Pseudo-potential band structure calculation of InSb ultra-thin films and its application to assess the n-metal-oxide-semiconductor transistor performance

Zhen Gang Zhu1, Tony Low1, Ming Fu Li1,2, Wei Jun Fan3, P Bai1, D L Kwong5 and G Samudra1

1 Silicon Nano Device Lab (SNDL), ECE Department, National University of Singapore (NUS), Singapore
2 Fudan University, Shanghai 201203, People’s Republic of China
3 School of EEE, Nanyang Technological University of Singapore, Singapore
4 Institute of High Performance Computing, Singapore
5 Institute of Microelectronics, Singapore

E-mail: eleshanr@nus.edu.sg

Received 14 October 2007, in final form 15 October 2007
Published 4 January 2008
Online at stacks.iop.org/SST/23/025009

Abstract

The band structure of InSb thin films with $\langle 100 \rangle$ surface orientation is calculated using the empirical pseudopotential method (EPM) to evaluate the performance of nanoscale devices using a InSb substrate. Contrary to the predictions by simple effective mass approximation methods (EMA), our calculation reveals that the $\Gamma$ valley is still the lowest lying conduction valley. Based on EPM calculations, we obtained the important electronic structure and transport parameters, such as effective mass and valley energy minimum, of InSb thin film as a function of the film thickness. Our calculations reveal that the ‘effective mass’ of $\Gamma$-valley electrons increases with the scaling down of the film thickness. We also provide an assessment of nanoscale InSb thin film devices using a non-equilibrium Green’s function under the effective mass framework in the ballistic regime.

1. Introduction

An ultra-thin body (UTB) metal-oxide-semiconductor fieldeffect transistor (MOSFET) structure is a promising candidate for scaling MOSFET devices into the nanometer regime because it has the excellent attribute of suppressing various short channel effects caused by the downsizing of the device gate length [1]. Recently, there have been experimental and theoretical efforts to evaluate the use of non-conventional channel orientation [2–4] or new channel materials (such as Ge, III–V compound semiconductors, due to their small $\Gamma$-valley electron masses) [4–6] to improve the MOSFET performance.

In particular, InSb is being considered as a new channel material [6] for state-of-the-art MOSFET devices because of its high bulk mobility (in fact, it has the highest electron and hole mobility among common III–V semiconductors [7, 8]). A theoretical investigation of InSb MOSFET devices based on a simple effective mass approximation (EMA) had been conducted in [9] to predict its device performance limits. It was predicted that the $\Gamma$ valley energy rises rapidly under body quantization, and the inversion charges in the thin film are transferred to the L valley. Therefore, the advantage of
the high injection velocity from the Γ-valley electrons is lost. Work in [9] used the bulk effective mass even in thin film regime. However, EMA may not be a reliable method in describing the electronic band structure of thin films due to unaccounted effects such as band coupling and non-parabolic dispersion. We also found that the bulk effective mass can only describe the energy dispersion in a very small range of \( k \)-space, which brings to question how accurately EMA can capture the quantization effect in a thin film (see the discussions on figure 7 in this paper).

In this work, we describe a more physically reliable and accurate model, the local empirical pseudopotential method (EPM) [10, 11] to study InSb thin films. Our results show that the charge transfer from the \( \Gamma \) to the L valley does not occur and the effective masses in transverse directions become larger than those in bulk material, which have a direct impact on the device transport properties. This paper serves to communicate in detail the models employed in our published conference paper [12]. In section 2, we give a detailed description of the theoretical background of EPM. We will discuss the introduction of a model potential to describe the atomic pseudopotential, which allows us to extend EPM to thin film calculations. In subsection 2.2, we derive the spin–orbit coupling contribution to the matrix element of the Hamiltonian used in the EPM calculation. In order to achieve accurate results in EPM calculations of thin film band structures, we introduce two parameters to take into account the volume renormalization and the spin–orbit coupling. In subsection 2.3, we discuss the methodology adopted to passivate the surface dangling bonds by using hydrogen (H) atom bonding. In section 3, we discuss the important features of the InSb thin film electronic structure. We also obtained important electronic parameters, such as effective mass and valley energy minimum of the InSb thin film as a function of the film thickness. In section 4, we provide an assessment of nanoscale InSb thin film double-gated MOSFET devices using a non-equilibrium Green's function under the effective mass regime. However, EMA may not be a reliable method in small band gap materials like InSb. Therefore we would also need to determine the pseudopotential of hydrogen atoms.

2. Theoretical background

The EPM has already been successfully applied in band structure calculations of metal, semiconductor or other materials [13–15]. It has also been extended to calculate band structures and electronic properties of Si quantum dots [16, 17], quantum wires [18, 19] and quantum films [20, 21]. In view of pedagogical clarity, we shall also give an overview of the essential theoretical aspects of EPM. Before going into the rigorous mathematical details, we shall first give an overall description of the method used in this work. The key aspects of applying the EPM to a thin film calculation are as follows. (1) A supercell (figure 1) has to be specified prior to any calculation. A sufficiently large vacuum layer has to be explicitly included in the supercell so as to eliminate any wavefunction overlap between different thin film layers. (2) We have to determine the suitable atomic pseudopotentials for the Sb and In atoms independently for our thin film calculations. (3) The effect of spin–orbit coupling needs to be included because it is sufficiently large to affect the band structure significantly (see table 3) in InSb and many other compound semiconductors. (4) Hydrogen passivation of the surface dangling bonds is incorporated into model potential. This is a very crucial treatment as the surface dangling bonds will lead to surface states in the band gap which will overwhelm the band structure of the thin film, especially small band gap materials like InSb.

2.1. Pseudopotential method

2.1.1. Empirical pseudopotential of bulk InSb. It is well known that the band structure can be derived by solving a secular equation [7]. The matrix element of the pseudopotential (PS) is

\[
\langle PW, G \rangle \hat{V}^{PS} \langle PW', G' \rangle = \sum_\tau \exp(-i(G - G') \cdot \vec{\tau}) U^{PS}_\tau(G - G'),
\]

where

\[
U^{PS}_\tau(G) = \frac{1}{\Omega} \int_{\Omega} U^{PS}_\tau(r) \exp(-iG \cdot r) \, dr,
\]

\( \Omega \) is the volume of the unit cell, \( N \) is the number of unit cells, \( G \) and \( G' \) are the reciprocal lattice vectors, \( U^{PS}_\tau(G) \) is the form factor and \( \langle PW, G' \rangle \) is the plane wave basis function. In a single crystal, there are usually several basis atoms associated.
validity. For InSb, we use a Table 1.

\[
\Delta G = \cos(\theta') - \sin(\theta')
\]

The symmetry and anti-symmetry structure factors are given, the difference of the individual atomic PSs of Sb and In. as h ow ni nt a b l e1) can be defined as the summation and empirical symmetry and anti-symmetry form factors (ESAFF), symmetry and anti-symmetry form factors for InSb as US.

Before we proceed with the calculation of the InSb thin film band structure, we need to derive the atomic pseudopotential (PS) of In and Sb atoms, respectively, or more accurately speaking, to construct an empirical PS. We can reliably be calibrated against experimental data. For bulk InSb, there are two atoms in a unit cell. Conventionally, bulk empirical PS (at special \( q^2 \) values, they are also known as empirical symmetry and anti-symmetry form factors (ESAFF), as shown in table 1) can be defined as the summation and the difference of the individual atomic PSs of Sb and In. The symmetry and anti-symmetry structure factors are given, respectively, as \( S^S = \cos(\theta') \) and \( S^A = \sin(\theta') \), where \( \theta' = \Delta G = G - G' \). It should be noted that if we define the symmetry and anti-symmetry form factors for InSb as \( U^S = (1/2)(U_{ss}^{PS} + U_{nn}^{PS}) \) and \( U^A = (1/2)(U_{ss}^{PS} - U_{nn}^{PS}) \), then the factor \( 1/\Omega \) appearing in equation (2) will be \( 1/\Omega_{atomic} \), where \( \Omega_{atomic} \) is the atomic volume in a unit cell.

By performing the EPM calculation of the bulk band structure iteratively, a suitable set of ESAFF is obtained such that it yields the ‘correct’ energy gaps at pertinent symmetry points [7, 8]. This set of ESAFF is shown in table 1.

Because of the spherical symmetry approximation of the local pseudopotential, the first few lowest energy shells are equal to 3, 4, 8, 11 and 12 at normalized \( q^2 \), where \( q \) is the wave vector in the unit 2\( \pi/a \) and \( a \) is a crystal lattice constant. Based on this set of ESAFF, we can devise a suitable interpolation scheme that allows us to reasonably predict a pseudopotential value at other \( q^2 \) values. This will then allow us to extend this EPM method to the calculation of the thin film band structure of an arbitrary thickness. An essential assumption we are invoking is that the atomic pseudopotential is transferable and thus will not be changed from the bulk to a thin film. This assumption is reasonable because its \textit{ab initio} calculation has demonstrated that the bulk potential will just be changed in only one atomic layer at the interface for the GaAs/AlAs superlattice [22] and will be exactly the same for all other atoms. Hence, the potential will be bulk-like potential except at one atomic layer at the surface of the thin film. Therefore, the bulk atomic potential can be used in the thin film calculation. However, further investigation reveals that we must add the renormalization factor to the pseudopotential to give the correct results for the thin film as shown in subsection 2.1.2.

2.1.2. Thin film and empirical model potential (EMP). For the (100) surface InSb thin film band structure calculation, we use a supercell as shown in figure 1. Figure 2(a) depicts the first Brillouin zone of bulk InSb and its subsequent projection onto the \( xy \) plane, i.e. the first Brillouin zone of the InSb thin film in 2D is shown in figure 2(b). In this paper, the
band structure of the InSb thin film will be plotted using the same coordinate system for the 2D reciprocal vector space as illustrated in figure 2(b).

In bulk InSb, the band structure is determined by the pseudopotential at several $q^2$ values as stated in the previous section and the shells with $q^2 < 3$ are not of much importance. For thin film InSb, the reciprocal primitive vector will be shorter than that of bulk and thus the radius of the lowest energy shell is smaller. This is the consequence of a larger primitive vector in the real space for (1 0 0) film (see figure 1). Therefore, the shells with $q^2 < 3$ are also very important for the band structure calculation and their effects must be properly predicted and included in our calculation. To account for them, we shall employ an empirical model potential that can give the continuous pseudopotential as a function of $q$. We use a model potential [16, 17, 19, 21]

$$V_{PS}^{at} = \frac{a_1(q^2 - a_2)}{a_3 \exp(a_4q^2)} - 1,$$

where the subscript 'atom' is used to distinguish the different atomic species’ pseudopotential. $a_1$, $a_2$, $a_3$ and $a_4$ are the fitting parameters such that they yield the correct form factor as determined by ESAff.

To obtain a unique atomic model potential of In and Sb, we need eight constraints for the eight fitting parameters. Seven constraints come from the seven nonzero ESAff parameters at $q^2$ equal to 3, 4, 8, 11 and 12 shown in table 1. The remaining one constraint is imposed by the InSb work function (WF), which gives one nonzero ESAff at $q^2$ equal to 0 [21]. This last constraint accounts for the surface property, which is important for a thin film [21]. In fitting these parameters, it should be noted that the physical unit of $q^2$ in equation (3) is $(2\pi/\alpha)^2$ and $V_{PS}^{at}$ is in the Rydberg atomic unit. The best-fitting parameters for the model potential are shown in table 2 and its variation with $q^2$ is shown in figure 3. In the fitting parameters to pseudopotentials, $a_1$ and $a_2$ for indium atom are very large numbers. However, we accepted these numbers as the fitting parameters giving an excellent fit to the discrete pseudopotentials (see figure 3 and table 1) and similar large values of these parameters also occur in previously reported calculations for other III and V materials [23]. From figure 3, we observe that the model potentials of Sb and In atoms are in a reasonable range. Also, the fits to $U^S$ and $U^A$ are quite good as is clear from table 1.

We perform the EPM calculation of bulk InSb using this new model potential and reproduce the required energy band minima derived from experiment, as tabulated in table 3. We shall point out that the model potential of one of the species of an atom in a semiconductor compound is actually the atomic model potential, i.e. the volume of the unit cell

### Table 2. Derived parameters for model potentials of atoms Sb and In.

| Parameters | Sb     | In     |
|-----------|--------|--------|
| $a_1$     | 0.2588 | 719.470|
| $a_2$     | 1.5832 | 2.0811 |
| $a_3$     | 1.9689 | 3.8136 |
| $a_4$     | 0.7159 | 0.9116 |

is divided by the number of atoms in this unit cell. So for the thin film calculations, there will be a renormalization with a different unit cell volume under the assumption that the overall pseudopotential of atoms is the superposition of the local pseudopotential of all the atoms in the thin film.

#### 2.2. Hydrogen passivation

In the thin film calculation, we must ensure that surface dangling bonds are correctly passivated with hydrogen (H) so that there will be no surface states within the band gap. To determine model potential for hydrogen, we tried model potential of H in [16, 19, 21] and calculated the band structure of eight atomic layer (atm) Si (1 0 0) thin film and compared the EPM results with those derived from ab initio calculation [24] (figure 4). The band structures obtained from both methods are very consistent, thus establishing the validity and reliability of this H model potential. In Si (1 0 0) thin film calculation, the renormalization of the supercell volume is set as a ratio of the number of H atoms in the supercell and the number of Si atoms in the same unit cell. Also we assumed that the atomic volume of H is the same as that of Si.

For the InSb thin film, the bond length of In–H is set as 1.75 Å (from experiment in [25]) and Sb–H as 1.711 Å (from ab initio energy minimization calculation [24], which is very close to 1.67 Å in [26]). The angle between the bond of In–H and the top surface (θ) (figure 1) should be determined carefully. Figure 5(a) shows the calculated direct InSb band gap ($E_G$) as a function of θ for a varying number of InSb
Figure 4. Energy bands of Si ⟨100⟩ thin film (eight atomic layers Si film + eight atomic layers vacuum with H passivation) calculated with \textit{ab initio} method (using CASTEP, local density approximation) and via the empirical pseudopotential method with a model potential. An excellent agreement for the various conduction valleys minima is obtained.

atomic layers. When the dangling bonds cannot be passivated completely by hydrogen, the surface states will appear in the band gap. As the energy gap is defined by the conduction band minimum and the valence band maximum, it varies dramatically in figure 5(a) when surface states are present. However, in the region of angles where the dangling bonds are mostly passivated, the variations are reasonable. For example, for 8 atm (1.3 nm), the biggest variation is 0.09 eV in \( \theta \in [0.6, 1.01] \); for 28 atm (4.5 nm) thick film, the biggest variation is 0.025 eV in the same range. The variations of band gap with the angle in these mostly passivated films are physically meaningful. In conclusion, the dramatic variation in band gap in figure 5(a) is, in fact, an indication of the presence of surface states and incomplete passivation.

To determine the stationary angle, the safest way is to calculate the total energy and find the minimum. The angle corresponding to the minimum is the stationary angle. However this procedure needs \textit{ab initio} calculation in which the model potential, the stationary angle, the bond length of H and other atoms and spin–orbit parameters should be derived with the final band structure simultaneously for an ultra-thin film. It is practically very difficult to determine so many quantities simultaneously even if we leave aside computational time issue. To overcome such difficulties, we shall determine the angle in a rough method which is reasonably justified through the following considerations.

(i) We choose the angle satisfying the condition \( \partial E / \partial \theta = 0 \), which reflects that extremum occurs here and varies with \( \theta \) adiabatically. This means that the rate of changing of \( E \) caused by varying of angle is zero at this angle and infinitesimal around this angle. And in principle, the realistic band gap should be present at the angle where there is no influence of varying angle on it. Thus this angle with the zero rate is more reasonable to be the stationary one than others in considering range of angles. Figure 5(a) confirms \( \theta = 0.77 \) rad as such a stationary value of \( \theta \) in this way. (ii) In a simplified and qualitative interpretation, the total energy is the summation of the energies of all electrons in all the occupied states from the bottom of the whole band to the Fermi level (if the temperature is zero). If we assume that the width between the valence band maximum and the bottom of the whole band is fixed with varying angle, a larger band gap may lead to a smaller total energy. (iii) Once surface states are eliminated, energy gap variations are relatively insensitive to the angle variation. For example, if we use a bond angle of 39° used in the symmetric Si(100) dihydride (we also use this configuration for passivation) through \textit{ab initio} calculation [27], the difference of energy gaps to \( \theta = 0.77 \) rad case is about 0.02 eV for 8 atm and about 0.005 eV for 28 atm. (iv) \( \theta = 0.77 \) rad is very close to \( \pi/4 \). Actually, \( \theta = 0 \) (in the plane) and \( \theta = \pi/2 \) (perpendicular to the plane) are both unstable angles, leaving the surface states

Figure 5. (a) The calculated InSb direct energy gap \( (E_G) \) as a function of H–In bond angle \( \theta \), plotted for InSb thin films of varying number of atomic layers (atm). The graph illustrates that there is a common \( \theta \) where the energy function is a stationary point. (b) The direct energy gap \( (E_G) \) at the bond angle \( \theta \), where the energy function is a stationary point, plotted as a function of the film thickness. An expression which fits these data points is given. It gives correct asymptotic behavior when the film thickness tends to bulk, yielding bulk band gap value of 0.43 eV.
incompletely passivated. The reasonable one may be around \( \pi/4 \) geometrically.

A fitting formula (similar to [16, 19, 21]) affirms that \( E_G \) approaches the bulk band gap of 0.43 eV as the film thickness tends to the bulk limit as shown in figure 5(b) (0.43 eV is the bulk band gap of InSb without the spin–orbit contribution, while it is 0.17 eV when the spin–orbit interaction is considered).

For the InSb thin film, we need to renormalize hydrogen atomic potential derived from the Si thin film calculation, \( V_H^P(\text{InSb}) = (\Omega_{\text{InSb}}/\Omega_1)^{\alpha} V_H^P(\text{Si}) \) using relative volume of \( H \) atoms compared to the volume of InSb, where \( \Omega_{\text{InSb}} \) and \( \Omega_1 \) are the supercell volumes of Si and InSb respectively. Since the InSb thin film has the same supercell structure as the Si thin film, its supercell renormalization is performed using their respective lattice constants as \( (\Omega_{\text{InSb}}/\Omega_1) = (a_{\text{Si}}/a_{\text{InSb}})^3 \).

2.3. Spin–orbit coupling (SOC)

SOC treatment is incorporated by calculating the matrix element of the spin–orbit coupling term in the Hamiltonian expressed in a plane wave representation. It can be reduced to an expression similar to the pseudopotential form factor, known as the spin–orbit form factors (SOFFs) [28–31]. In this section, we will clarify the formulation used for the SOC treatment. A more mathematically detailed derivation shall be presented in the appendix.

If we consider a supercell composed of different species of basis atoms at each lattice site, the spin–orbit Hamiltonian can be specified as

\[
H_{kx,ks}^{\text{SO}} = -i \sigma_{x,s} \cdot (\vec{k}' \times \vec{k}) \sum_{j=1}^{m} S_j (\Delta \vec{k}) \lambda_j,
\]

where \( \lambda_j \) is a parameter for \( j \)th species, \( m \) is the number of species, \( \Delta \vec{k} = \vec{k}' - \vec{k} \) is the change in the wave vector, \( \sigma_{x,s} \) are the matrix elements of Pauli matrices, \( s \) and \( s' \) denote the spin states; up spin or down spin. The structure factor is given as

\[
S_j (\Delta \vec{k}) = n_{\text{num}} \sum_{\text{atom}} \exp(i \cdot \Delta \vec{k} \cdot \vec{r}_j),
\]

where the summation is over \( n_{\text{num}} \) atoms of \( j \)th species (see the appendix for more detail).

The parameters \( \lambda_j \) are determined by the \( B \) function defined in the appendix (see equation (A.19)). The spin–orbit parameter \( \eta \) (see equation (A.15)) is a property of isolated atoms, so it is independent of the crystal wave vector. The defined \( B \) function is important in the calculation of \( \lambda_j \) and its behavior is depicted in figure 6.

In figure 6, we plot the variation of the \( B \) function (i.e. Equation (A.19)) as a function of wave vector for Si and In atoms in (a) and the difference of the \( B \) function between these two types of atoms in (b). For our plot in figure 6, we just show the outermost core state, i.e. \( n = 4 \) and \( l = 1 \) (\( p \) orbit). Because of the normalization, the \( B \) function tends to 1 as \( k \rightarrow 0 \).

Although the explicit form of the spin–orbit coupling matrix element is presented in the appendix, there are some additional steps that must be taken into consideration in order to obtain the correct result. Thus, we shall briefly elaborate the procedures involved in the SOC treatment. First of all, the SOFFs are obtained via a similar methodology as EMAFF; by iterating our EPM calculations until the correct energy gaps are obtained for bulk InSb. SOFFs have to be normalized with respect to supercell volume when used for a thin film calculation. Besides the usual volume renormalization, one has to account for another renormalization. Conventionally, the bulk band structure of InSb is derived by considering a two-atom unit cell in which the primitive vectors are non-orthogonal and the mutual angles between them are all \( \pi/3 \). In thin film calculation, a larger supercell is constructed in which the primitive vectors are mutually orthogonal. Hence this angular effect renormalization and volume renormalization must be considered in our treatment.

The spin–orbit Hamiltonian is divided into two parts; one part dealing with the angular part and other the magnitude part. When the supercell is changed from bulk 2-atom unit cell to \( \{100\} \) thin film supercell, the volume renormalization will appear in the second part like what has been done on form factors of pseudopotential of Si and In. For the angular part renormalization, we treat it as an unknown parameter which we derived using the following procedure. We removed the vacuum slab from the thin film supercell. This new supercell is therefore a unit cell for a bulk crystal. We then adjust the angular part renormalization parameter so that it reproduces the correct bulk band structure and spin–orbit splitting (as calculated from a conventional bulk crystal unit cell). This procedure has to be repeated for each film thickness. After this renormalization, the vacuum slabs are replaced again and the thin film band structure can be derived from EPM.

3. Features of InSb electronic structures

Figure 7(a) shows the calculated band structure of bulk InSb from our EPM. We derived the following effective masses: \( m_L = 2.45 m_0 \) (the longitudinal mass of \( L \) valley), \( m_X = 3.90 m_0 \) (the longitudinal mass of \( \Delta \) valley) and \( m_F = 0.016 m_0 \) (the isotropic mass of \( \Gamma \) valley). The open circles in figure 7(a) are data points fitted to the band minima by using a parabolic dispersion with effective masses as stated. Figure 7(b) shows the ‘effective mass’ in the vicinity of \( \Gamma \) valley (\( m_F \)) calculated by taking the second derivative of energy with respect to the
wave vector $k$. We note that the parabolic assumption for the energy dispersion at $\Gamma$ is only valid for a very small $k$ range. Consider a 1.3 nm InSb film, the wave vector spread according to the uncertainty principle is $\sim 0.04(2\pi/a)$. From the plot, the parabolic assumption for EMA is only applicable up to a $k$ range of $\sim 0.01(2\pi/a)$. Hence, in the ultra-thin film regime, a parabolic EMA is not a reasonable assumption and becomes a highly unreliable method for the calculation of size quantization effect. In addition, the small band gap also entails a considerable amount of coupling between the conduction and valence bands, which renders the uncoupled EMA approach unreliable.

Figure 8 shows the calculated band structure of thin film InSb from our EPM. These thin films still retain the direct band gap properties in contrast to the EMA calculation [9], which has $L$ valley as the lowest lying conduction valley when the film thickness is below 5 nm. Figure 9(a) shows the variation of the energy minima at $\Gamma$, $L$ and $\Delta$ valleys as a function of the film thickness. We also plot the fitting curves for the energy minima as a function of the thickness $x$ in this figure. The fitting formulae are $0.168 \times 2 + 150/(x + 9.7)^{1.83}$, $1.033 \times 95 + 92/(x + 5.5)^{1.94}$ and $0.704 \times 39 + 94/(x + 5.2)^{1.94}$. There was no crossing of the energy minima down to six atomic layers, in contrast to the predictions by EMA methods. In addition, we note that the lowest lying $\Gamma$ valley is separated from the other valleys with a gap more than 0.3 eV, signifying that electrons are dominantly occupying $\Gamma$ valley. Another startling contrast with results of EMA is the band gap. Although, the band gap is enhanced by the quantization effect, this increase with the reducing thickness is much slower than the well-known $d^{-2}$ behavior predicted by EMA in an infinite deep well model (e.g. at the $\Gamma$ valley, see figure 9(a)). However, EPM also yields the result that the conduction band quantization effect is larger than that of valence band in agreement with prediction by EMA (figure 9(b)). From a logic device point of view, one would desire to have a larger band gap to curb the increasing band-to-band tunneling current with each new generation CMOS devices. The fact that InSb is a direct band gap material may aggravate the problem. However, if one can achieve a sufficiently thin film with $\sim 8$ atm, which offers a band gap more than 1 eV, band-to-band tunneling current should still be tolerable.

Figure 10 shows the contour plot of the conduction and valence band energy dispersions for eight atomic layer InSb thin film. We confirm that the $\Gamma$ valley in thin film still retains its isotropy unlike the case of Si [4]. Figure 11 shows the Si and InSb thin film effective mass fitted from their 2D energy dispersion. The InSb $\Gamma$ valley is isotropic with increasing effective mass as the film thickness is scaled down. The Si $\Gamma$ valley in thin film actually is originated from the bulk $\Delta$ valley, projected onto the 2D $k$-space. We observe that anisotropy in Si becomes more prominent with the decrease of the film thickness, with the $\Gamma K$ direction effective mass diverging for each of the two degenerate bands. The larger $\Gamma K$ effective mass in Si is for the lower band of the two degenerate bands, and the smaller $\Gamma K$ effective mass corresponds to the top band.
Figure 9. (a) Variation of the band gap with the number of atomic layers. (Top scale shows the thickness of thin film in nm in (a) and (b)). (b) The InSb thin film valence band maximum (VBM) and conduction band minimum (CBM) as a function of the number of atomic layers. Bulk values are approached at thicker films.

Figure 10. Energy contour plot for eight atomic layer InSb thin film for the first conduction and first valence band are plotted in (a) and (b), respectively. Values of the energy are indicated on each contour line. Important symmetry points are also indicated. It is apparent that the \Gamma\textsubscript{1} conduction valley is relatively isotropic in nature even in the ultra-thin film regime.

Figure 11. Electron effective mass of Si and InSb \Gamma\textsubscript{1} conduction valley. The inset shows the constant energy ellipsoid for Si projected onto the 2D \textit{k}-space, with the \Gamma\textsubscript{1} valley being doubly degenerate.

In fact, we expect the effective mass to also increase with wave vector as depicted in figure 7(b). However, even if effective mass is not a rigorously derived quantity due to non-parabolic valleys, it serves as a very useful ‘figure of merit’. Interestingly, the isotropic effective mass of InSb \Gamma\textsubscript{1} valley increases with the decrease of the film thickness. At \sim 1 nm of the InSb film, the effective mass is \sim 0.1m_{0}. Increase in \textit{m}_{\Gamma\textsubscript{1}} would retard the quantization effect in the InSb thin film. The isotropy of the electron mass, which translates to the isotropy of the electron transport property, should be advantageous as it affords engineer with more flexibility in orientating the n and p-MOSFET devices to yield the most optimum transport direction on the same substrate.

4. Ballistic limit of InSb nanoscale devices

We calculated the InSb thin film devices’ ballistic current limit using the non-equilibrium Green’s function (NEGF) method [32], under the framework of simple effective mass method. Since the energy dispersion for (100)-surface InSb thin film is relatively isotropic (figure 10), one can employ a decoupled 2D treatment to the problem and calculation is done in the framework of the mode–space NEGF approach [32]. The effective masses used for the various InSb devices are as
derived in figure 11. For Si devices, due to the anisotropy, we simply assumed an effective mass of 0.20nm. A double-gated device structure shown in figure 12 is employed. A channel length of 20 nm is used. An EOT of 1 nm and a metal gate is employed. No wavefunction penetration into oxide is assumed. The N-type source/gate is employed. No wavefunction penetration into oxide is assumed. The N-type source/gate is employed. No wavefunction penetration into oxide is assumed.

The boundary conditions for the potential at the contact are set to floating, by imposing the condition that the potential yields a slightly larger ballistic current compared to their InSb counterparts. The main reason is due to the larger density-of-states mass in Si and the fact that it is doubly degenerate.

The drive current increases with the decrease in the film thickness. A NEGF of NanoMOS [32] from Purdue Computational Electronics group for this work.

Appendix A. Treatment of spin–orbit coupling

The first starting point is the formulae in [28], such as equation (A.3) in it. To explicitly write down the formula of SOC, we shall calculate one of the terms of equation (A.4) in [28] denoted as (II) here. We consider

\[
(II) = - \sum_{l} \langle \hat{k} | b_l | \Lambda \rangle \langle \Lambda | \hat{k} \rangle \\
= - \frac{1}{N_{cell} \Omega_{cell} \sum_{n\ell m} \int d\vec{r}_1 \int d\vec{r}_2 \int d\vec{r}_3 \, c^{-\vec{r}_1 \cdot \vec{r}_2} \Psi_{n\ell m} \times \| (\hat{\vec{r}}_1 - \vec{R}_j - \vec{t}_1), \theta_1, \phi_1 \rangle \Psi^{*}_{n\ell m} (\hat{\vec{r}}_2 - \vec{R}_j - \vec{t}_2, \theta_2, \phi_2) \times \xi \langle \hat{\vec{r}}_3 - \vec{R}_j - \vec{t}_3 \rangle \delta (\hat{\vec{r}}_2 - \vec{r}_1) e^{\vec{r}_1 \cdot \vec{r}_2}. \tag{A.1}
\]

where \( n, l, m \) are the quantum numbers characterizing the core states, \( t = n l m \), \( \vec{R}_j \) is the point located in the unit cell, \( \vec{t}_i \) are the coordinates of atoms in the basis, \( \xi (|\vec{r}|) = \sum_{n l m} V_{n l m} e^{-\vec{r} \cdot \vec{r}} \), with \( V \) being the potential of atomic nuclei, \( \vec{t} \) is the angular momentum operator, \( \delta \) is the Dirac \( \delta \) function, \( \Omega_{cell} \) is the volume of the unit cell and \( N_{cell} \) is the number of unit cells in the crystal. By performing the integral over \( \vec{r}_3 \), and then let \( \vec{r}_1 = \vec{R}_j - \vec{t}_1 \) and \( \vec{r}_2 = \vec{R}_j - \vec{t}_2 \). Then we have

\[
(II) = - \sum_{l=nlm} \left[ \frac{1}{N_{cell} \Omega_{cell} \sum_{i} \int d\vec{r}_1 \, e^{-\vec{r}_1 \cdot \vec{r}_2} \Psi_{n\ell m} (\vec{r}_1, \theta_1, \phi_1) \times \left\langle \int d\vec{r}_2 \, e^{\vec{r}_1 \cdot \vec{r}_2} \Psi^{*}_{n\ell m} (\vec{r}_2, \theta_2, \phi_2) \times \xi \langle \vec{r}_3 - \vec{R}_j - \vec{t}_3 \rangle \delta (\vec{r}_2 - \vec{r}_1) e^{\vec{r}_1 \cdot \vec{r}_2} \right\rangle \right]. \tag{A.2}
\]

We denote the first square bracket in equation (A.2) by [A], the second one by [B] and the third square bracket by [C]. Each of these terms will now be individually analyzed. [A] includes the local position information of all atoms in a unit cell, as given by

\[
[A] = \frac{1}{\Omega_{cell}} \sum_{i} e^{\Delta \vec{k} \cdot \vec{r}_i}. \tag{A.3}
\]

where \( \Delta \vec{k} = \vec{k} - \vec{k} \) is the change in the wave vector. When we consider a single-element crystal, there is only one kind of atom. If we set \( n_{cell} \) as the number of all atoms in a unit cell, then we have

\[
[A] = \frac{1}{n_{cell}} \sum_{i} e^{\Delta \vec{k} \cdot \vec{r}_i} = \frac{S(\Delta \vec{k})}{\Omega_{atom}}, \tag{A.4}
\]

where

\[
S(\Delta \vec{k}) = \frac{1}{n_{cell}} \sum_{i} e^{\Delta \vec{k} \cdot \vec{r}_i} \tag{A.5}
\]

is the counterpart of the structure factor in [28] and \( \Omega_{atom} \) is the atomic volume in a unit cell so that \( \Omega_{atom} = \Omega_{cell}/n_{cell} \).

5. Conclusion

The band structure of III–V material InSb thin films is calculated using the empirical pseudopotential method (EPM). Γ valley in InSb remains the lowest lying conduction valley (a desirable trait if high mobility characteristic is required) despite size quantization effects but its isotropic effective mass increases with the decrease in the film thickness. A NEGF calculation of the InSb and Si double-gated devices reveals that they have comparable ballistic drive current.
The second square bracket $[B]$ in equation (2.2) is

$$
[B] = \sum_{l=0}^{\infty} C_l \int d \vec{r} j_l(k' r) Y_{l k}(\theta, \phi) R_{nl}(r) Y_{l m}^*(\theta, \phi) \frac{d \vec{r}}{r} \times \int_0^{2 \pi} \sin \theta Y_{l k}^*(\theta, \phi) Y_{l m}(\theta, \phi) \frac{d \phi}{2 \pi}
$$

(A.6)

where $C_l = (-1)^l i^l [4 \pi (2l + 1)]^{1/2}, Y_{lm}^*$ are the spherical harmonics and $j_l$ are the spherical Bessel functions. We use the formula $e^{i \vec{r} \cdot \vec{r}'} = \sum_{l=0}^{\infty} i^l [4 \pi (2l + 1)]^{1/2} j_l(k'r) Y_{l 0}(\theta, \phi)$ and $Y_{l 0}(\theta, \phi)$ are the spherical harmonics with the rotational index $m = 0$ in any coordinate system with $z$ in the $z$-direction. Because of the orthogonality of spherical harmonics, the second square bracket in equation (A.6) gives $\delta_{ll'} \delta_{mm'}$, i.e., $\delta_{ll'} \delta_{mm'}$. Hence equation (A.6) simplifies to

$$
[B] = (-1)^l \Omega^{-1/2} B_{nl1}(\vec{k'}) \delta_{nl0},
$$

(A.7)

where $B_{nl1} = \Omega^{-1/2} \int d \vec{r} e^{i \vec{r} \cdot \vec{r}'} [4 \pi (2l + 1)]^{1/2} j_l(k'r) R_{nl}(r) r^2 \frac{d \vec{r}}{r}$ is the counterpart of the $B$ function defined in [28] and $R_{nl}$ is the radial part of core wavefunction.

The third square bracket $[C]$ in equation (2.2) as given by

$$
\Omega^{-1/2}[C] = \Omega^{-1/2} \int d \vec{r} \psi_{nlm}^+(\theta, \phi) \xi(r) \xi(r) \frac{d \vec{r}}{r}
$$

(A.8)

where the angular part is

$$
\langle Y_{lm}^* \mid Y_{lk} \rangle = \frac{\delta_{ll'} \delta_{mm'}}{r}.
$$

(A.9)

Here $m$ in spherical harmonics of the angular part must be constrained by $\delta_{nl0}$ in equation (A.7). The same core states characterized by $nlm$ are expressed in the different coordinates. The term $r \vec{r}_i | \vec{b}_j \rangle$ in equation (A.1) means that the core states are projected into $\vec{r}_i$ coordinate system in which the $z$-direction is in $\vec{k}$. Similarly in $| \vec{b} \rangle$, they are projected into $r_\alpha$ (i.e. $\vec{r}_\alpha$) coordinates, in which the $z$-direction is in $\vec{k}$. So we can write $Y_m$ in equation (A.9) as $Y_{lk}$. Then we get the angular part to be [28]

$$
\langle Y_{lk} \mid Y_{lk} \rangle = \Omega^{-1/2} \int \frac{d \vec{P}_l}{(d \cos \alpha)} \frac{\vec{k} \times k}{k^2} \frac{d \vec{r}}{r}
$$

(A.10)

where $P_l$ are the associated Legendre polynomials, $\alpha$ is the mutual angle between the vector $\vec{k}$ and $\vec{k'}$.

From equations (A.10), (A.9) and (A.8), we get

$$
\Omega^{-1/2}[C] = i A_{nl} \Omega^{-1/2} \int d \vec{r} e^{i \vec{r} \cdot \vec{r}'} [4 \pi (2l + 1)]^{1/2} R_{nl}(r) \xi(r) j_l(k'r) r^2 \frac{d \vec{r}}{r}
$$

(A.11)

where $A_{nl} = \Omega^{-1/2} \int d \vec{r} e^{i \vec{r} \cdot \vec{r}'} [4 \pi (2l + 1)]^{1/2} R_{nl}(r) \xi(r) j_l(k'r) r^2 \frac{d \vec{r}}{r}$, which is similar to the ‘A’ function in ([28]). Then we can get the term (II) as

$$
(II) = (-i) S(\Delta \vec{k}) \frac{\vec{k} \times \vec{k'}}{k k'}
$$

\(\times \sum_{nl} (-1)^l B_{nl1}(\vec{k'}) A_{nl}(\vec{k}) \frac{d P_l}{d \cos \alpha} \),

(A.12)

where the summations shall be over the quantum numbers $n$ and $l$.

The other terms in (A3) in [28] can also derive from the similar calculations. Then the final expression of SOC can be written as

$$
H_{\vec{k}, \vec{k'}}^{SO} = -\hat{\sigma}_{s \bar{s}} \cdot (\vec{k} \times \vec{k'}) [i S(\Delta \vec{k}) \lambda_{atom}],
$$

(A.13)

where $\lambda_{atom} = (1/k k') \sum_{nl} (-1)^l \left[ B_{nl1}(\vec{k'}) A_{nl}(\vec{k}) + A_{nl}(\vec{k'}) B_{nl1}(\vec{k}) \right]

- \sum_{nl'} B_{nl1}(\vec{k'}) B_{nl1}(\vec{k}) \eta_{nl'} \frac{d P_l}{d \cos \alpha},

(A.14)

where the first term comes from the plane wave (p)–core state (c), the second term comes from c–p and the third term comes from the c–c term.

\[ \eta_{nl'} = \int_0^{\infty} R_{nl}(r) \xi(r) R_{nl}(r) r^2 \frac{d r}{r} \]

(A.15)

is about the spin–orbit splitting of isolated atoms. It is just determined by local wavefunctions of $e$ states. We shall discuss two cases.

**Case one: two different atoms in a unit cell.** We set $\gamma = \Delta \vec{k} - \vec{\tau}$, $\lambda_1$ and $\lambda_2$ are corresponding to the two atoms. We get [29]

$$
H_{\vec{k}, \vec{k}'}^{SO} = -i \hat{\sigma}_{s \bar{s}} \cdot (\vec{k} \times \vec{k'}) \chi(k' \gamma \vec{k})
$$

(A.16)

where $\chi(k' \gamma \vec{k}) = (\lambda_1 + \lambda_2)/2$ are the symmetric (antisymmetric) contributions to the spin–orbit Hamiltonian.

**Case two: multi-type and many atoms in a unit cell.** If we have $m$ species atoms in a unit cell and $n_1, n_2, \ldots, n_m$ are the numbers of atoms for each type atom in a unit cell. We have $\sum_{i=1}^{n_m} n_i = n_{num}$. We get

$$
H_{\vec{k}, \vec{k}'}^{SO} = -i \hat{\sigma}_{s \bar{s}} \cdot (\vec{k} \times \vec{k'}) \sum_{j=1}^{m} S_j(\Delta \vec{k}) \lambda_j,
$$

(A.17)

where $\lambda_j$ is for $j$th species, and structure factor is $S_j(\Delta \vec{k}) = n_{num} \sum_{n} \exp(i \vec{d} \cdot \vec{\tau}_n)$, here summation over $n_j$ means over the atoms of $j$th species.

If we only keep the third term c–c in equation (A.14) (as in [29–31]) and only consider the main contribution from the outermost $l = 1$ core state, i.e. $p$ orbit. Also note the relation $d P_l/d(\cos \alpha) = 1$, for $l = 1$, we have

$$
\lambda_{atom} = (kk')^{-1} B_{nl1}(\vec{k'}) B_{nl1}(\vec{k}) \eta_{nl1},
$$

(A.18)

where $n$ is selected as the quantum number characterizing the outermost $p$ core state and $\eta_{nl1}$ is the spin–orbit parameter for various atoms, we write it as $\eta_{nl1}$. If we redefine the $B$ function by introducing a normalization constant $C$ as

$$
B_{nl}(\vec{k}) = C \int_0^{\infty} j_l(k'r) R_{nl}(r) r^2 \frac{d r}{r},
$$

(A.19)

where $C$ is determined by the limit $\lim_{k \to 0} k^{-1} B_{nl}(k) = 1$. And in this procedure, empirical adjustable parameters $\mu$ and $\alpha$ describing the ratio of the spin–orbit contributions for free atoms $A$ and $B$; see [29]. To evaluate the $B$ function
explicitly, we may determine the expressions of the core state wavefunction \( R_{nl}(r) \) and we use the formula in [34]

\[
\psi_{nlm} = \sum_p C_{nlp} \chi_{plm},
\]

(A.20)

where \( C_{nlp} \) are the coefficients of expansion and \( \chi_{plm} \) are Slater-type orbits with integer quantum numbers, namely

\[
\chi_{plm}(r, \theta, \phi) = \tilde{R}_{lp}(r)Y_{m}^{l}(\theta, \phi),
\]

(A.21)

where

\[
\tilde{R}_{lp}(r) = [(2n_{lp})!]^{-1/2}(2\zeta_{lp})^{n_{lp}+1/2}r^{n_{lp}-1}e^{-\zeta_{lp}r},
\]

(A.22)

where \( C_{nlp} \) and \( \zeta_{lp} \) can be found in the tables in [34]. So the functions \( \psi \) in equation (A.20) are

\[
\psi_{nlm} = \left[ \sum_p \tilde{R}_{lp}(r)C_{nlp} \right] \times Y_{m}^{l}(\theta, \phi) = R_{lp}(r)Y_{m}^{l}(\theta, \phi).
\]

(A.23)

In terms of these functions, we can calculate the integral as

\[
I(k) = \int_{0}^{\infty} j_{l}(kr)R_{nl}(r)r^{2}dr = \sum_p C_{nlp} \int_{0}^{\infty} j_{l}(kr)\tilde{R}_{lp}(r)r^{2}dr
\]

\[
= \sum_p C_{nlp}N_{lp}I_{l}^{p}(k),
\]

(A.24)

where \( N_{lp} = [(2n_{lp})!]^{-1/2}(2\zeta_{lp})^{n_{lp}+1/2} \) and \( I_{l}^{p}(k) = \int_{0}^{\infty} j_{l}(kr)r^{n_{lp}+1}e^{-\zeta_{lp}r}dr \). The integral of \( I_{l}^{p}(k) \) can be easily derived when \( l = 0 \). It is

\[
I_{l=0}^{p}(k) = \frac{(n_{lp} - 1)!}{k^{2}\zeta_{lp}^{2} + k^{2}n_{lp}^{2}},
\]

(A.25)

where \( v = \arctan(k/\zeta_{lp}) \). If we calculate the value of the limit \( k \to 0 \), it shall be pointed out that the spherical Bessel functions \( j_{l}(kr) \) can be formulated \( j_{l}(x) = \sqrt{\pi}J_{l+1/2}(x) \), where \( J_{l+1/2}(x) \) is a Bessel function of the first kind. Then we can get

\[
\lim_{k \to 0} I_{l=0}^{p}(k) = \frac{k(n_{lp} + 2)!}{3}\zeta_{lp}^{n_{lp}+3}.
\]

(A.26)

This result is used to derive the normalization constant in equation (A.19).

References

[1] ITRS 2004 http://www.itrs.net/
[2] Bin Yu et al 2002 IEDM Tech. Digest 10...02

[3] Komoda T et al 2004 IEDM Tech. Digest 217
[4] Low T, Feng Y P, Li M F, Samudra G, Yeo Y C, Bai P, Chan L and Kwong D L 2005 ISDRS (Semiconductor Device Research Symp., 2005 International) pp 384–5
[5] Fischetti M V and Lau X 1989 IEDM Tech. Digest 481–4
[6] Ashley T et al 2004 ICSICT (Solid-State and Integrated Circuits Technology, 2004, Proc. 7th Int. Conf. vol 3, p 2253
[7] Li M F 1994 Modern Semiconductor Quantum Physics (Singapore: World Scientific)
[8] Levinshtein M, Rumyantsev S and Shur M 1996 Handbook Series on Semiconductor Parameters (Singapore: World Scientific)
[9] Petha A, Krishnamohan T, Kim D, Sae Rooney O h, Wong H, Philip C S, Nishi Y and Saraswat K C 2005 IEDM p 605
[10] Harrison W A 1966 Pseudopotentials in the Theory of Metals (Reading, MA: Benjamin)
[11] Heine V 1970 Solid State Physics vol 24 ed H Ehrenreich, F Seitz and D Turnbull (New York: Academic)
[12] Zhu G Z, Low T, Li M F, Fan W J, Bai P, Kwong D L and Samudra G 2006 IEDM Tech. Digest S31P2
[13] Fong C Y and Cohen M L 1970 Phys. Rev. Lett. 24 306
[14] Cohen M L and Bergstresser T K 1966 Phys. Rev. 141 789
[15] Cohen M L, Lin P J, Roessler D M and Walker W C 1967 Phys. Rev. B 50 155992
[16] Wang L W and Zunger A 1994 J. Phys. Chem. B 98 2158
[17] Wang L W and Zunger A 1994 J. Phys. Chem. 100 2394
[18] Jian-Bai X and Cheah K W 1997 Phys. Rev. B 55 15688
[19] Chin-Yu Y, Zhang S B and Zunger A 1994 Phys. Rev. B 50 14405
[20] Jian-Bai Xia and Cheah K W 1997 Phys. Rev. B 56 14925
[21] Zhang S B, Chin-Yu Yeh and Zunger A 1993 Phys. Rev. B 48 112049
[22] Van de Walle C G and Martin R M 1987 Phys. Rev. B 35 8154
[23] Van de Walle C G and Martin R M 1989 Phys. Rev. B 39 1871
[24] Magri R and Zunger A 2002 Phys. Rev. B 65 165302
[25] Segall M D, Lindan P J D, Probert M J, Pickard C J, Hasnip P J, Clark S J and Payne M C 2002 J. Phys.: Condens. Matter 14 2717
[26] Raghavachari K, Fu Q, Chen G, Li L, Li C H, Law D C and Hicks R F 2002 J. Am. Chem. Soc. 124 15119
[27] Balázs G, Breunig H J, Lork E and Offermann W 2001 Organometalics 20 2666
[28] Nordrup J E 1991 Phys. Rev. B 44 1419
[29] Weisz G 1966 Phys. Rev. 149 504
[30] Chelikowsky J R and Cohen M L 1976 Phys. Rev. B 14 556
[31] Chelikowsky J R and Cohen M L 1984 Phys. Rev. B 30 4828
[32] Walter J P, Cohen M L, Petroff Y and Balkanski M 1970 Phys. Rev. B 1 2664
[33] Bloom S and Bergstresser T K 1970 Phys. Stat. Sol. 3 141
[34] Clementi E and Roetti C 1974 At. Data Nucl. Data Tables 14 177–478