Electronic structures and surface states of the topological insulator $Bi_{1-x}Sb_x$

Hai-Jun Zhang,[1] Chao-Xing Liu,[2] Xiao-Liang Qi,[3] Xiao-Yu Deng,[1] Xi Dai,[1] Shou-Cheng Zhang,[3] and Zhong Fang[1]
[1]*Beijing National Laboratory for Condensed Matter Physics and Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*
[2]*Department of Physics, McCullough Building, Stanford University, Stanford, California 94305-4045, USA*
[3]*Center for Advanced Study, Tsinghua University, Beijing 100084, China*
(Received 9 March 2009; revised manuscript received 14 July 2009; published 13 August 2009)

We investigate the electronic structures of the alloyed $Bi_{1-x}Sb_x$ compounds based on first-principles calculations including spin-orbit coupling (SOC), and calculate the surface states of semi-infinite systems using maximally localized Wannier function. From the calculated results, we analyze the topological nature of $Bi_{1-x}Sb_x$, and found the followings: (1) pure Bi crystal is topologically trivial. (2) Topologically nontrivial phase can be realized by reducing the strength of SOC via Sb doping. (3) The indirect bulk band gap, which is crucial to realize the true bulk insulating phase, can be enhanced by uniaxial pressure along *c* axis. (4) The calculated surface states can be compared with experimental results, which confirms the topological nature. (5) We predict the spin-resolved Fermi surfaces and showed the vortex structures, which should be examined by future experiments.

DOI: 10.1103/PhysRevB.80.085307 PACS number(s): 71.15.Dx, 71.18.+y, 73.20.At, 73.61.Le

I. INTRODUCTION

In an ordinary insulator, the valence and conduction bands are separated by an energy gap, making it electrically inert. Therefore, the ordinary insulator is not sensitive to the change in boundary condition. Recently a new class of insulator, namely, topological insulator (TI), was proposed.[1–6] TI also has a bulk energy gap, which is usually generated by spin-orbit coupling (SOC); however, it is different from the ordinary insulator in the sense that topologically protected gapless states, robust against disorder, appear at the edge or surface of a finite sample within the bulk energy gap. Thus the TI has conducting channels along its edge or surface. The quantum spin Hall (QSH) insulator, such as HgTe/CdTe quantum well,[2,3,7] is an example of two-dimensional (2D) TI. The conducting edge channels of HgTe/CdTe quantum wells have been theoretically predicted[2] and experimentally observed.[3,7] From the theoretical point of view, the TI can be distinguished from the ordinary insulator by the Z_2 topological invariants,[1,5,6,8] and the existence of gapless spin-filtered edge states on the sample boundary is guaranteed for TI. The edge states come in Kramers's doublets, and time-reversal (TR) symmetry ensures the crossing of their energy band at time-reversal invariant momenta (TRIM). Since these band crossings on the edge are protected by TR, they cannot be removed by any perturbation respecting the TR symmetry, such as nonmagnetic impurities. It is expected that the robust gapless spin-filtered surface (edge) states have novel applications in spintronics.

Besides the 2D QSH insulator, the TI can also exist in three-dimensional (3D) material.[4,8–13] Similar to the edge states in 2D QSH insulator, in 3D TI, topological surface state protected by TR, which can be described by odd number of Dirac points, emerges at the surface of the finite 3D sample.[4,6,8,14] Compared with the 2D TI, the 3D TI and its surfaces can be readily investigated by angle-resolved photoemission spectroscopy (ARPES) and scanning tunnel microscope experiments. The 3D TI also displays the remarkable topological magnetoelectric effect.[6] Therefore searching for realistic 3D TI is now becoming an attractive and challenging subject.

It was first suggested that the semiconducting alloy of bismuth and antimony ($Bi_{1-x}Sb_x$) is an example of such 3D TI.[8,11] Based on the tight-binding (TB) model of Liu and Allen,[15] Fu and Kane[8] developed a theory to analyze the topological nature of the surface state in $Bi_{1-x}Sb_x$ alloy.[11] Experimentally, Hsieh *et al.*[10] observed the surface states by high-momentum-resolution ARPES and demonstrated the topological nature of the surface states by counting the number of the Fermi surface crossings from the zone center to the boundary. However, clear discrepancies exist between the theory and the experiment about the surface states, although their final conclusions are consistent with each other. On the other side, the surface states of pure Bi or Sb have been intensely studied experimentally and theoretically,[16–21] but there are still fewer careful studies of their alloy. Therefore, in this paper, we present a systematic study of the surface states of $Bi_{1-x}Sb_x$ alloy, based on quantitative first-principles calculations. We show that pure Bi is topologically trivial because the SOC is too strong. The effective role of Sb doping is to reduce the strength of SOC and revert the band ordering at *L* point of the Brillouin Zone (BZ). Finally strong topological insulator can be reached by Sb doping. By constructing the maximally localized Wannier function (MLWF) from the *ab initio* schemes, we calculate the surface states of semi-infinite system, and analyze the shape of Fermi surfaces as well as the spin-resolved local density of states. These results are compared with the experiment of Hsieh *et al.*[10] In addition, although direct band gap exists in $Bi_{1-x}Sb_x$ system, indirect band gap can be only realized for a very narrow doping range, and the material has long been regarded as typical semimetallic system. In order to make the bulk material insulating completely, based on our calculations, we predict that an efficient way to enhance the indirect band gap is to apply uniaxial pressure along the *c* axis.

The paper is organized as follows. In Sec. II we discuss the crystal structure and symmetry of $Bi_{1-x}Sb_x$. In Sec. III we study the transition between the topological nontrivial and trivial phases and present a schematic phase diagram as a

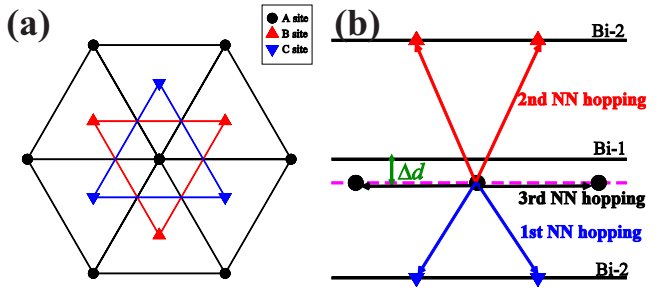

FIG. 1. (Color online) (a) Along the (111) direction of rhombohedral A7 structure, there are three possible atomic positions (A, B, and C) for the triangle plane. Black filled dots denote *A* site, red filled up triangles denote *B* site, and blue filled down triangles denote *C* site. (b) The schematic plot of Bi-layers projected onto the plane paralleling to the (111) axis. After the dimerization of two Bi layers, the Bi-2 moves to the dashed site. Δd denotes the magnitude of the dimerization. first and second NN hopping are the interlayer hopping between the different sublattices, while third NN hopping is the intralayer hopping within the same sublattice.

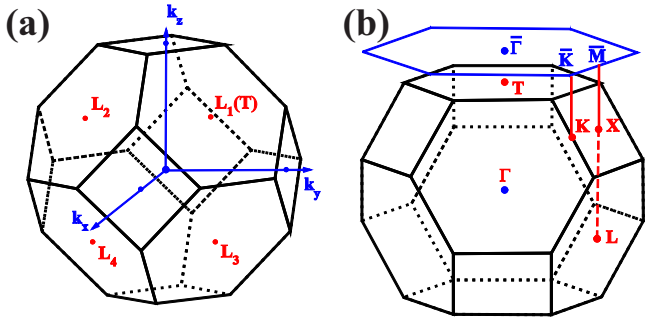

FIG. 2. (Color online) (a) BZ of fcc structure. (b) 3D BZ of rhombohedral A7 structure and its projection onto the [111] surface. The A7 BZ can be obtained from the fcc BZ by two steps: (1) rotating the (111) direction of cubic to be along the *c* axis; (2) slightly distortion of the BZ along the *c* axis. The L_1 point in fcc BZ is changed to be *T* point in the A7 BZ, which is now inequivalent to $L_{2,3,4}$ due to the distortion. *T* (Γ), *X*(l), and *K* in 3D BZ are projected to $\bar{\Gamma}$, $\bar{M}$, and $\bar{K}$ in the 2D BZ of [111] surface.

function of SOC. In Sec. IV we develop an accurate method based on MLWF to obtain the surface Green's Function, and study the topologically nontrivial surface states of $Bi_{1-x}Sb_x$. In Sec. V we provide a brief discussion and conclusion.

II. STRUCTURE AND SYMMETRY

Bi and Sb have the same rhombohedral A7 crystal structure[15] with space group $R\bar{3}m$. The A7 structure can be regarded as a distorted fcc NaCl structure. For fcc NaCl structure, there are two sets of sublattices, say Bi-1 for Na sites and Bi-2 for Cl sites, which both form the fcc structure. In such fcc structure, Bi-1 and Bi-2 sublattices are equivalent and we can shift Bi-1 sublattice by (1/2,0,0), (0,1/2,0), or (0,0,1/2) of cubic structure to obtain Bi-2 sublattice. Along the (111) direction of fcc structure, Bi triangle layers are stacked with the sequence of *ABCABCABC*…, where *A*, *B*, and *C* denote three different atomic positions for triangle plane, as shown in Fig. 1. Without distortions, there exist two kinds of space-inversion center. One is located at the Bi layer center and each Bi sublattice is space inverted to itself (called type-I inversion), while the other one is located at the middle way of two Bi layers and each Bi sublattice is space inverted to another sublattice (called type-II inversion).

Starting from fcc NaCl structure, two steps are required to obtain rhombohedral A7 structure. One is the stretching along the cubic (111) direction, while the second is the relative shift of inter-Bi-layer distances along the (111) direction, or in other words dimerization of two Bi layers. The second step breaks the type-I inversion symmetry but preserve the type-II inversion. Therefore, after distortions, the inversion symmetry can only transfer the atoms from one sublattice to the other, which is very important for our following discussion.

The BZ of fcc structure and rhombohedral A7 structure are shown in Fig. 2. For fcc structure, there are four equivalent *L* points, which are located at $L_1=(\pi,\pi,\pi)$, $L_2=(\pi,-\pi,\pi)$, $L_3=(\pi,\pi,-\pi)$, and $L_4=(\pi,-\pi,-\pi)$ of BZ. After two kinds of distortions along the cubic (111) direction, the fcc structure changes to the rhombohedral A7 structure, which breaks the equivalence between L_1 (which is denoted as *T* point in BZ of rhombohedral A7 structure) and $L_{2,3,4}$.

III. EFFECT OF ALLOYING AND PHASE DIAGRAM

Although pure Bi and Sb have been studied extensively, the alloyed system is not carefully considered yet. In this section, we will present a simple TB model by taking the strength of SOC λ and dimerization Δd as two key parameters to describe the effect of alloying. Based on this TB model, we obtained a schematic phase diagram for the topological nature of the compounds, which is instructive for us to understand the main physics. Then in the next section, surface states of alloyed system will be studied from accurate *ab initio* calculations based on virtual crystal approximation (VCA).

A. Effect of alloying

In order to take into account the effect of alloying, certain kinds of approximations have to be introduced. The first step, which is conventionally followed, is to assume the uniform distribution and neglect the disorder effects. However, this approximation is not sufficient. To further simplify our understanding, we emphasize the following factors: (1) Sb is located just on top of Bi in the periodical table, therefore, they have the same number of valence electrons and form the same A7 crystal structure. (2) Even 30% Sb alloying into Bi will only modify the lattice parameters by around 1%.[15] (3) The atomic SOC strength of Sb is weaker than Bi by a factor of 3. Therefore, we believe the strongest effect of Sb alloying into Bi is to reduce the SOC strength, and we can neglect the effect coming from the change in lattice parameters. Following this strategy, we construct a simple TB model to understand the main physics.

We consider one *s* and three *p* orbitals of each Bi atoms, together with the two sublattices and two spin degree of free-

dom, and totally there are 16 orbitals, denoted as $|s_1^{\uparrow}\rangle$, $|s_1^{\downarrow}\rangle$, $|p_{1x}^{\uparrow}\rangle$, $|p_{1y}^{\uparrow}\rangle$, $|p_{1z}^{\uparrow}\rangle$, $|p_{1x}^{\downarrow}\rangle$, $|p_{1y}^{\downarrow}\rangle$, $|p_{1z}^{\downarrow}\rangle$, $|s_2^{\uparrow}\rangle$, $|s_2^{\downarrow}\rangle$, $|p_{2x}^{\uparrow}\rangle$, $|p_{2y}^{\uparrow}\rangle$, $|p_{2z}^{\uparrow}\rangle$, $|p_{2x}^{\downarrow}\rangle$, $|p_{2y}^{\downarrow}\rangle$, and $|p_{2z}^{\downarrow}\rangle$. Here z axis is taken along the (111) direction of fcc structure or the (001) direction of hexagonal cell and subscript number denotes different sublattices. The hopping parameters are defined in Fig. 1, with the expression

$$V_{pp\sigma} = \tilde{V}_{pp\sigma}[1 + \alpha(\Delta d - \Delta d_0)]$$

$$V_{pp\pi} = \tilde{V}_{pp\pi}[1 + \beta(\Delta d - \Delta d_0)]$$

$$V_{ss\sigma} = \tilde{V}_{ss\sigma}[1 + \gamma(\Delta d - \Delta d_0)]$$

$$V_{sp\sigma} = \tilde{V}_{sp\sigma}[1 + \delta(\Delta d - \Delta d_0)] \quad (1)$$

for the first nearest-neighbor (NN) hopping,

$$V'_{pp\sigma} = \tilde{V}'_{pp\sigma}[1 - \alpha(\Delta d - \Delta d_0)]$$

$$V'_{pp\pi} = \tilde{V}'_{pp\pi}[1 - \beta(\Delta d - \Delta d_0)]$$

$$V'_{ss\sigma} = \tilde{V}'_{ss\sigma}[1 - \gamma(\Delta d - \Delta d_0)]$$

$$V'_{sp\sigma} = \tilde{V}'_{sp\sigma}[1 - \delta(\Delta d - \Delta d_0)] \quad (2)$$

for the second NN hopping and

$$V''_{pp\sigma} = \tilde{V}''_{pp\sigma}$$

$$V''_{pp\pi} = \tilde{V}''_{pp\pi}$$

$$V''_{ss\sigma} = \tilde{V}''_{ss\sigma}$$

$$V''_{sp\sigma} = \tilde{V}''_{sp\sigma} \quad (3)$$

for the third NN hopping. Here the parameters $\tilde{V}_m$, $\tilde{V}'_m$, and $\tilde{V}''_m$ are for pure Bi which are taken from the Liu-Allen model,[15] where m is $pp\sigma$, $pp\pi$, $ss\sigma$, or $sp\sigma$, respectively. Since the intralayer Bi-Bi distance is larger than interlayer Bi-Bi distance, V_m and V'_m are larger than V''_m. Linear dependence of Δd due to the dimerization is assumed and Δd_0 is for the experiment structure of pure Bi. When $\Delta d=0$, this system has no dimerization, leading to $V_m=V'_m$. Then a set of linear equations is obtained, which can be used to determine the value of the parameters α, β, γ, and δ.

Besides the hopping terms, the atomic SOC Hamiltonian can also be taken into account

$$H_{\mathrm{SOC}} = \frac{\hbar}{4m^2c^2}[\nabla \mathbf{V} \times \mathbf{P}] \cdot \sigma, \quad (4)$$

where $\nabla \mathbf{V}$ is the potential gradient, $\mathbf{P}$ is the momentum, and σ is Pauli spin matrices. In the atomic limit, the matrix elements of H_{SOC} in our atomic basis set can be explicitly written down, as shown in Liu's paper.[15] In the H_{SOC} matrix elements λ is the spin-orbit coupling parameter.

Therefore, the final Hamiltonian is given as the function of two variables Δd and λ with the form

$$\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22}, \end{bmatrix} \quad (5)$$

where $H_{11}=H_{22}$, $H_{12}=H_{21}^{\dagger}$ are 8×8 matrices. H_{11} includes intrasublattice hopping and on-site SOC interaction, while H_{12} represents interlayer hopping. Since the Hamiltonian has the type II inversion symmetry at T and L points, a unitary transformation is applied here to rewrite the Hamiltonian in the new basis with unambiguous parity, the odd parity basis, $\frac{1}{\sqrt{2}}(|s_1^{\sigma}\rangle - |s_2^{\sigma}\rangle)$, $\frac{1}{\sqrt{2}}(|p_{1x}^{\sigma}\rangle + |p_{2x}^{\sigma}\rangle)$, $\frac{1}{\sqrt{2}}(|p_{1y}^{\sigma}\rangle + |p_{2y}^{\sigma}\rangle)$, and $\frac{1}{\sqrt{2}}(|p_{1z}^{\sigma}\rangle + |p_{2z}^{\sigma}\rangle)$ and the even parity basis, $\frac{1}{\sqrt{2}}(|s_1^{\sigma}\rangle + |s_2^{\sigma}\rangle)$, $\frac{1}{\sqrt{2}}(|p_{1x}^{\sigma}\rangle - |p_{2x}^{\sigma}\rangle)$, $\frac{1}{\sqrt{2}}(|p_{1y}^{\sigma}\rangle - |p_{2y}^{\sigma}\rangle)$, and $\frac{1}{\sqrt{2}}(|p_{1z}^{\sigma}\rangle - |p_{2z}^{\sigma}\rangle)$. With the new basis, the Hamiltonian is changed to be $\tilde{H}$,

$$\begin{bmatrix} \tilde{H}_{11} & 0 \\ 0 & \tilde{H}_{22} \end{bmatrix} \quad (6)$$

which is block diagonal, because the odd and even parity states will not mix in a system with space-inversion symmetry.

B. Phase diagram

The topological nature of the system can be determined from the parity of the occupied bands at TRIM.[8] The band gap for $Bi_{1-x}Sb_x$ is near T and L points, therefore, here we focus on one T and three L points in BZ. The parity of the occupied bands for T and L points can be easily obtained since the eigenstates of $\tilde{H}_{11}$ have odd parity while those of $\tilde{H}_{22}$ have even parity. In Fig. 3, the energy levels of six p bands for both L and T points are plotted as a function of Δd, where two different values of λ are chosen, $\lambda=1.5$ eV corresponding to the value of Bi and $\lambda=0.5$ eV corresponding to the value of Sb. Three lowest levels of total six p bands should be occupied, namely, the conduction band and valence band are the third lowest band and fourth lowest band, respectively, which have opposite parities. When increasing the dimerization parameter Δd, at T point the band gap increases rapidly and there is no leveling crossing between the conduction band and valence band, while at L point, the band gap is quite small and the sequence of conduction band and valence band can even change with $\lambda=0.5$ eV. When $\Delta d=\Delta d_0$, as indicated by the dashed line in Fig. 3, the occupied valence bands for Bi ($\lambda=1.5$ eV) and Sb ($\lambda=0.5$ eV) have different parities at L point but the same at T point, therefore, we conclude the topological natures of Bi and Sb are different. This result is the same as that of Fu and Kane,[8] in which they claim that Z_2 invariants $(\nu_0;\nu_1\nu_2\nu_3)$ are (0;000) for Bi, which is topologically trivial but (1;111) for Sb, which corresponds to topological nontrivial phase.

By determining the gap closing line of the conduction and valence bands at L points, we can obtain the phase diagram of the alloy $Bi_{1-x}Sb_x$ as a function of SOC parameter λ and dimerization parameter Δd, as shown in Fig. 4. Since the topological nature of the system can only be changed by closing bulk gap, in the region where Sb locates, the system has the same nontrivial topological behavior to Sb, while in the other region, the system is topological trivial, which is

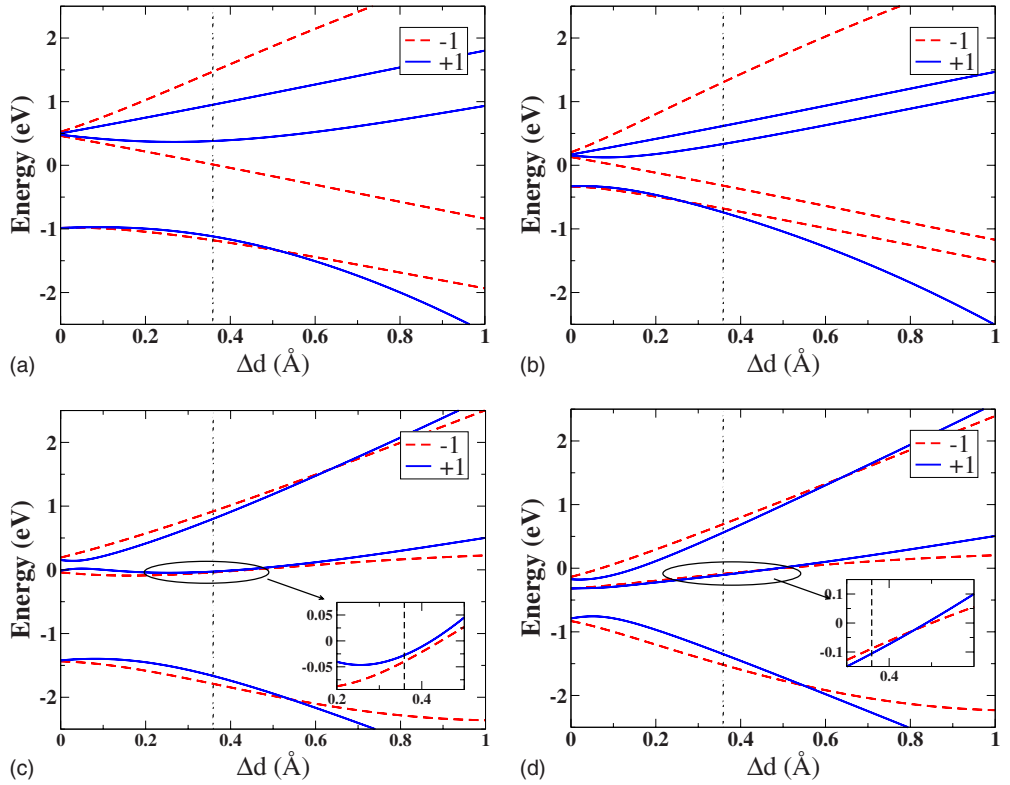

FIG. 3. (Color online) Energy bands are plotted as a function of the dimerization parameter (Δd) with SOC parameter λ taken as 1.5 and 0.5 eV. (a) and (b) are for T point while (c) and (d) for L point. The solid blue lines denote the states with parity +1 while the dashed red ones are the states with parity −1. The dashed black line represents experimental dimerization ($\Delta d = \Delta d_0$).

the same as Bi. It can be also understood from the parity analysis that the Bi is topologically trivial because both L and T points have reverted bands ordering, namely, parity are −1 for both L and T. However, by reducing the SOC strength (i.e., Sb doping), the parity of L is recovered to be +1 while T remains to be −1, therefore, topological nontrivial system is realized. In the next section, accurate simulation based on *ab initio* calculations will be presented.

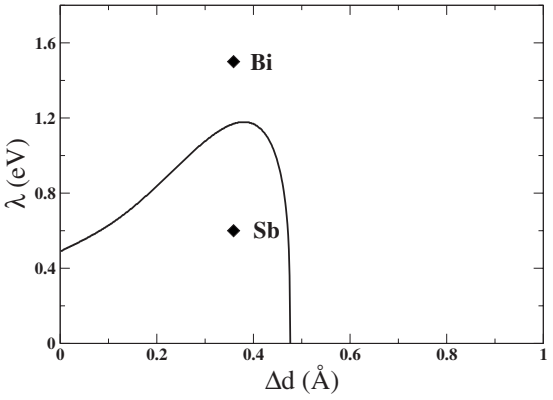

FIG. 4. The phase diagram of the system with two variables: the dimerization parameter Δd and SOC parameter λ. In the region where Sb locates, the parity of L and T points are +1 and −1, respectively, in other words, the system is in the topological nontrivial phase with Z_2 invariants (1;111). In the other region where Bi locates, however, the parity of L and T points are all −1, and the Z_2 invariants are (0;000).

IV. *AB INITIO* CALCULATIONS AND SURFACE STATES

A. *Ab initio* method and surface Green's function

The *ab initio* calculation is carried out by our BSTATE (Beijing Simulation Tool for Atom Technology)[22] code with plane-wave pseudopotential method. The generalized gradient approximation of PBE type[23] is used for the exchange-correlation potential. Especially in Bi's pseudopotential, we take into account $5d^{10}$ electrons as the valence band by ultrasoft pseudopotential scheme. The k-mesh is taken as $12 \times 12 \times 12$ and the cut-off energy is 340.0 eV for the self-consistent calculation. For pure Bi, the optimized lattice parameters are $a=4.669$ Å, $c=12.1506$ Å, and $d=0.2341$ Å, which are in good agreement with previous calculations.

To take into account the effect of alloying, VCA is necessary. There are several ways to do this, particularly in the pseudopotential approach, the simple linear combination of Sb and Bi pseudopotentials can be used, and the corresponding pseudopotential of alloyed virtual atom is regenerated by solving the atomic problem again.[22] Unfortunately, this procedure is not accurate enough for our purpose here. As already suggested by previous studies,[22] such VCA procedure can be used for those states far away from the Fermi level, however, for those states very close to the Fermi level, the error bar is big. The system we study here ($Bi_{1-x}Sb_x$) is very sensitive to the p orbitals, such VCA pseudopotential cannot give sufficient accuracy. In order to have an accurate VCA scheme, we need to consider the particularity of our system. As we already explained in the last section, the alloyed $Bi_{1-x}Sb_x$ have the same crystal structure and almost the same

structure parameters. The main effect of Sb alloying is to tune the SOC strength, λ. Therefore we may have a simple yet accurate VCA scheme. Here we take Bi's parameters for simplicity and tune λ in Bi's pseudopotential to simulate the doping parameter x of the alloyed $Bi_{1-x}Sb_x$. In such a way, since we do not need to the solve the atomic problem again, the pseudopotential is accurate enough.

We are interested in the surface states of the semi-infinite system, a method based on MLWF is developed to calculate the surface states of semi-infinite system. The *ab initio* MLWF (Refs. 24 and 25) method can be regarded as an exact TB method with its parameters calculated from *ab initio* self-consistent electronic-structure calculations. First, the semi-infinite $Bi_{1-x}Sb_x$ system can be divided into two parts: the bulk part and the surface part. The bulk part Hamiltonian is constructed with the MLWFs from bulk $Bi_{1-x}Sb_x$ *ab initio* calculations, while the surface part Hamiltonian is constructed with MLWFs from $Bi_{1-x}Sb_x$ film with slab calculations. With these MLWFs' hopping parameters, iterative method[26,27] is adopted to solve the surface Green's function of the semi-infinite system $G_{nn}^{l\alpha,l\alpha}(\mathbf{k}_\parallel, \epsilon+i\eta)$, where n and l denote the supercell along z direction and the atomic bilayer plane within one supercell, respectively. α gives the orbital index in one atomic bilayer and $\mathbf{k}_\parallel$ is a good quantum number in semi-infinite system.

The charge density of states (DOS) and spin DOS (Ref. 9) are related to the surface Green's function with the expression

$$N_n^l(\mathbf{k}_\parallel, \epsilon) = -\frac{1}{\pi}\mathbf{Im}\sum_\alpha G_{nn}^{l\alpha,l\alpha}(\mathbf{k}_\parallel, \epsilon + i\eta) \tag{7}$$

and

$$S_{n,\sigma}^l(\mathbf{k}_\parallel, \epsilon) = -\frac{1}{\pi}\mathbf{Im}\sum_{\alpha,\beta} G_{nn}^{l\alpha,l\beta}(\mathbf{k}_\parallel, \epsilon + i\eta) O_{l\beta,l\alpha}^\sigma$$

$$O_{\beta,\alpha}^\sigma = \langle l\beta|\hat{s}^\sigma|l\alpha\rangle, \tag{8}$$

respectively, where $\hat{s}^\sigma$ is the spin ($s_{x,y,z}$) operator. When $n=0$, $l=0$, $N_0^0(\mathbf{k}_\parallel, \epsilon)$ and $S_{0,\sigma}^0(\mathbf{k}_\parallel, \epsilon)$ give the local DOS and local spin DOS at the surface and in the following we use $N(\mathbf{k}_\parallel, \epsilon)$ and $S_\sigma(\mathbf{k}_\parallel, \epsilon)$ for short.

B. Surface state and Fermi surface

The local DOS $N(\mathbf{k}_\parallel, E)$ at the surface is plotted for two different SOC parameters, λ=1.28 eV and λ=1.1 eV for the topologically trivial and nontrivial situations, respectively. In Fig. 5 with λ=1.1 eV, developing from the $\bar{\Gamma}$ point, there exists two surface bands connected to the $\bar{M}$ point, which are denoted as Σ_1 and Σ_2, respectively. At $\bar{M}$ point, Σ_2 band returns to valence band while Σ_1 band merges into the conduction band. Therefore, those surface states cross the Fermi energy five times in total (odd number), which indicates the topologically nontrivial nature of this phase. On the contrary, in Fig. 6 (for λ=1.28 eV), both Σ_1 and Σ_2 bands return to the valence band at $\bar{M}$ point, and they cross the Fermi level four times (even number). This indicates that the system is topological trivial. $Bi_{1-x}Sb_x$ alloy with effective SOC parameter λ being 1.28 eV belongs to the same topological phase of pure Bi bulk crystal. The surface state in Fig. 6 is also consistent with pervious works.[19,28] In an ARPES experiment,[20] the surface states of single-crystal Bi was observed to be different from our theoretical results. However, in Ref. 20 the authors also discussed that the interaction between the topmost bilayer and the substrate becomes very weak because of breaking the transformation symmetry at the surface. Thus the system they studied is a Bi bilayer film rather than a bulk crystal, so that there is no direct discrepancy between the experimental result and our calculation.

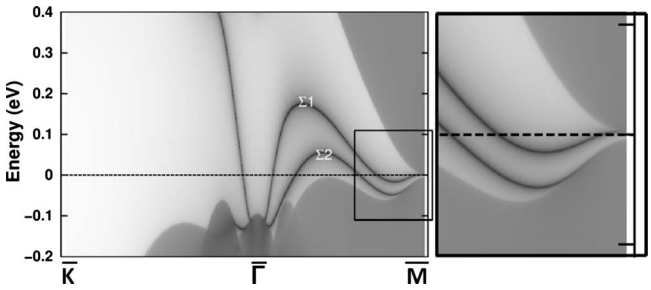

FIG. 5. Left panel: the surface local DOS for λ=1.1 eV. The dark gray regions denote the continuous bulk bands with a small gap of about 10 meV at $\bar{M}$ point, and the white gray regions denote energy bulk gap. Two surface bands (Σ_1 and Σ_2) disperse within the bulk gap. The black dashed line indicates the Fermi energy, which intersects five times with the two surface bands from $\bar{\Gamma}$ to $\bar{M}$. Right panel: the region framed by the black rectangle in the left panel is zoomed in. One surface state (Σ_1) goes up to merge the conduction band while the other one (Σ_2) goes back to the valence band.

The shape of the Fermi surface for the two different phases is plotted in Figs. 7 and 8. There are one $\bar{\Gamma}$ point and three $\bar{M}$ points in the surface BZ, and they are all TRIM. The main difference between Figs. 7 and 8 is around the $\bar{M}$ point. For both cases, the $\bar{\Gamma}$ point is enclosed by one Fermi arc, however, the $\bar{M}$ point is different: it is enclosed by one Fermi arc at λ=1.28 eV, and it is not for λ=1.1 eV. Therefore by counting the total number of TRIM enclosed by Fermi sur-

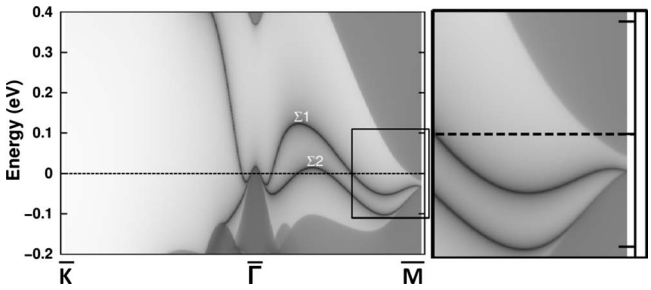

FIG. 6. Left panel: the surface local DOS for λ=1.28 eV. The two surface states (Σ_1 and Σ_2) disperse in a different way in the present case. At the $\bar{M}$ point, both Σ_1 and Σ_2 are connected to the valence band. Therefore the Fermi energy intersects four times with the surface bands between $\bar{\Gamma}$ and $\bar{M}$. Right panel: the region framed by the black rectangle in the left panel is zoomed in.

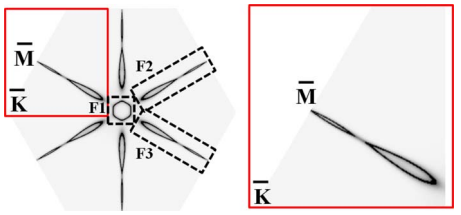

FIG. 7. (Color online) Left panel: the Fermi surface plot for $\lambda = 1.1$ eV. The white gray hexagonal region is the 2D BZ of [111] surface for A7 structure. $\bar{\Gamma}$ is enclosed by a hexagonal electron pocket. There are other six hole pockets and six electron pockets surrounding. Right panel: the region framed by red rectangle in the left panel is zoomed in. We can clearly see that the outermost six small electron pockets do not enclose $\bar{M}$.

face in the BZ, it is even number for $\lambda = 1.28$ eV (one $\bar{\Gamma}$ plus three $\bar{M}$), and odd number for $\lambda = 1.1$ eV (only one $\bar{\Gamma}$ point).

Here we compare our results with that from TB analysis[8] and that from experiment.[10] As shown in Fig. 9(a), we find five crossing points between the surface bands and the Fermi energy along the line from $\bar{\Gamma}$ to $\bar{M}$, this is the same to those observed in the experiment of Hsieh *et al.*[10] (Fig. 9), however, in a simple TB model,[11] the number of crossing is three [Fig. 9(b)]. A small discrepancy is found between our *ab initio* calculation and the experiment of Hsieh *et al.*[10] near $\bar{M}$ point, as shown in Figs. 9(a) and 9(d). In the experiment of Hsieh *et al.*,[10] a third surface band Σ_3 appears near $\bar{M}$ point and be degenerate with Σ_1 band at $\bar{M}$ point, however, in our calculation, there is no such band and Σ_1 band will go up and merge with the conduction band. This discrepancy may come from additional trivial surface states due to the imperfect surface. In our calculations, we only treat the perfect surface without any defects, absorbate, and etc. However, in the presence of surface imperfection, additional trivial surface state as show in Fig. 9(c) may appear and intersect with the nontrivial surface state, this will further complicate the experimental situation. Nevertheless, this possibility is not studied in our present calculations, and the discrepancy remains to be justified by future studies.

In addition to the energy resolution, we are also able to calculate the spin-resolved surface states. As an example, we carry out the surface state's calculation for the semi-infinite $Bi_{1-x}Sb_x$ system's top surface, and show the spin-resolved surface state in Fig. 10. The spin orientation of the surface

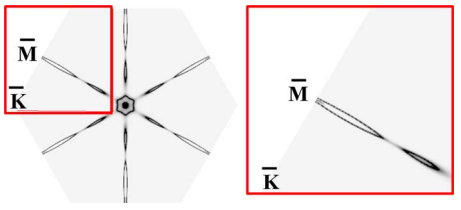

FIG. 8. (Color online) Left panel: the Fermi surface plot for $\lambda = 1.28$ eV, which is similar to Fig. 7, except that the outermost six electron pockets enclose the $\bar{M}$ point. Right panel: the region framed by red rectangle in the upper panel is zoomed in.

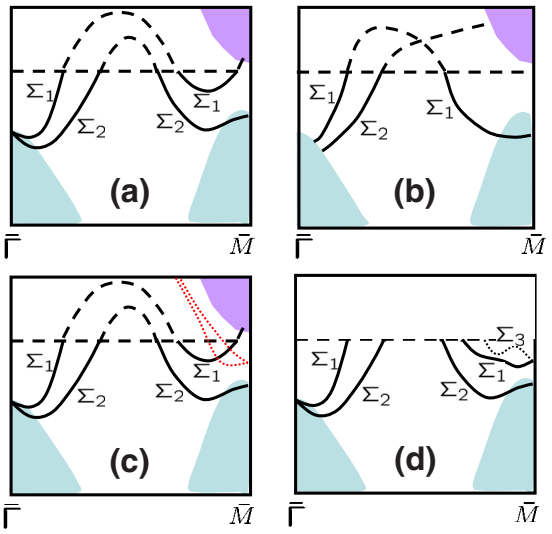

FIG. 9. (Color online) Schematic picture for the comparison of the surface bands obtained from (a) our *ab initio* calculation, (b) TB model [from the work of Teo *et al.* (Ref. 11)], and (d) ARPES experiment results. In ARPES experiment, an additional Σ_3 surface band [dotted line in (d)] becomes degenerate with Σ_2 band at $\bar{M}$ point. This additional band may come from the hybridization between the topological surface states and the other trivial surface states, as suggested by the red dotted line in (c).

states at the Fermi level are plotted for three regions (F1, F2, and F3) of 2D BZ with $\lambda = 1.1$ eV. Please note the magnitude of Sz component is actually very weak (around one percent of in-plane component), we plot the Sz distribution just for the completeness of the results. Clearly, vertex structure is found for the electron pocket around $\bar{\Gamma}$ point (F_1), which confirms the topological nature of surface state. Because top surface's normal is along the $+z$ direction, we can confirm that the chirality of the vertex structure around $\bar{\Gamma}$ is

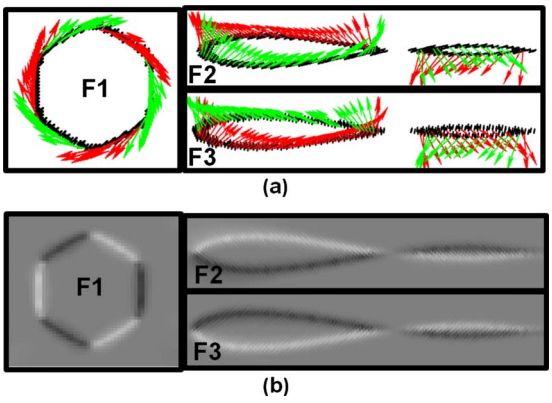

FIG. 10. (Color online) The spin-resolved Fermi surface for the semi-infinite $Bi_{1-x}Sb_x$'s top surface. The arrow in (a) indicates (S_x, S_y); different colors in (a) and (b) represent S_z along different directions. The green color in (a) and dark gray in (b) mean that the S_z is along the $+z$ direction; the red color in (a) and white gray in (b) mean that the S_z is along the $-z$ direction. The three pieces of Fermi surface F_1, F_2, and F_3 are marked in Fig. 7. λ is taken as 1.1 eV here.

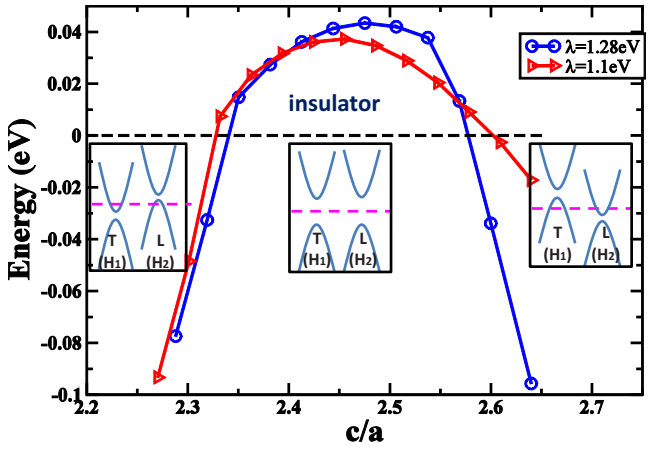

FIG. 11. (Color online) Indirect energy band gap. We define the indirect energy gap as the difference between the CBB and the VBT. The blue line and red line present the indirect energy gap's functions of c/a for $\lambda=1.28$ and 1.1 eV. H_1 and H_2 points are not high symmetry points and locate in the mirror plane in BZ. H_1 is near to T point and H_2 is near to L point. In the inset, we show the schematic figures indicating the different energy band structures.

left handed and agrees with the recent spin-resolved ARPES experiment.[29,30]

C. Indirect band gap

In the above discussions, we call $Bi_{1-x}Sb_x$ bulk as "insulator" because there exists a direct band gap between the conduction and the valence bands. Unfortunately, $Bi_{1-x}Sb_x$ is actually a semimetal (not true insulator) for most of the doping range x, namely, there exists finite overlap between the conduction band bottom (CBB) and the valence band top (VBT). If we define the true gap E_g (indirect gap) as the energy difference between the CBB and the VBT, E_g is negative for most of the x, and it is positive only for $0.07<x<0.22$.[11] On the other hand, to identify the TI nature, except the parity arguments as discussed above, it is crucially important to have a full bulk gap throughout the BZ. Therefore, a serious question for $Bi_{1-x}Sb_x$ is "can we make the indirect gap E_g as positive as possible?" or, in other words, "can we widen the range of doping x where system is truly insulating?" Here we will show that applying the uniaxial pressure is an efficient way to open up the indirect band gap E_g.

Figure 11 shows the calculated indirect band gap E_g as function of c/a ratio with fixed volume. The c/a ratio can be tuned either by c-axis pressure or by forming thin film matched to substrate with different lattice parameters. For both sides of the topological phases ($\lambda=1.28$ eV or 1.1 eV), a broad positive E_g region can be obtained by reducing c/a ratio slightly (around 3% reduction from its experimental value $c/a=2.6$). From the elastic constant C33 measured experimentally for Bi,[31–33] we can estimate that about 1.2GPa pressure will reduce the c/a ratio by such an amount (3%). For both $\lambda=1.28$ and 1.1 eV, the CBB is located at L point, however, the VBT is located at T point for $\lambda=1.28$ eV, and at H_1 point (around T point in the mirror plane) for $\lambda=1.1$ eV. Despite of the different positions of VBT, the effect of c-axis pressure is always to raise the energy levels round L point, and lower the levels around the T point. For $\lambda=1.1$ eV, energy level at H_1 point goes lower, and energy level at H_2 point near the L point in the mirror plane goes upper. Therefore the positive indirect band gap is realized as schematically illustrated in the insets of Fig. 11.

V. CONCLUSIONS

As a summary, we develop a method to study the alloyed $Bi_{1-x}Sb_x$ system and present a phase diagram to describe the topological nature of the system. We show that Bi is topologically trivial because the SOC is too strong. By alloying with Sb, the effective SOC strength is reduced and the topologically nontrivial phase is realized. By accurate *ab initio* calculations and MLWF, we calculate the surface states of semi-infinite system. The results are compared with recent experiments. We predict the spin-resolved Fermi surface which can be tested by spin-resolved ARPES. Finally, we suggest an efficient way to tune the indirect band gap by uniaxial pressure, such that true bulk insulating state can be realized for a broad doping range.

ACKNOWLEDGMENTS

It is a pleasure to thank N. Marzari for solving problems related to the construction of MLWF. This work is supported by the National Science Foundation of China, the Knowledge Innovation Project of the Chinese Academy of Sciences, and the 973 project of the Ministry of Science and Technology of China. X.L.Q. and S.C.Z. are supported by the U.S. Department of Energy, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515.

[1] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).
[2] B. A. Bernevig, T. L. Hughes, and S. C. Zhang, Science **314**, 1757 (2006).
[3] M. Konig, H. Buhmann, L. W. Molenkamp, T. Hughes, C.-X. Liu, X.-L. Qi, and S.-C. Zhang, J. Phys. Soc. Jpn. **77**, 031007 (2008).
[4] L. Fu, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **98**, 106803 (2007).
[5] J. E. Moore and L. Balents, Phys. Rev. B **75**, 121306(R) (2007).
[6] X.-L. Qi, T. L. Hughes, and S.-C. Zhang, Phys. Rev. B **78**, 195424 (2008).
[7] M. Konig, S. Wiedmann, C. Brune, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang, Science **318**, 766 (2007).
[8] L. Fu and C. L. Kane, Phys. Rev. B **76**, 045302 (2007).
[9] X. Dai, T. L. Hughes, X.-L. Qi, Z. Fang, and S.-C. Zhang, Phys. Rev. B **77**, 125319 (2008).
[10] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and

M. Z. Hasan, Nature (London) **452**, 970 (2008).
[11] J. C. Y. Teo, L. Fu, and C. L. Kane, Phys. Rev. B **78**, 045426 (2008).
[12] H. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, Nat. Phys. **5**, 438 (2009).
[13] Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nat. Phys. **5**, 398 (2009).
[14] Q. Liu, C.-X. Liu, C. Xu, X.-L. Qi, and S.-C. Zhang, Phys. Rev. Lett. **102**, 156603 (2009).
[15] Y. Liu and R. E. Allen, Phys. Rev. B **52**, 1566 (1995).
[16] K. Sugawara, T. Sato, S. Souma, T. Takahashi, M. Arai, and T. Sasaki, Phys. Rev. Lett. **96**, 046411 (2006).
[17] P. Hofmann, Prog. Surf. Sci. **81**, 191 (2006).
[18] C. R. Ast and H. Höchst, Phys. Rev. Lett. **90**, 016403 (2003).
[19] Y. M. Koroteev, G. Bihlmayer, J. E. Gayone, E. V. Chulkov, S. Blügel, P. M. Echenique, and P. Hofmann, Phys. Rev. Lett. **93**, 046403 (2004).
[20] C. R. Ast and H. Höchst, Phys. Rev. B **67**, 113102 (2003).
[21] T. Hirahara, T. Nagao, I. Matsuda, G. Bihlmayer, E. V. Chulkov, Y. M. Koroteev, P. M. Echenique, M. Saito, and S. Hasegawa, Phys. Rev. Lett. **97**, 146803 (2006).
[22] Z. Fang and K. Terakura, J. Phys.: Condens. Matter **14**, 3001 (2002).
[23] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[24] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).
[25] I. Souza, N. Marzari, and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2001).
[26] M. P. L. Sancho, J. M. L. Sancho, and J. Rubio, J. Phys. F: Met. Phys. **14**, 1205 (1984).
[27] M. P. L. Sancho, J. M. L. Sancho, J. M. L. Sancho, and J. Rubio, J. Phys. F: Met. Phys. **15**, 851 (1985).
[28] T. Hirahara, K. Miyamoto, I. Matsuda, T. Kadono, A. Kimura, T. Nagao, G. Bihlmayer, E. V. Chulkov, S. Qiao, K. Shimada, H. Namatame, M. Taniguchi, and S. Hasegawa, Phys. Rev. B **76**, 153305 (2007).
[29] D. Hsieh, Y. Xia, L. Wray, D. Qian, A. Pal, J. H. Dil, J. Osterwalder, F. Meier, G. Bihlmayer, C. L. Kane, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Science **323**, 919 (2009).
[30] A. Nishide, A. A. Taskin, Y. Takeichi, T. Okuda, A. Kakizaki, T. Hirahara, K. Nakatsuji, F. Komori, Y. Ando, and I. Matsuda, arXiv:0902.2251 (unpublished).
[31] Y. Eckstein, A. W. Lawson, and D. H. Reneker, J. Appl. Phys. **31**, 1534 (1960).
[32] A. J. Lichnowski and G. A. Saunders, J. Phys. C: Solid State Phys. **9**, 927 (1976).
[33] J. Dominec, J. Phys. C: Solid State Phys. **18**, 5793 (1985).