to the strong, short and covalent bonds existing among first nearest neighbors. In turn, the vertical spacing between two bilayers, the \(\text{inter-bilayer distance}\), is denoted by \(d_{jk} = d_{23}, d_{34}\), etc and is related to the weak, long and metallic bonds existing among second nearest neighbors.

Comparison of the results of the structural parameters for Bi(111) with those in bulk, as obtained by our DFT calculations, are summarized in table 2. Here it will be important to mention only that the changes in \(d_{ij}\) and \(d_{jk}\) within the four topmost bilayers are well converged for slabs of ten or more bilayers (up to 39 BLs). Moreover, the bulk-like interlayer spacings is recovered around the fifth bilayer, in the sense that at such distance from the surface, changes of \(d_{ij}\) and \(d_{jk}\) with respect to the bulk values are less than 0.1%.

4. Electronic structure of bulk Bi and Bi(111) thin films

4.1. Ab initio projected band-structure of bulk Bi and the emergence of the A7-distortion and SOC surface states

Here we analyze the \(ab\) \(initio\) and fully relativistic projected band-structure of bulk Bi on the SBZ of Bi(111) in order to introduce and discuss all the surface states known for Bi(111) [22], those in the A7-distortion gap and those arising only because of the SOC. Figure 4(top) shows the projected band-structure of bulk Bi, which includes SOC, projected along the \(\Gamma M\) and \(\Gamma K\) directions of the SBZ of Bi(111) (see figure 2(c)) and in the energy range within which the various gaps and surface states are expected [22, 23]. It shows the gap between the conduction and valence bands caused by the A7-distortion around \(E_F\)—the A7-distortion gap—and the three gaps arising once the SOC is included, as we shall see below. We label the latter gaps as SOC gap 1, SOC gap 2, and SOC
A7-distortion gap appearing along high-symmetry lines of the SBZ with and without SOC, indicating (Bottom) Band-structure of a 39-bilayer slab along the high-symmetry lines of the surface Brillouin zone, \( \Gamma M \) and \( \Gamma K \).

Figure 4. (Top) Projected band-structure of bulk Bi on the SBZ of Bi(111) from first-principles and including the SOC along the high-symmetry lines of the surface Brillouin zone, \( \Gamma M \) and \( \Gamma K \).

Turning to the surface states, figure 4(bottom) shows that without SOC we only have the surface states in the A7-distortion gap (blue arrows). Once SOC is included in the calculation (black lines) surface states emerge as all SOC gaps, as shown in the figure. Importantly, energies of the above SOC surface states are in agreement with ARPES measurements [22]. Comparison between the band structure with (black lines) and without SOC (green lines) reveals that two metallic surface-state branches exist in the A7-distortion gap along \( \Gamma M \) independently from the SOC (here the word ‘branch’ denotes that the displayed curves do not represent a single state). These states are thus the A7-distortion metallic surface states. Also, it will be important for later discussion to note that in figure 4(bottom) the surface-states branches that peel off from the upper bulk bands (those in the A7-distortion gap and the SOC gap 1) come in pairs, whereas the surface-states that peel off from the lower bulk band (those in the SOC gap 2 and 3) are single branches.

Before going into details on the A7-distortion metallic surface states, which are the focus of this investigation, let us identify the features that are intrinsic to the geometric structure of Bi (the A7-distortion) and how the SOC shapes them. Indeed, figure 4(bottom) reveals that: (1) the splitting of the A7-distortion metallic surface states into two branches from \( \Gamma \) to \( \frac{1}{2} M \) arises only when the SOC is included. Yet, even though the SOC largely splits them starting at small \( k \)-points, they do split into two branches even without SOC, although starting at larger \( k \)-points. (2) While from \( \Gamma \) to \( \frac{1}{2} M \) the SOC shifts up the energy of the high-energy metallic branch (H1,2) and slightly down that of the low-energy metallic branch (L1,2), from \( \frac{1}{2} M \) to \( M \) the SOC increases the energy of both branches. The latter is the reason for which they do not overlap with the bulk band. (3) The role of the SOC to enhance the splitting between the two metallic branches consistently decreases from \( \Gamma \) to \( M \). In fact, for the 15.5 nm film (39 BL), the SOC rather decreases the splitting within the \( k \)-region from \( \frac{1}{2} M \) to \( M \). For example, we find that for the 15.5 nm film the splitting between the two metallic branches at \( M \) without SOC is 90 meV, whereas that with SOC is only 34 meV.

Note that for this inversion-symmetric film of 15.5 nm, with and without SOC, the metallic branches are doubly degenerate. Namely, although the metallic branches appear separated along \( \Gamma M \), each metallic branch in figure 4(bottom) is composed of two degenerate states: Kramers’ pairs. Thus, the SOC does not lift the Kramers’ degeneracy anywhere inside the SBZ. Because the film is inversion-symmetric and the branches are doubly degenerate, that means that their splitting arises without requiring inversion asymmetry and without breaking the Kramers’ degeneracy, which demonstrates that the splitting of the metallic surface states is not of the Rashba type. Moreover, it is clear for the same reasons that, despite the 34 meV splitting, the degeneracy at \( M \) is guaranteed for inversion-symmetric films because \( M \) is a time-reversal invariant point regardless of the presence of inversion symmetry. Since each A7-distortion metallic branch is doubly degenerate,
Figure 5. Band-structure of Bi(111) around $E_F$ for slabs of various thicknesses (7, 12, 21, 27, and 39 BLs) along $\Gamma\text{-}M$ comparing the calculated A7-distortion metallic surface-states against the ARPES measurements of Hengsberger et al [13] (Exp. 1) and Ast and Höcht [31] (Exp. 2) made on the Bi(111) surface single crystals and against the calculated projected band structure of bulk Bi on the SBZ of Bi(111) (black lines).

there are actually four metallic surface-state states along $\Gamma\text{-}M$. Branch $H_{1,2}$ is composed of Kramers’ degenerate states $H_1$ and $H_2$; likewise, $L_{1,2}$ is composed of Kramers’ degenerate states $L_1$ and $L_2$. Importantly, at least for inversion-symmetric films, $H_1$ and $L_1$ are primarily localized on one side of the film, whereas $H_2$ and $L_1$ are primarily localized on the other side.

In summary, while the SOC tunes the A7-distortion gap, widely splits the A7-distortion metallic surface-state branches near $\Gamma$, and reshapes their dispersion along $\Gamma\text{-}M$, the origin of the A7-distortion gap, the metallic surface states in it, and their splitting for most of the $\Gamma\text{-}M$ direction does not lie in the SOC. Moreover, the SOC does not lift the Kramers’ degeneracy and, therefore, despite the splitting of the metallic branches at $M$, the required degeneracy at $M$ is satisfied.

4.2. Splitting and delocalization of the A7-distortion metallic surface states as a function of film thickness

In figure 5 we examine the evolution and convergence of the dispersion of the A7-distortion metallic branches as a function of film thickness by comparing the results for inversion-symmetric slabs of 7, 12, 21, 27, and 39 BL. We find that 7 BL are enough to converge the dispersion of the two metallic branches from $\Gamma$ to $\frac{\pi}{7}M$, whereas 12 BL are sufficient for convergence of the dispersion of the two metallic branches from $\Gamma$ to $\frac{\pi}{12}M$. We note that the differences between the two dispersion curves—7 BL and 12 BL—are minimal from $\frac{\pi}{12}M$ to $\frac{\pi}{7}M$. This shows that the coupling between the two surfaces of the 7 BL film plays a small but increasing role in the splitting from $\frac{\pi}{12}M$ to $\frac{\pi}{7}M$ that is taken away by using the 12 BL film. Next, we can see that 27 BL suffice for converging the dispersion of the two metallic branches from $\Gamma$ to $\frac{\pi}{27}M$. This shows that the splitting between the two branches from $\Gamma$ to $\frac{\pi}{27}M$ for 39 BL arises in the presence of inversion symmetry, is independent of the coupling between the two surfaces, thus independent of thickness of the slab, and hence intrinsic to Bi(111). Finally, we can see that from $\frac{\pi}{10}M$ to $M$ the dispersion is not converged for 27 BL.

Importantly, figure 5 shows that there is good agreement between our calculations for an inversion-symmetric film consisting of 39 BL and the ARPES measurements of Hengsberger et al [13], Ast and Höcht (figure 1(a) of [31]), and Ohkubo et al [11], for the Bi(111) surface of single crystals regarding both the dispersion and the fact that the two metallic branches do not become degenerate at $M$. This agreement supports that, should the crystal be inversion-asymmetric because of the nature and limitations of the experiment, (1) the splitting of the metallic branches along $\Gamma\text{-}M$ takes place regardless of inversion symmetry and (2) the splitting at $M$ between the two metallic branches comes up not only for simulated inversion-symmetric thin films but also for the Bi(111) surface of single crystals. As mentioned earlier in section 4.2, we find a splitting of $\sim$34 meV at $M$ for the 39 BL film. This value is in good agreement with ARPES measurements on the Bi(111) surface of single crystals, which report a splitting of $\sim$25 meV [11] and $\sim$37 meV [31]. We shall denote the energy splitting at $M$ as $\Delta E$ and discuss it in section 4.4.

Figure 5 also displays the projected band structure of bulk Bi (black lines) to closely analyze the behaviour of the A7-distortion metallic branches with respect to the projected bulk bands. Our fully relativistic DFT calculations indicate that the width of the A7-distortion gap of bulk Bi around $M$ is $\sim$120 meV and spans from 15 to 135 meV below $E_F$, in stark contrast to the $\sim$20–30 meV gap predicted by tight-binding calculations (spanning at most from 20 to 50 meV below $E_F$) [26, 31]. In addition, contrary to the belief that the surface states of the Bi(111) surface of single crystals overlap with the bulk band (based on tight-binding calculations [26, 30]) [6, 31, 36], our DFT calculations for slabs thinner than 8 nm ($\sim$21 BL) show that the metallic branches do not overlap with the bulk bands at or in the vicinity of $M$. Moreover, there is no indication that the metallic branches tend to cross the bulk band; on the contrary, from $\frac{\pi}{10}M$ to $M$, they tend to slowly move away from the bulk bands. Thus, our results in figure 5 also indicate that ARPES measurements around $M$ [11, 13, 31] are within the gap and are not surface resonances (i.e. hybridization of surface states with quantum well states [33]) but surface states. Our calculations, however, do confirm that near $M$ the metallic branches are increasingly delocalized among subsurface and bulk-like layers. This explains why ARPES measurements of the metallic branches around $M$ render a rather low intensity.

4.3. Evolution of the metallic branches along the $\Gamma\text{-}M$ as a function of the SOC strength: comparison of inversion-symmetric and -asymmetric films

We have seen that each of the two degenerate states of each branch is localized on one or the other surface of the slab (for the most part along $\Gamma\text{-}M$, as we shall see later). However,
Figure 6. Evolution of the A7-distortion metallic branches (red lines) of a 24 BL Bi(111) as a function of the strength of the SOC for films (Top) with inversion and (Bottom) without inversion symmetry. The black lines correspond all other states of the band-structure. The diameter of the dots represent how much the metallic band is localized on a single surface of the film. In the case of the slab without inversion symmetry, the red dots correspond to the unperturbed surface (see text).

one should keep in mind that only one surface is investigated experimentally. In order to understand the effect of the SOC and the inversion symmetry on the band-structure of Bi(111) films, and compare with experimental observations, we consider the effect of the SOC on a slab with inversion symmetry (figure 6(top panel)) and another with without inversion symmetry (figure 6(bottom panel)).

Figure 6(top panel) shows the evolution of the splitting between and degree of localization of the metallic branches of a 24 BL Bi(111) with inversion symmetry as a function of the strength of the SOC. In order to do that, we must distinguish between the two surfaces of the slab. We will refer to them as the ‘top’ and ‘bottom’ surface. The states located on, say, the top surface are then highlighted with red dots whose diameter
represents to what extent the state is localized on the surface. The other Kramers’ state at each k-point (located in the bottom surface) cannot be appreciated because it is degenerate with the bands located on the top surface. As expected from available measurements and previous calculations [31, 35], both metallic branches bear some localization weight on any of the two surfaces of a film. As expected from section 4.1, the splitting between the two metallic branches increases almost all along $\Gamma M$ and decreases near $M$ with increasing strength of the SOC. Figure 6(top panel) also shows that with increasing strength of the SOC, the localization of the metallic branches on the topmost bilayer slightly decreases around the $\Gamma$ point and slightly increases toward the zone boundaries $M$ and $K$. Also, figure 6(top panel) shows that the localization of the metallic states at the surface BL reduces as $k$ approaches $M$.

Let us now consider the effect of the SOC in a slab without inversion symmetry. Figure 6(bottom panel) displays the evolution of the metallic branches of a 24 BL Bi(111) without inversion symmetry as a function of the strength of the SOC. In this case, the symmetry is broken by increasing the inter-bilayer distance $d_{jk}$ between the bottom bilayer and the rest of the film. Again, the red dots indicate how much the metallic states are localized on the top bilayer. By comparing figures 6(top) and (bottom), one can see that the dispersion of the bands on the top bilayer and their dependency on the strength of the SOC are nearly the same with and without inversion symmetry. In addition, can see that, without inversion symmetry, the metallic branches display extra splittings. In the absence of the SOC, the branches resulting from the extra splitting are still doubly degenerate. Hence, inversion asymmetry alone does not lift the Kramers’ degeneracy either. However, with broken symmetry and once the SOC is included, the obtained splitting finally lifts the Kramers’ degeneracy and reveals the four states composing the two metallic branches. That is indeed the lifting of the Kramers’ degeneracy, the ‘spin’ or Rashba splitting. Still, Kramers’ pairs are on opposite sides of the slab. Therefore, once again, the splitting found for the top bilayer bears no relation to the lack of inversion symmetry or the lifting of the Kramers’ degeneracy (Rashba effect). Importantly, figure 6(bottom) shows that in this inversion-asymmetric film, the resulting Rashba splitting does satisfy the translational-symmetry-required degeneracy at $M$. We shall come back to analyze inversion-asymmetric slabs when we discuss the Rashba splitting in Bi(111) in section 5.

4.4. Origin of the two A7-distortion metallic surface-state branches along $\Gamma M$

In figure 5 we have two A7-distortion metallic branches, $H_{1,2}$ and $L_{1,2}$, that display a k-dependent splitting between them, although each branch corresponds to two degenerate states. So the splitting in question here is not that of Kramers’ pairs. Note also that the splitting cannot be entirely ascribable to SOC, as demonstrated in figure 4(bottom). Thus, based on the fact that the metallic surface states exist regardless of the SOC and/or the lack of inversion symmetry, we can start by recognizing that the surface states of Bi(111) are first of all a pair of Shockley states arising because of the abrupt termination of the periodic potential for the p-states of Bi. Yet, why does the surface display two branches and not one? The answer is not one reason but at least three: SOC, anisotropy of the structure, and coupling between the two surfaces of the film. Moreover, the importance of each varies depending on the k-region along $\Gamma M$. Below, we shall discuss each $k$-region in detail.

As shown in section 4.1, the splitting of the metallic branches from $\Gamma$ to $\frac{1}{2}M$ arises only once the SOC is included. Still, what does the SOC split then? The splitting does not lift the Kramers’ degeneracy and yet it is caused by the SOC, thus we propose that from $\Gamma$ to $\frac{1}{2}M$, we have a $m_j = \pm 1/2$—$m_j = \pm 3/2$ splitting, as described in figure 3. In the following, we provide some arguments and indirect evidence that the two metallic branches (accounting for a total of four states) have a $m_j = \pm 1/2$ and $m_j = \pm 3/2$ character from $\Gamma$ to $\frac{1}{2}M$.

The bulk states shown in figure 4(top) appear as purely p-states of bulk Bi if projected on spherical harmonics (s-states have much lower energy $\sim 8$ eV below $E_F$) [25]. However, it would be necessary to project the wavefunction onto the eigenfunctions of the $J^2$ and $J_z$ operators to obtain their main total-angular-momentum and magnetic-quantum-number character, which is a computational task that we could not afford. Nonetheless, by analyzing the splittings of the bulk bands along $\Gamma\Gamma$ and $\Gamma XX$ occurring upon inclusion of the SOC [25] and comparing them with those sketched in figure 3(b), it is possible to identify that along those directions the lower single p-band must have a $j = 1/2$ character and the upper pair of p-bands must have a $j = 3/2$ character. The latter in turn must split into $m_j = \pm 1/2$- and $m_j = \pm 3/2$-states. Therefore, one may conclude that in figure 4(top), the projected bulk bands around $\Gamma$ of the SBZ that span from $\sim 2$ to $\sim 3$ eV correspond to the $j = 1/2$-states, whereas the upper occupied bands (those up to $\sim 1.5$ eV below $E_F$) correspond to $j = 3/2$-states that split into $m_j = \pm 1/2$- and $m_j = \pm 3/2$-states (even though in other regions of the BZ the energy ordering of the $j = 1/2$ and $j = 3/2$ states may be swapped, in which case the two contributions would also overlap [23]). One can thus speculate that around $\Gamma$ the bulk bands up to $\sim 1.5$ eV below $E_F$ must be $j = 3/2$ states of a $m_j = \pm 1/2$ and $m_j = \pm 3/2$ character that overlap energetically for the most part once projected on the SBZ, whereas those below $\sim 1.5$ eV have a $j = 1/2$ character. In line with these arguments, note that the surface states that peel off from the lower band ($j = 1/2$) are a single branch while the surface states that peel off from the upper band (overlapping of $m_j = \pm 1/2$ and $m_j = \pm 3/2$) come in pairs (see figure 4(bottom)). Among the branches that come in pairs there is the pair of A7 surface-state metallic branches we are interested in. Thus, the surface states derived from these three major bands ($j = 1/2$, $m_j = \pm 1/2$, and $m_j = \pm 3/2$), which do not overlap with anything else, might acquire a relatively pure $j, m_j$ character.

The $m_j = \pm 1/2$–$m_j = \pm 3/2$ splitting is known in quasi-2D semiconductors as the splitting between the light-hole (LH) and heavy hole (HH) branches (the LH-HH splitting), in reference to the effective masses attached to carriers because of their dispersion [46, 50]. In the case of the A7-distortion metallic branches of Bi(111), the high-energy metallic branch, $H_{1,2}$,
has a larger dispersion than the low-energy branch, $L_{1,2}$, as shown in figure 5. We would thus say that the $H_{1,2}$ branch (light) corresponds to carriers lighter than the $L_{1,2}$ branch (heavy). In line with this analogy (with the band-structure of semiconductors), we also consider the anisotropy of the charge density to determine the character of the metallic branches. Namely, in quasi-2D semiconductors it is known that the charge density associated to light and heavy branches is anisotropic with respect to the direction perpendicular to the surface. Specifically, the in-plane ($xy$) charge density is larger for a HH-branch ($m_j = \pm 1/2$) than for a LH-branch ($m_j = \pm 3/2$), whereas the charge density along the direction perpendicular to the surface ($z$) is larger for a LH-branch than for a HH-branch [63]. To analyze the latter features in the A7-distortion metallic branches, figure 7 displays our calculation of the evolution along $\Gamma M$ of the $p_y$ (left) and $p_z$ (right) local density of states corresponding to low-energy metallic branch (heavy) and the high-energy branch (light) for the 39 BL film. As expected, figure 7 shows that, especially from $\Gamma$ to $1/2 M$, the projected LDOS $p_y$ attached to $L_{1,2}$ (heavy) is significantly larger than that of $H_{1,2}$ (light), whereas the projected LDOS $p_z$ attached to $H_{1,2}$ (light) is significantly larger than that of $L_{1,2}$ (heavy).

Till now, we have provided similarities between the two split metallic branches of Bi(111) and the LH-HH splitting found in semiconductors [50], suggesting that high- and low-energy branches have $m_j = \pm 1/2$ and $m_j = \pm 3/2$ character, respectively. We now turn to analyze the local magnetic moment of the two metallic branches of Bi(111), which can gauge the $m_j = \pm 1/2$ and $m_j = \pm 3/2$ character of the two metallic branches of Bi(111). Namely, for a quantization axis normal to the surface ($z$), states with a predominant $m_j = \pm 1/2$-character are expected to have a larger contribution to the in-plane magnetization than states with a predominant $m_j = \pm 3/2$-character. Of course, in the case of $L_{1,2}$ and $H_{1,2}$, the total magnetic moment is zero for both because of Kramers’ degeneracy, but we shall focus only on their state that has positive magnetic moment: H1 and L1. Also, we analyze the $m_y$ magnetization because it dominates the in-plane magnetization along $\Gamma M$. Figure 8 displays our calculated evolution of the $m_y$ magnetization of L1 and H1 as projected onto the surface bilayer of the 39 BL film along $\Gamma M$. Indeed, as expected and in agreement with SARPES measurements [34], figure 8 shows that L1 (heavy) has a larger in-plane magnetization all along $\Gamma M$ than H1 (light), reasserting that the low-energy branch has a predominant $m_j = \pm 1/2$-character and the high-energy branch has a predominant $m_j = \pm 3/2$-character. In summary, our results suggest that the splitting between the two metallic branches from $\Gamma$ to $1/2 M$ corresponds to a $m_j = \pm 1/2$–$m_j = \pm 3/2$ splitting caused by the SOC, but in which the Kramers’ pairs are still degenerate.

We now turn to the splitting from $1/2 M$ to $\pi M$. Note that the value $k = \frac{\pi}{10} M$ is chosen in relation to our 39 BL slab because up to that $k$-point the dispersion of its A7-distortion metallic branches is convergent. In section 4.1, we found that the splitting between the two branches from $1/2 M$ to $\pi M$ is not caused by the SOC because it is clearly established even without SOC (figure 4(b)), but is considerably enhanced by it. We also found in section 4.2 that the splitting within this $k$-region...
cannot be attributed to the coupling between the two surfaces of the film because the splitting remains the same as the number of bilayers increases from 21 to 39 BL. However, figure 7 shows that within this region the $H_{1,2}$ branch is still predominantly $p_z$ and the $L_{1,2}$ is predominantly $p_{x,y}$. Thus, we conclude that, even though the SOC effect is not as important as it is near the $\Gamma$ point, these $p$-states inherently split because the material is anisotropic (because of its layered structure) and the potential landscape for in-plane states is different from that for out-of-plane states. In-plane states are mostly $p_z$ states and have lower energy because they are more involved in the covalent bonding among Bi atoms. Out-of-plane states are mostly $p_z$ states and have higher energy because they are less involved in covalent bonding. Thus, the splitting of $p$-states at low symmetry $k$-points near the zone boundaries is expected for the $p$-states of any layered material even if the SOC is irrelevant and even for the bulk, as suggested in figure 3.

Lastly, we turn to the splitting near $M$. Our calculations in section 4.2 showed that it remains all the way up to $M$ even for films of 39 BL. Importantly, since each branch is doubly-degenerate, as demanded by the time-reversal and translational symmetries, one must note that there is no $a$ priori reason to expect them to meet at $M$. Nevertheless, it is a matter of concern because, as shown in figure 5 and in agreement with previous calculations [33, 35], the splitting between the two branches is dependent on the thickness of the slab: the thicker the slab, the smaller the splitting. This hints that the electronic structure of Bi(111) has not converged even at 27 BL or perhaps not even at 39 BL, which to our knowledge seems to be unique to Bi. Thus, the question becomes, is the splitting at $M$ native to Bi(111) or a long-range effect only noticeable for thin films? To answer this question we must analyze the splitting at $M$ on the same footing we used to analyze the splitting along the rest of the $\overline{\Gamma M}$ direction. Thus, we first note that without SOC the splitting between the branches reaches its largest value precisely between $\frac{\pi}{2}M$ and $M$. Yet once the SOC is applied, the splitting is quenched in this $k$-region. Therefore, from $\frac{\pi}{2}M$ and $M$, the SOC is uniting the metallic branches or at least not splitting them. Also, within this region, both branches have almost the same contribution from $p_z$ and $p_{x,y}$ states, as shown in figure 7, for which the anisotropy of the Bi–Bi bonding plays a marginal role. Importantly, within this region the states become significantly delocalized among deeper bulk-like layers. Thus, as noted by several authors [12, 16, 35, 36], the splitting is influenced by the coupling between the opposite surfaces of a slab. However, awareness of this interaction does not inform whether the gap will vanish or converge to a non-zero value at a given thickness.

Figure 5 shows that the splitting $\Delta E$ between the two metallic branches decreases as a function of thickness, in agreement with previous calculations [33, 35]. Our limited data suggests that the bands might meet at $N_B \sim 72$ BLs (figure 9) and then behave differently for thicker slabs if the two sides of the film remain coupled. We did not pursue the calculation for $N_B \sim 72$ BL for several reasons. First of all, it is an extremely demanding calculation. But more importantly, we have shown already that the key aspects to understand the origin and the ARPES measurements of the metallic states of Bi(111) for most of the $\overline{\Gamma M}$ direction are not buried in the behavior of the metallic states at $M$. A non-linear regression fitting of our 11 data points (figure 9) indicates that, although $\Delta E$ varies as $\frac{1}{N_B}$ within $N_B = 7$ and $N_B = 39$, as found in [35], there is an additional constant factor $C = 40 \text{ meV}$ ($\Delta E = \frac{B}{N_B} - C$) that invalidates that the two bands are degenerate only in the limit of an infinite number of layers. Our fitting in figure 9 is of course limited to the range of the available data, but it certainly shows that an extrapolation based on a limited amount of data points may not be appropriate to predict the behavior of much thicker slabs. Importantly, although our calculations cannot discern whether the splitting at $M$ is entirely or only partially caused by the coupling between the two surfaces of the film, we must remind the reader that ARPES measurements confirm the splitting at $M$ even for single crystals [11, 13, 31], which indicates that either the splitting is not entirely related to the coupling between the two surfaces caused by small thickness of the film or, perhaps more interestingly, the coupling remains independently of thickness.

The thickness, if any, at which the two branches become decoupled still has physical significance because one would like to know the converged electronic structure of Bi(111), particularly if such thickness is finite. On the other hand, regardless of whether the splitting at $M$ is intrinsic to Bi(111) or not, it is important to know the origin of the coupling between the two surfaces from the point of view of the charge localization that allows the two surfaces to remain coupled at distances beyond 15 nm. Analysis of the spatial localization of the two metallic branches, $H_{1,2}$ and $L_{1,2}$, shows that, for the 39 BL slab, precisely from $\frac{\pi}{2}M$ and $M$ (the $k$-region within which the dispersion in not converged at this thickness), each of the four metallic states (H1, H2, L1, and L2) is no longer confined to a single surface but bears an appreciable contribution on both surfaces of the slab. Thus, within this small $k$-region, the metallic states are not only delocalized among deeper layers down to vanishingly small values but actually gain an appreciable localization at the other surface. Therefore, the spatial overlapping of...
Figure 10. A7 metallic surface states H1 and L2 projected on the ‘top’ surface of an inversion-symmetric 39 BL film indicating the polarization of their magnetic moment along ΓM: red for positive (‘up’) and black for negative (‘down’) magnetic moment.

these states afforded by their relocalization on both surfaces as \( k \) approaches \( M \) may cause a hybridization (H1 with L1 and H2 with L2) that perpetuates the gap between the two metallic branches. Another consequence of the relocalization of the metallic states near \( M \) is discussed in the following section.

4.5. Vanishing of the magnetic moment near \( M \) and discontinuity avoidance

Spin-resolved ARPES measurements on ultra-thin films (7 BL [36] and 10 BL [34]) supported on Si(111) have shown that the two A7-distortion metallic states display opposite magnetic-quantum-number polarization and this has been confirmed by DFT calculations. Thus, even though the two-fold degeneracy at \( M \) is strictly satisfied, there is an apparent nonphysical discontinuity and indetermination of the magnetic moment on either side of the film (top or bottom) at the \( M \)-point for each metallic band displayed in figure 4(b) of [34]. To appreciate this magnetic discontinuity and indetermination at \( M \), in figure 10 we display our calculated dispersion of the A7-distortion surface states H1 and L2 from \( Γ \) to \( M \) and from \( M \) to \( Γ \) of the contiguous SBZ for the 39 BL film. Note that H1 and L2 are localized on one side of the slab, say the ‘top bilayer’, and have opposite magnetic moment, as indicated by the black and red colors and arrows. Thus, one can see that the calculated magnetic moment of H1 discontinuously switches from positive to negative, whereas that of L2 switches from negative to positive at the zone boundary, \( M \). Spin-resolved ARPES measurements on Bi(111) ultra-thin films have detected that the polarization switches [36] but have not attested the discontinuity, instead SARPES measurements indicate that the net spin polarization of each A7-distortion state vanishes near \( M \) [34, 36].

The question that arises is: why does the spin polarization vanish? It is indeed in part because the A7-distortion surface states become increasingly delocalized as they approach \( M \), but that does not entirely explain it because while the spin polarization vanishes, the surface states are still fairly detectable by ARPES. In this section, we shall resolve the nonphysical discontinuity and indetermination of the magnetic moment to put our understanding of the A7 metallic branches at the \( M \) point on firm grounds, and at the same time show that the solution to this issue provides insight into the vanishing of the spin polarization obtained from the above spin-resolved ARPES measurements. Namely, no matter how thin or how thick the sample is, the magnetization vanishes because the Kramers’ pairs, which for the most part of the SBZ are localized on opposite sides of the slab, gradually relocalse on both surfaces as \( k \) approaches \( M \), thus cancelling each other’s magnetization. To visualize this phenomenon, in figure 11 we have extracted from our calculations the evolution of the magnetic moment and spatial localization (‘top’ or ‘bottom’ bilayer) of the four A7-distortion surface states H1, H2, L1, and L2 states as they disperse from \( M \) to \( Γ \) for the inversion-symmetric film of 39 BL. First, we note in figure 11 that, in
agreement with ARPES measurements, the localization of the A7-distortion surface states on the ‘top’ and ‘bottom’ bilayers become increasingly smaller near \( M \) and so does the magnetic moment. Next, note that near \( M \) each surface (‘top’ and ‘bottom’) is host to all four A7 surface states: H1, H2, L1, and L2. Since experimentally only one surface is sampled, we shall concentrate the discussion only on the ‘top’ bilayer, though the same arguments apply for the ‘bottom’ bilayer. Thus, as shown in figure 11, the top bilayer hosts mainly the H1 and L2 states for the most part along the \( \Gamma M \) line. The high-energy band H1 has ‘up’ polarization (figure 11(a)), whereas the low-energy band L2 has ‘down’ polarization (figure 11(b)). Most importantly, note that, close to \( M \), it is no longer true that the top bilayer gets magnetic-moment contributions only from the H1 and L2 states since their corresponding Kramers’ pairs, H2 and L1, increasingly contribute to its net magnetic moment. Note now that the contributions of H1 and H2 and those of L1 and L2 have opposite polarization because they are Kramers’ pairs. Therefore, the contributions of H2 and L1 cause a further and increasing reduction of the magnetic moment at the top bilayer such that it becomes exactly zero at \( M \). Thus, the fact that the magnetization vanishes at \( M \) not only avoids the magnetic moment discontinuity but also reproduces the evanescence of the magnetic moment near \( M \) found in SARPES experiments.

Regarding the observed evanescence of the spin polarization near \( M \) [34, 36], we propose that (1) the experiment simultaneously detects both spin textures around \( M \) and at the same energy (within experimental resolution) because, at least around \( M \), SARPES has actually been able to sample the two states composing each Kramers’ pair (which at other regions of the \( \Gamma M \) line are strictly on opposite sides of the film) and (2) the interaction of the Bi film with the substrate causes a negligible Rashba splitting around \( M \), as we shall discuss in more detail in section 5. In summary, the fact that the spin polarization vanishes at \( M \) not only precludes the magnetic-moment discontinuity and indetermination but is also a signature that the Kramers’ degeneracy is not detectably lifted by the presence of the substrate.

4.6. Origin of the spin polarization of the two A7-distortion metallic branches

Spin-resolved ARPES measurements on ultra-thin films have shown that the two metallic branches display opposite magnetic-quantum-number polarization [34, 36]. We have already shown that the spin texture observed in such states cannot be attached to the lifting of Kramers’ degeneracy. We must then explain why the two metallic states render opposite magnetic-moment texture experimentally if they are in principle not correlated; i.e. they are not split Kramers’ pairs.

We propose that the metallic branches have been observed to have opposite magnetic polarization (or opposite spin texture) because the hybridization leading to the covalent bonds among the Bi atoms of the surface bilayer lowers the energy and makes the surface atoms non-magnetic. Namely, covalent bonds are formed at the expense of quenching unpaired spins. Isolated Bi atoms, for example, have three unpaired electrons and nevertheless bulk Bi is non-magnetic precisely because the three covalent bonds that each atom makes lower the energy more than retaining any unpaired spin. Likewise, the net magnetization of the surface bilayer is expected to be zero because no covalent bonds are broken by the creation of the surface, and hence each surface atom retains essentially unaltered the relatively strong covalent bonds that any bulk atom makes. Thus, just as in the bulk phase, the covalent bonds of surface atoms cannot afford magnetic polarization. This behavior also falls in line with the fact that Bi is the material with largest diamagnetism. That being said, let us now analyze the calculated magnetic moment of the metallic bands within the regions where it does not vanish (i.e. away from \( M \)). Although the SOC does not lift the Kramers’ degeneracy, Kramers’ pairs (which have opposite magnetic quantum number) cannot quench each other’s magnetic moment because, as mentioned earlier, they are primarily located on opposite sides of the slab away from \( M \). So, for example, if the H1 state of the top bilayer yields a ‘spin-up’ polarization, the only possibility to cancel the net magnetization at the surface is that the L2 state (also localized at the top bilayer) acquires the opposite spin polarization.

5. Rashba splitting of Bi(111) metallic states resulting from three perturbations

We have shown that the origin of the observed splitting between the two metallic branches corresponds to the lifting of the \( m_1 = \pm 1/2 - m_2 = \pm 3/2 \) degeneracy by the SOC and is influenced by the anisotropy of the structure as well as by the coupling between the two surfaces of the film. This does not mean that the Rashba splitting is not possible but that it is just not intrinsic to Bi(111) or to the surface creation per se. In this section we shall discuss the Rashba splitting of the two metallic branches of Bi(111) produced by three different perturbations to the inversion symmetry.

5.1. The Rashba splitting by two structural perturbations

To test the Rashba-type splitting—i.e. the splitting between the Kramers’ pairs when the system lacks inversion symmetry—and compare with experiments on ultrathin films [33], we have broken the inversion symmetry of a 12 BL Bi slab by (1) increasing by 0.3 Å the relaxed inter-bilayer distance between the surface and subsurface bilayers of one side of the slab (\( d_{23} \)); and (2) increasing by 0.1 Å the relaxed intra-bilayer distance between the two layers composing the surface bilayer (\( d_{12} \)). We have chosen these perturbations because Si(111) was the substrate on which the Bi films were grown in the experiments of [34, 36]. Si(111) is expected to cause a weak perturbation at the interface with the Bi film because (1) the grown Bi film has no dangling bonds and (2) Bi atoms have slightly higher electronegativity than the atoms of the semiconducting Si substrate, for which no significant ionic or covalent bond between the Si substrate and the Bi film is expected.

The resulting electronic structure of the 12 BL film with the inter-bilayer (purple lines) and intra-layer (red lines) structural perturbation are shown in figure 12 and compared with that
of an inversion-symmetric film of the same thickness (black lines). As expected, both structural perturbations cause the two metallic branches split and display the four non-degenerate states. That is the Rashba splitting of the A7-distortion metallic branches because it does lift Kramers’ degeneracy. Not surprisingly, the metallic surface-state branches of the ‘perturbed’ surface are much less sensitive to the 0.3 Å inter-bilayer perturbation affecting the weak metallic-like bonds (figure 12(top)) than to the 0.1 Å intra-bilayer perturbation affecting the strong covalent bonds (figure 12(bottom)). The latter also comes hand in hand with the fact that the Rashba splitting is larger for the intra-bilayer perturbation. The electronic structure corresponding to the above perturbations have several noteworthy features which will help understand experimental results and previous calculations for thin films in appendix [33, 34, 36]. Specifically, we note that (1) although both of the above structural perturbations clearly break inversion symmetry, none of them depopulates the metallic states of the perturbed surface or affect significantly the dispersion of the A7 metallic states of the ‘unperturbed’ side; (2) comparison between the metallic bands of the ‘unperturbed’ surface of the inversion-asymmetric slab and those of the surface of the inversion-symmetric slab (see e.g figure 12(bottom)) shows that they are almost identical; (3) although both films in figure 12 are inversion asymmetric, the metallic states of the ‘unperturbed’ surface are expected to still display a gap at $M$, almost identical to that of the inversion-symmetric slab; (4) in both inversion-asymmetric films, the resulting Rashba splitting clearly satisfies the degeneracy at $M$ required by translational symmetry; (5) Kramers’ pairs lie mostly on opposite sides of thin films, as mentioned before.

The above five points explain why Bi(111) thin films supported on Si(111), which are clearly inversion asymmetric, still display the splitting at $M$ in agreement with DFT calculations [33]. This shows that the splitting at $M$ appears not only for simulated inversion-symmetric slabs, and (presumably inversion-asymmetric) single crystals, but also for asymmetric supported thin films. Our calculations for inversion-asymmetric films reveal that for a perturbation such as those shown in figure 12, the $M$-degeneracy would be undetectable for two reasons: (1) the Kramers’ pair of each measured band is primarily localized on the other side and (2) near $M$—i.e. the only small $k$-region in which the splitting of the Kramers’ pairs could be observed on a single side of the film (as shown section 4.5), the Rashba splitting is too small to be detected with available resolution. Furthermore, the latter also explains why the measured spin polarization [34, 36] vanishes near $M$: two states of opposite spin polarization are detected at the approximately the same energy (within experimental resolution) and thus the spin polarization vanishes. Our results in figure 12 also show that it is not necessarily correct to attach inversion symmetry to systems in which the degeneracy at $M$ is not seen [36]. An experiment performed on the ‘unperturbed’ surface of a inversion-asymmetric film bearing such ‘weak’ perturbations on the other side would not be able to distinguish whether the sample is inversion symmetric or not, certainly not based on the splitting at $M$. In other words, the splitting of the metallic branches at $M$ observed on one side of a film or sample is a necessary but not sufficient condition to attest inversion symmetry. Therefore, the idea that interaction between the two sides of a film makes the metallic branches ‘electronically inversion symmetric’ [36] is not on firm grounds. Our calculations suggest that as long as the perturbation breaking the inversion symmetry let the states of the perturbed surface to be occupied, the experiment will keep detecting the splitting of the branches at $M$, in no violation of the degeneracy at $M$ because this is satisfied regardless, as shown in figure 12. In other words, the splitting of the metallic branches is not of the Rashba type neither far from $M$ nor close to it. Thus, one can conclude that if two spin-textured states do not meet at $M$ (experimentally or computationally), then that indicates that the splitting along $\Gamma M$ does not correspond to the lifting of the spin or Kramers’ degeneracy. Interestingly, though, the Rashba splitting for relatively mild perturbations could be measured if the two sides of a perturbed film were investigated.
In particular, the states $H_1$ and $L_2$ of the ‘clean’ surface tend to merge around $M$, the surface states $H_2$ and $L_1$ of the $H$-host surface also tend to merge at $M$ but are shifted to higher energies. However, once $H$ establishes a chemical bond with Bi, the ‘original’ Kramers’ pairs, $H1$ and $L1$, completely dissociate, $H_1$ and $L_2$ merge at $M$, and $H_2$ and $L_1$ become depopulated and shifted to the conduction band. Thus, the splitting brought up by $H$-adsorption on $Bi(111)$ is that shown with red arrows in figure 13.

Since the metallic states $H_1$ and $L_2$ of the $H/Bi(111)$ films in figure 13 are not degenerate, have opposite magnetic moment, are both localized on the ‘clean’ surface, resemble the dispersion of the branches of inversion-symmetric Bi(111) films, and merge at $M$, it has been argued that $H/Bi(111)$ thin films of 11 BL [26, 35] mimic a semi-infinite $Bi(111)$ slab, and therefore the splitting between $H_{1,2}$ and $L_{1,2}$ observed by ARPES is of the Rashba type. Our results, however, show that the clean surface of $H/Bi(111)$ is not only an inadequate model of the Bi(111) surface of single crystals, but also a fundamentally different system. First, the metallic states of the bottom surface become unoccupied states (shifted to the conduction band) because $H$ takes charge from the host Bi(111) surface. In that sense, $H$-adsorption may seem to decouple the two sides of a thin film, yet such ‘decoupling’ is at the expense of a total depletion of the metallic states of the $H$-host surface and, more importantly, a partial depletion of the ‘clean’ surface’ states because the four metallic states are localized on both surfaces simultaneously near $M$ for inversion-symmetric films, as shown in section 4.5. The perturbation is so strong that it not only breaks the inversion symmetry causing a splitting of the Rashba type, but causes a complete dissociation of the original Kramers’ pairs. Furthermore, $H$-adsorption not only dissociates the metallic states but also shifts bulk-like states. Besides, it has been found that the presence of $H$ strongly affects the bonds of the ‘clean’ surface as it largely modifies its phonon dispersion [64]. Second, one can see in figure 13 that the ‘decoupling’ brought by $H$ is relative because it does not improve the convergence or resolve the thickness dependence of the calculated dispersion of the A7-distortion surface states in any way. Specifically, by comparing figures 5 and 13 for 12 and 21 BL, one can see that the $H_1$ and $L_2$ states of $H/Bi(111)$ films miss the almost perfect convergence that is achieved from $\Gamma$ to $\frac{1}{2}M$ for 12 and 21 BL inversion-symmetric Bi(111) films. More importantly, $H$-adsorption clearly worsens the thickness-dependence problem of inversion-symmetric Bi(111) films to simulate the metallic states because those of $H/Bi(111)$ are even more thickness dependent and along a larger $k$-region than those of inversion-symmetric Bi(111) films. Third, the dispersion of A7-distortion metallic states of Bi(111) for Bi single crystals [11, 31]—having perhaps the two sides essentially decoupled by millions of BL—display A7-distortion metallic states whose dispersion is better simulated by the inversion-symmetric films of 39 BL than by the 21 BL $H/Bi(111)$ film, including the fact that the metallic states of the $H$-adsorbed films lack the splitting at $M$ between the A7 metallic branches reported in all ARPES experiments for the Bi(111) surface of single crystals.

**Figure 13.** Band structure around $E_F$ of $H/Bi(111)$ slabs for slabs of 6, 12, and 21 BL, comparing its single-degenerate A7-distortion surface states against the doubly-degenerate ones of the inversion-symmetric slab of 39 BL and against the projected band-structure of bulk Bi (black lines). The splitting of the Kramers’ pairs (H1&H2 and L1&L2) caused by the presence of $H$ is indicated for the 21 BL slab (see text).

---

5.2. The Rashba splitting by adsorbing $H$ on Bi(111)

Figure 13 displays the electronic structure of hydrogen-adsorbed Bi(111) films of 6 BL (blue lines), 12 BL (green lines), and 21 BL (cyan lines), and compares them with the electronic structure of the inversion-symmetric film of 39 BL (orange lines) and with the projected bulk band (black lines). Atomic hydrogen (H) is adsorbed atop each surface atom of one side of the Bi(111) films, the ‘bottom’ surface. Note that the metallic states of Bi/Bi(111) display a dispersion similar to that of the metallic states of Bi(111) because they derive from the $H_{1,2}$ and $L_{1,2}$ branches. Since $H$ is placed on the ‘bottom’ surface, the metallic states of Bi/Bi(111) correspond to the $H_1$ and $L_2$ states of inversion-symmetric films, are singly degenerate, and have opposite magnetic polarization. Finally, notice that the thickness dependence observed in figure 13 shows that a sufficiently thick film with an inversion asymmetry caused by something such as $H$-adsorption should have metallic states strictly lying in the A7 gap and well localized all along $\Gamma M$. It would be interesting to monitor how the band structure of, for instance, the 21 BL film evolves from that shown in figure 5 to that in figure 13 as $H$ approaches one of the Bi(111) surfaces. However, it may not be necessary because we know that in inversion-symmetric films the states of each Kramers’ pair are localized on opposite surfaces for the most part, whereas in the H-terminated film while the metallic states are both localized on the ‘clean’ surface of the film, the unoccupied surface states around the conduction band are localized on the H-adsorbed surface (figure 13). The localization of each of the states in both scenarios suggests that as long as $H$ does not form a chemical bond with Bi but only breaks inversion symmetry, the splitting of the ‘original’ Kramers’ pairs ($H_1$&$H_2$ and $L_1$&$L_2$) sets in in a fashion similar to that shown in figure 12.
One may argue that calculating the band structure of H/Bi(111) and comparing it with that of Bi(111) while including SOC in both calculations may confound the search for the origin of the splitting observed by ARPES experiments. This view suggests that one should first simulate the semi-infinite slab by adsorbing H without SOC and then see the splitting caused when the SOC is incorporated. One concern about this approach is that in an experiment one cannot switch on and off the SOC, since it is there for all atoms at all times. In other words, Bi(111) and H/Bi(111) without SOC do not exist. The only thing over which we may have control is enforcing inversion symmetry/asymmetry. Nonetheless, the computational experiment of performing the calculation of H/Bi(111) without SOC ultimately corroborates that its clean surface is fundamentally different from those of inversion-symmetric Bi(111) films and does not provide a convergent dispersion of the metallic states shown in green in figure 4 (bottom). Namely, the presence of H on the bottom surface does not bring the H1,2 and L1,2 branches together near M or finally merge them at M. On the contrary, it makes a drastic splitting of similar magnitude as that shown in figure 13 in which the surface states localized on the H-host surface are largely shifted up in energy. Now, one may still argue that the purpose of adsorbing H is to eliminate the metallic states of the H-host surface for which those states should be ignored altogether. For example, in figure 1 of [26], the calculation without SOC was interpreted as implying that Bi(111) has only one metallic branch (likely because the other one is shifted way above the Fermi energy). However, Bi(111) without SOC having a single metallic branch is not supported by our findings in sections 4.1 and 4.2. Specifically, we have shown that the splitting of the A7 metallic branches from Γ to \( \sim \frac{5}{12} M \) for the 39 BL film (including SOC) is converged and thus thickness independent (figure 5). Therefore, the splitting from \( \sim \frac{5}{12} M \) to \( \sim \frac{7}{12} M \) between H1,2 and L1,2 when the SOC is not included is also thickness independent. Thus, the conjecture that the A7 metallic states of a semi-infinite slab is a single branch is not sustained, which once again indicates that H/Bi(111) is an inadequate system to model the electronic structure of Bi(111).

Nevertheless, the band structure of H/Bi(111) without SOC does provide a valuable outlook. The splitting between the H1 and L2 states of H/Bi(111) can be considered of the Rashba type because H1\&H2 and L1\&L2 completely dissociate to form ‘new’ Kramers’ pairs H1\&L2 and H2\&L1, as shown in figure 13 for the 21 BL H/Bi(111) film. Thus, one may say that the splitting between the H1 and L2 in figure 13 is the Rashba splitting of the only A7 metallic branch of H/Bi(111), but not of Bi(111). Of course, in this profoundly different scenario of inversion asymmetry in which some Kramers’ pair are dissociated and others created, the way in which the system satisfies translation symmetry (the M degeneracy) is also different. Namely, for this type of inversion asymmetry, one should really expect to find the degeneracy at M via ARPES on a single side of the film. These results suggest that, depending on the nature of the perturbation breaking the symmetry, splitting of Kramers’ pairs may be found on opposite sides of the slab (as in the mild structural perturbations in section 5.1) or on the same side (if there is charge depletion of the surface states of one side). Interestingly, if the Bi(111) surface were indeed that in which the Kramers’ pairs are located on the same surface, that would mean that the entanglement of the Kramers’ pair only exists for thin films. Yet, so far, there is no evidence for that.

H-adsorption on Bi(111) is just one of the many possible perturbations that may render inversion asymmetry. The drastic splitting of the Kramers’ pairs seen in H/Bi(111) is specific to the depletion of the surface states from one side of the film but not necessarily to H since other species might have the same effect. In fact, this way of producing a Rashba splitting is not alien to those used in semiconductors. For example, the charge depletion of the H-host surface creates a scenario similar to that in which an inversion charge-density layer is created on the surface of semiconductors by an external electric field [51, 52] while the other side undergoes charge depletion. Therefore, should such strong perturbation and the charge reallocation within the film be possible by mere chemisorption, such systems might be more interesting to study than Bi(111) per se, in part also because in cases such as that of H/Bi(111) films, the only Kramers’ pair occupied is localized on the ‘clean’ side and there is a slightly stronger localization of the metallic states on the clean bilayer of a H/Bi(111) film that in the topmost bilayer of an inversion-symmetric Bi(111) film.

**6. Summary**

We present fully relativistic density-functional-theory calculations of the electronic structure of Bi(111) films to revisit the interpretation of the metallic surface states of Bi(111) and some of the unresolved issues around thin films and their limit to single crystals. We investigate the spin–orbit pseudogaps and surface states in the valence band of the projected band-structure of bulk Bi and obtain for the first time some of the actual spin–orbit surface states from first principles. The evolution of the electronic band structure as a function of thickness indicates that calculations for films as thick as \( \sim 9 \) nm reproduce fairly well the measured band-structure of single crystals within experimental resolution. Our first-principles calculations show that the metallic surface states near the zone edge (the M point) are not resonances in the sense that they do not overlap with the bulk band at the zone boundary and do not meet at M for films of up to 15.5 nm. They lie strictly inside the A7-distortion conduction-valence band gap, which at M is at least three times wider than predicted by tight-binding calculations. Thus, our results confirm that available ARPES measurements of the metallic surface-state bands on the Bi(111) surface of single crystals have sampled the vicinity of the M point without overlapping with the valence band [11, 13, 31]. We also show that the intensity of the peaks is weakened because the states become increasingly delocalized as k approaches M. This also shows that juxtaposing tight binding calculations of the bulk-projected band against first-principles calculations of the surface states [6, 26] is not
reliable, as well as the conclusions drawn from the assumption that the metallic branches overlap with the bulk band. Our calculations together with ARPES experiments [11, 13, 31] show that the band structure obtained by adsorbing H on one side of a film only 4 nm (11 BLs) is not representative of that of the Bi(111) surface of a single crystal. H adsorption does not help converge the electronic structure to that of thicker slab (of up to 15.5 nm). On the contrary, it disrupts the dispersion of the surface states that is already obtained without H.

We find that the splitting of the two A7-distortion metallic surface-state branches along ΓM arises for inversion-symmetric films even if the SOC is not included in the calculations, for which the splitting is not totally ascribable to the SOC. The splitting between the two branches can be explained in terms of the crystal anisotropy, long-range interactions, and SOC. We show that near Γ the SOC causes a m_j = ±3/2−m_j = ±1/2 splitting. In the middle of the ΓM direction, the splitting arises because of the crystal anisotropy and it is largely enhanced by the SOC. Near M, however, the splitting between the two branches is strongly influenced by the long-range interaction between the two sides of the film. Since the splitting exists in inversion-symmetric films and does not lift Kramers’ degeneracy, its presence is not caused by the Rashba effect.

We show that the splitting between the two metallic branches at M as a function of thickness cannot be extrapolated to find the thickness at which such splitting would vanish. Nevertheless, we show that the splitting at M is not in conflict with the translational symmetry in inversion-symmetric or inversion-asymmetric films because (1) the splitting found on a single side of a film does not correspond to the lifting of the Kramers’ degeneracy and (2) the two-fold degeneracy at M is guaranteed for inversion-symmetric and asymmetric films. Thus, the splitting at M requires no further explanation [35, 36] or closure. Moreover, the expected degeneracy at M has not been observed for inversion-asymmetric thin films [34, 36] or for the surface of single crystals [11, 31]. We find that the only reason for which the splitting between the two metallic branches at M could be of concern—the magnetic discontinuity and indetermination at M—is resolved since the magnetization at M is zero. We show that the spin texture of the metallic branches is reduced near M and exactly vanishes at M because the states composing each Kramers’ pair become increasingly localized at both sides of the slab cancelling out each other’s magnetic moment. The fact that the spin polarization vanishes at M not only precludes the magnetic-moment discontinuity and indetermination but is also a signature that the Kramers’ degeneracy is hardly lifted for supported thin films, since otherwise the degenerate Kramers’ states would have different energies and would be distinguishable by their spin texture.

We provide a rationale for the observed opposite spin-texture of the two metallic bands, in light of the fact that they cannot be attached to a broken Kramers’ degeneracy and that they are not correlated. We propose that the states have opposite magnetic polarization because the Bi surface atoms have no true dangling bonds, hence the strong hybridization within the surface bilayer lowers the energy and makes it non-magnetic. Specifically, each surface atom retains essentially unaltered the relatively strong covalent bonds that any bulk atom has, which cannot afford magnetization. Thus the two bands localized at each surface of a film, although non-correlated, must have opposite spin. We show that lifting of the Kramers’ degeneracy (Rashba splitting) takes place only if the inversion symmetry is broken and its magnitude is dependent upon the strength of the perturbation that breaks it. In other words, the Rashba splitting is, in general, not intrinsic to Bi(111) but to the perturbation imposed on the sample to break inversion symmetry. Our calculations illustrate that the extent of the Rashba-splitting on Bi(111) depends on the magnitude of the perturbation. We show that for structural perturbations that do not involve charge polarization, donation or sharing, the Rashba splitting is so small that it could not be detected within the current experimental resolution, which explain the measurements in [33, 34]. In turn, the band structure yielded by H/Bi(111), although not representative of that of pristine Bi(111), is representative of the Rashba splitting caused by the strong perturbation imparted by chemisorption, which depletes the metallic states of one side of the Bi film. Our calculations for inversion-asymmetric films indicate that if ARPES experiments observe that the two metallic states of Bi(111) do not meet at M, then the measured states are not Kramers’ pairs even if Kramers’ degeneracy is lifted by the presence of a substrate.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: http://marisolalcantaraortigoza.info/bi111-electronic-structure-data/.

Acknowledgments

This work was supported by DOE, Grant DE-FG02-07ER46354. Computations were performed at the Stokes HPCC facility at UCF Institute for Simulation and Training (IST) and local machines at UCF. MAO thanks Rolf Heid and Klaus Peter Bohnen for facilitating her stay at Karlsruhe Institute of Technology where the bulk of the calculations were carried out. We are indebted to Rolf Heid, Klaus Peter Bohnen, Sergey Stolbov, and Christian Ast for countless insightful discussions.

ORCID iDs

Marisol Alcántara Ortigoza 🌐 https://orcid.org/0000-0003-3579-4894
Talat S Rahman 🌐 https://orcid.org/0000-0003-3889-7776

References

[1] Shoenberg D 1939 Proc. R. Soc. A 170 341
[2] Jérome D, Rice T M and Kohn W 1967 Phys. Rev. 158 462
[3] Brandt N B and Chudinov S M 1972 J. Low Temp. Phys. 8 339
[4] Miura N, Hiruma K, Kido G and Chikazumi S 1982 Phys. Rev. Lett. 49 1339
[5] Weitzel B and Micklitz H 1991 Phys. Rev. Lett. 66 385
[6] Hofmann P 2006 Prog. Surf. Sci. 81 191
[7] Vossloh C, Holdenried M and Micklitz H 1998 Phys. Rev. B 58 12422
[8] Muntyanu F M, Gilewski A, Nekov K, Zaleski A and Christoph V 2008 Solid State Commun. 147 183
[9] Tian M, Kumar N, Chan M H W and Mallouk T E 2008 Phys. Rev. B 78 045417
[10] Hsieh D et al 2009 Science 323 919
[11] Ohtsubo Y, Pertelitti L, Goerbig M O, Fevre P L, Bertran F and Taleb-Ibrahim A 2013 New J. Phys. 15 033041
[12] Chang T-R, Lu Q, Wang X, Lin H, Miller T, Chiang T-C and Bian G 2019 Crystals 9 510
[13] Hengsberger M, Segovia P, Garnier M, Purdie D and Baer Y 2000 Eur. Phys. J. B 17 603
[14] Ast C R and Höchst H 2001 Phys. Rev. Lett. 87 177602
[15] Wells J W, Handrup K, Kallehauge J F, Gammelgaard L, Chang T-R, Lu Q, Wang X, Lin H, Miller T, Chiang T-C and Bian G 2019 Crystals 9 510
[16] Varykhalov A, Marchenko D, Scholz M R, Rienks E D L, Kim T K, Bihlmayer G, Sánchez–Barriga J and Rader O 2012 Phys. Rev. Lett. 108 066804
[17] Tamtögl A, Kraus P, Mayrhofer-Reinhartshuber M, Campi D, Bernasconi M, Benedek G and Ernst W E 2013 Phys. Rev. B 87 035410
[18] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[19] Alcántara Ortigoza M, Heid R and Bohnen K-P 2011 personal communication.
[20] Patthey F, Schneider W D and Micklitz H 1994 Phys. Rev. B 50 11827
[21] Bangert E, v. Klitzing K and Landwehr G 1974 Phys. Rev. B 10 46403
[22] Patthey F, Schneider W D and Micklitz H 1994 Phys. Rev. B 49 11293
[23] Tanaka A, Hatano M, Takahashi K, Sasaki H, Suzuki S and Sato S 1999 Phys. Rev. B 59 1786
[24] Thomas J, Jezequel G and Pollini I 1999 J. Phys.: Condens. Matter 11 9571
[25] Liu Y and Allen R E 1995 Phys. Rev. B 52 1556
[26] Ast C R and Höchst H 2003 Phys. Rev. B 67 113102
[27] Mönig H, Sun J, Koroteev Y M, Bihlmayer G, Wells J, Chulkov E V, Pohl K and Hofmann P 2004 Phys. Rev. Lett. 93 046403
[28] Patthey F, Schneider W D and Micklitz H 1994 Phys. Rev. B 49 11293
[29] Golin S 1968 Phys. Rev. 166 643
[30] Anisichkiv V, Falicov L M and Yndurain F 1986 Phys. Rev. B 33 4352
[31] Golin S 1968 Phys. Rev. 166 643
[32] Anisichkiv V, Falicov L M and Yndurain F 1986 Phys. Rev. B 33 4352
[33] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[34] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[35] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[36] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[37] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[38] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[39] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[40] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[41] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[42] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[43] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[44] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[45] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[46] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375
[47] Alcántara Ortigoza M, Sklyadneva I Y, Heid R, Chulkov E V, Hortamani M and Wiesendanger R 2012 New J. Phys. 14 375