Tight-binding Hamiltonian from first-principles calculations

**Cai-Zhuang Wang · Wen-Cai Lu · Yong-Xin Yao ·
Ju Li · Sidney Yip · Kai-Ming Ho**

1 Introduction

The tight-binding method attempts to represent the electronic structure of condensed matter using a minimal atomic-orbital like basis set. To compute tight-binding overlap and Hamiltonian matrices directly from first-principles calculations is a subject of continuous interest. Usually, first-principles calculations are done using a large basis set or long-ranged basis set (e.g. muffin-tin orbitals (MTOs)) in order to get convergent results, while tight-binding overlap and Hamiltonian matrices are based on a short-ranged minimal basis representation. In this regard, a transformation that can carry the electronic Hamiltonian matrix from a large or long-ranged basis representation onto a short-ranged minimal basis representation is necessary to obtain an accurate tight-binding Hamiltonian from first principles.

The idea of calculating tight-binding matrix elements directly from a first-principles method was proposed by Andersen and Jepsen in 1984 [1]. They developed a scheme which transforms the electronic band structures of a crystal calculated using a long-ranged

C.-Z. Wang (✉) · Y.-X. Yao · K.-M. Ho
US Department of Energy, Ames Laboratory, Ames, IA 50011, USA
e-mail: wangcz@ameslab.gov

C.-Z. Wang · Y.-X. Yao · K.-M. Ho
Department of Physics and Astronomy, Iowa State University, Ames, IA 50011, USA

W.-C. Lu
State Key Laboratory of Theoretical and Computational Chemistry, Jilin University, Changchun 130021,
People's Republic of China

J. Li
Department of Materials Science and Engineering, University of Pennsylvania, Philadelphia,
PA 19104, USA

S. Yip
Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge,
MA 02139, USA

S. Yip
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge,
MA 02139, USA

basis set of muffin-tin orbitals (MTO's) into a much shorter-ranged tight-binding representation. In the framework of this transformation, tight-binding matrix elements can be calculated by first-principles LMTO method and casted into an effective two-center tight-binding Hamiltonian called LMTO-TB Hamiltonian [1]. More recently, an improved version of such a "downfolding" LMTO method, namely the order-N MTO [2], has also been developed which allows the LMTO-TB Hamiltonian matrix elements to be extracted more accurately from a full LMTO calculation [3].

Another approach to determine the tight-binding Hamiltonian matrix elements by first-principles calculations was developed by Sankey and Niklewski [4] and by Porezag et al. [5]. In their approach, the matrix elements are calculated directly by applying an effective one-electron Hamiltonian of the Kohn-Sham type onto a set of pre-constructed atomic-like orbitals. The accuracy of the tight-binding Hamiltonian constructed in this way depends on the choice of atomic-like basis orbitals. More recently, McMahan and Klepeis [6] have developed a method to calculate the two-center Slater-Koster hopping parameters and effective on-site energies from minimal basis functions optimized for each crystal structure, in terms of $\mathbf{k}$-dependent matrix elements of one-electron Hamiltonian obtained from first-principles calculations.

All of the above mentioned work was derived from a description of electronic structures using a fixed minimal basis set, except the work of McMahan and Klepeis [6]. It should be noted that while a fixed minimal basis set can give a qualitative description of electronic structures, it is too sparse to give an accurate description of the energetics of systems in varying bonding environments. A much larger basis set would be required in the first-principles calculations in order to get accurate and convergent results if the basis set is going to be kept fixed for various structures. Thus, it is clear that in order for a minimal basis set to have good transferability, it is important to focus our attention on the changes that the basis must adopt in different bonding environments.

In the past several years, we have developed a method for projecting a set of chemically deformed atomic minimal basis set orbitals from accurate *ab initio* wave functions [7–12]. We call such orbitals "quasi-atomic minimal-basis orbitals" (QUAMBOs) because they are dependent on the bonding environments but deviate very little from free-atom minimal-basis orbitals. While highly localized on atoms and exhibiting shapes close to orbitals of the isolated atoms, the QUAMBOs span exactly the same occupied subspace as the wavefunctions determined by the first-principles calculations with a large basis set. The tight-binding overlap and Hamiltonian matrices in the QUAMBO representation give exactly the same energy levels and wavefunctions of the occupied electronic states as those obtained by the fully converged first-principles calculations using a large basis set. Therefore, the tight-binding Hamiltonian matrix elements derived directly from *ab initio* calculations through the construction of QUAMBOs are highly accurate.

In this article, we will review the concept and the formalism used in generating the QUAMBOs from first-principles wavefunctions. Then we show that tight-binding Hamiltonian and overlap matrix elements can be calculated accurately by the first-principles methods through the QUAMBO representation. By further decomposing the matrix elements into the hopping and overlap parameters through the Slater-Koster scheme [13], the transferability of the commonly used two-center approximation in the tight-binding parameterization can be examined in detail. Such an analysis will provide very useful insights and guidance for the development of accurate and transferable tight-binding models. Finally, we will also discuss a scheme for large scale electronic structure calculation of complex systems using the QUAMBO-based first-principles tight-binding method.

2 Quasi-atomic minimal-basis-sets orbitals

The method to project the QUAMBOs from the first-principles wave functions has been described in detail in our previous publications [7–12]. Some of the essential features of the method will be reviewed here using Si as an example. If the Si crystal structure contains N silicon atoms and hence $4N$ valence electrons in a unit cell, the total number of minimal sp^3 basis orbitals per unit cell will be $4N$. In our method, the $4N$ QUAMBOs (A_α) are spanned by $2N$ occupied valence orbitals which are chosen to be the same as those from the first-principles calculations, and by another $2N$ unoccupied orbitals which are linear combinations of a much larger number of unoccupied orbitals from first-principles calculations. The condition for picking such $2N$ unoccupied orbitals is the requirement that the resulting QUAMBOs deviate as little as possible from the corresponding $3s$ and $3p$ orbitals of a free Si atom (A_α^0). The key step in constructing the above mentioned QUAMBOs is the selection of a small subset of unoccupied orbitals, from the entire virtual space, that are maximally overlapped with the atomic orbitals of the free atom A_α^0.

Suppose that a set of occupied Bloch orbitals $\phi_\mu(\mathbf{k},\mathbf{r})$ ($\mu = 1, 2, \ldots, n_{occ}(\mathbf{k})$) and *virtual* orbitals $\phi_v(\mathbf{k},\mathbf{r})$ ($v = n_{occ}(\mathbf{k}) + 1, n_{occ}(\mathbf{k}) + 2, \ldots, n_{occ}(\mathbf{k}) + n_{vir}(\mathbf{k})$), labeled by band μ or v, and wave vector $\mathbf{k}$, have been obtained from first-principles calculations using a large basis set, our objective is to construct a set of quasi-atomic orbitals $A_\alpha(\mathbf{r} - \mathbf{R}_i)$ spanned by the occupied Bloch orbitals $\phi_\mu(\mathbf{k},\mathbf{r})$ and an optimal subset of orthogonal virtual Bloch orbitals $\varphi_p(\mathbf{k},\mathbf{r})$

$$A_\alpha(\mathbf{r} - \mathbf{R}_i) = \sum_{\mathbf{k},u} a_{\mu\alpha}(\mathbf{k},\mathbf{R}_i)\phi_\mu(\mathbf{k},\mathbf{r}) + \sum_{\mathbf{k},p} b_{p\alpha}(\mathbf{k},\mathbf{R}_i)\varphi_p(\mathbf{k},\mathbf{r}) \tag{1}$$

where

$$\varphi_p(\mathbf{k},\mathbf{r}) = \sum_v T_{vp}(\mathbf{k})\phi_v(\mathbf{k},\mathbf{r}), \ (p = 1, 2, \ldots, n_p(\mathbf{k}) < n_{vir}(\mathbf{k})) \tag{2}$$

The orthogonal character of $\varphi_p(\mathbf{k},\mathbf{r})$ gives $\sum_v T_{vp}^*(\mathbf{k})T_{vq}(\mathbf{k}) = \delta_{pq}$, in which T is a rectangular matrix which will be determined later.

The requirement is that A_α should be as close as possible to the corresponding free atom orbitals A_α^0. Mathematically, this is a problem of minimizing $\langle A_\alpha - A_\alpha^0 | A_\alpha - A_\alpha^0 \rangle$ under the side condition $\langle A_\alpha | A_\alpha \rangle = 1$. Therefore the Lagrangian for this minimization problem is

$$L = \langle A_\alpha - A_\alpha^0 | A_\alpha - A_\alpha^0 \rangle - \lambda \left(\langle A_\alpha | A_\alpha \rangle - 1 \right) \tag{3}$$

The Lagrangian minimization leads to

$$A_\alpha(\mathbf{r} - \mathbf{R}_i) = D_{i\alpha}^{-1/2} \left[\sum_{\mathbf{k},\mu} \langle \phi_\mu(\mathbf{k},\mathbf{r}) \mid A_\alpha^0(\mathbf{r} - \mathbf{R}_i) \rangle \phi_\mu(\mathbf{k},\mathbf{r}) \right.$$

$$\left. + \sum_{\mathbf{k},p} \langle \varphi_p(\mathbf{k},\mathbf{r}) \mid A_\alpha^0(\mathbf{r} - \mathbf{R}_i) \rangle \varphi_p(\mathbf{k},\mathbf{r}) \right] \tag{4}$$

where

$$D_{i\alpha} = \sum_{\mathbf{k},\mu} \left| \langle \phi_\mu(\mathbf{k},\mathbf{r}) \mid A_\alpha^0(\mathbf{r} - \mathbf{R}_i) \rangle \right|^2 + \sum_{\mathbf{k},p} \left| \langle \varphi_p(\mathbf{k},\mathbf{r}) \mid A_\alpha^0(\mathbf{r} - \mathbf{R}_i) \rangle \right|^2 \tag{5}$$

For this optimized A_α, the mean-square deviation from A_α^0 is

$$\Delta_{i\alpha} = \langle A_\alpha - A_\alpha^0 | A_\alpha - A_\alpha^0 \rangle^{1/2} = [2(1 - D_{i\alpha}^{1/2})]^{1/2} \tag{6}$$

It is clear from Eqs. (5) and (6) that the key step to get quasi-atomic minimal-basis-set orbitals is to select a subset of virtual orbitals $\varphi_p(\mathbf{k},\mathbf{r})$ which can maximize the matrix trace

$$S = \sum_{i,\alpha,\mathbf{k},p} \langle \varphi_p(\mathbf{k},\mathbf{r}) | A_\alpha^0(\mathbf{r} - \mathbf{R}_i) \rangle \langle A_\alpha^0(\mathbf{r} - \mathbf{R}_i) | \varphi_p(\mathbf{k},\mathbf{r}) \rangle \tag{7}$$

The maximization can be achieved by first diagonalizing the matrix

$$B_{\nu\nu'}^{\mathbf{k}} = \sum_{i,\alpha} \langle \phi_\nu(\mathbf{k},\mathbf{r}) \mid A_\alpha^0(\mathbf{r} - \mathbf{R}_i) \rangle \langle A_\alpha^0(\mathbf{r} - \mathbf{R}_i) \mid \phi_{\nu'}(\mathbf{k},\mathbf{r}) \rangle \tag{8}$$

for each $\mathbf{k}$-point, where ν and ν' run over all unoccupied states up to a converged upper cutoff. The transformation matrix T which defines the optimal subset of virtual Bloch orbitals $\varphi_p(\mathbf{k},\mathbf{r})$ ($p = 1, 2, \ldots, n_p(\mathbf{k})$) by Eq. (2) is then constructed using the $\sum_k n_p(\mathbf{k})$ eigenvectors *with the largest eigenvalues* of the matrixes $B^{\mathbf{k}}$, each of such eigenvectors will be a column of the transformation matrix T. Given $\varphi_p(\mathbf{k},\mathbf{r})$, the localized QUAMBOs are then constructed by Eqs. (4) and (5). As one can see from the above formalism development that the key concept in this QUAMBO construction is to keep the bonding states (occupied state) intact and at the same time searching for the minimal number of anti-bonding states (which are usually not the lowest unoccupied states) from the entire unoccupied subspace. The bonding states that kept unchanged and the anti-bonding states constructed from the unoccupied states can form the desirable localized QUAMBOs.

Figure 1 shows the s- and p- like QUAMBOs of Si in diamond structure with different bond lengths of 1.95 Å, 2.35 Å and 2.75 Å, and in fcc structure with bond lengths of 2.34 Å, 2.74 Å, and 3.14 Å, respectively. The QUAMBOs are in general non-orthogonal by our construction as discussed above. One can see that the QUAMBOs constructed by our scheme are indeed atomic-like and well localized on the atoms. These QUAMBOs are different from the atomic orbitals of the free atoms because they are deformed according to the bonding environment. It is clear that the deformations of QUAMBOs are larger with shorter interaction distances. When the bond length increases to be 2.75 Å, the QUAMBOs are very close to the orbitals of a free atom.

As we discussed above, the effective one-electron Hamiltonian matrix in the QUAMBO representation by our construction preserves the occupied valence subspace from the first-principles calculations so that it should give the exact energy levels and wavefunctions for the occupied states as those from first-principles calculations. This property can be seen from Fig. 2 where the electronic density-of-states (DOS) of Si in the diamond structure calculated using QUAMBOs are compared with that from the original first-principles calculations. It is clearly shown that the electronic states below the energy gap are exactly reproduced by the
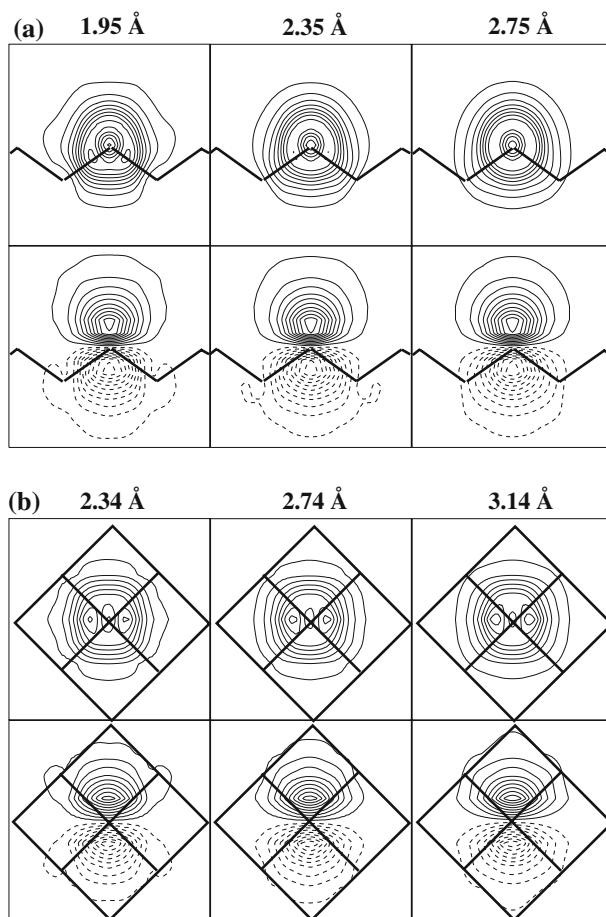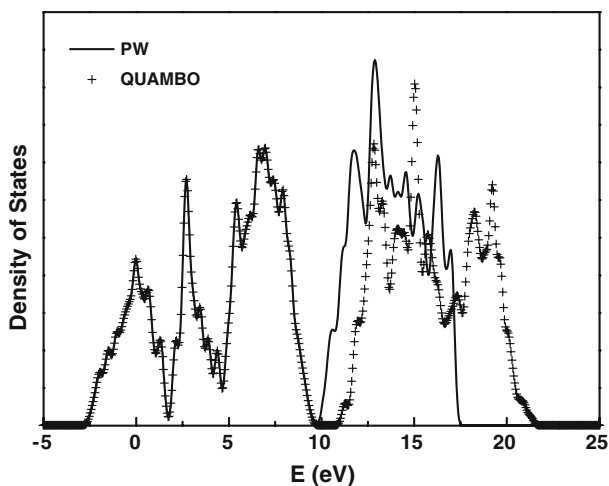
Fig. 1 Non-orthogonal *s*- and *p*- like QUAMBOs in Si (a) diamond structure in the (110) plane for three bond lengths 1.95 Å, 2.35 Å and 2.75 Å, and (b) fcc structure in the (100) plane for three bond lengths 2.34 Å, 2.74 Å, and 3.14 Å

QUAMBOs, while the unoccupied states have been shifted upwards so that the energy gap between the valence and conduction states increases from ∼0.7 eV to ∼1.8 eV. This shift is expected because the QUAMBOs contain admixtures of eigenstates from the higher energy spectrum.

It should be noted that the formalism for the QUAMBOs construction discussed in the section is based on the wavefunctions from first-principles calculations using all-electrons or norm-conserving pseudopotentials [14]. The formalism for constructing the QUAMBOs from first-principles calculations using ultra-soft pseudopotential (USPP) [15] or projector augmented-wave (PAW) [16], as implemented in the widely used VASP code [17,18], is similar and has been recently worked out by Qian et al. [12]. Moreover, Qian et al. also adopt a projected atomic orbital scheme [19–21] which replaces the unoccupied subspace from the first-principles calculations in the above formula with a projection of the unoccupied part of the atomic orbitals, and improve the efficiency and stability of the QUAMBO construction procedure [12].

Fig. 2 Electronic density of states of diamond Si obtained by using the QUAMBOs as basis set, compared with those from the corresponding LDA calculations using the PW basis set

3 Tight-binding matrix elements in terms of QUAMBOs

Once the QUAMBOs have been constructed, overlap and effective one-electron Hamiltonian matrices in representation of QUAMBOs are readily calculated from first-principles.

$$S_{i\alpha,j\beta} = \langle A_\alpha(\mathbf{r} - \mathbf{R}_i) \,|\, A_\beta(\mathbf{r} - \mathbf{R}_j) \rangle \tag{9}$$

$$H_{i\alpha,j\beta} = \langle A_\alpha(\mathbf{r} - \mathbf{R}_i) |H| A_\beta(\mathbf{r} - \mathbf{R}_j) \rangle \tag{10}$$

H in Eq. 10 can then be expressed by using the corresponding eigenvalues ε_n and eigenfunctions ϕ_n from original DFT calculations, i.e., $H = \sum_n \varepsilon_n |\phi_n\rangle \langle\phi_n|$, and thus the matrix elements $H_{i,\alpha,j\beta}$ can be calculated easily.

Note that in our approach the electronic eigenvalues and eigenfunctions of the occupied states from first-principles calculations are exactly reproduced by the QUAMBO representation. Although the overlap and effective one-electron Hamiltonian matrices in terms of the QUAMBOs are in a minimal basis representation, the matrices obtained from our method go beyond the traditional two-center approximation. Therefore, the Slater-Koster tight-binding parameters [13] obtained by inverting such first-principles matrices are expected to be environment-dependent.

In order to examine how the overlap and hopping integrals are dependent on the environment and to see how serious the error the two-center approximation will make in traditional tight-binding approaches, we have performed calculations for 3 types (i.e, diamond, simple cubic (sc), and face-centered cubic (fcc)) of crystal structures of Si with several different bond lengths for each type of structures in order to study the tight-binding parameters in different bonding environments. Based on the overlap and effective one-electron Hamiltonian matrix elements from our QUAMBO scheme, the Slater-Koster overlap integrals $s_{ss\sigma}$, $s_{sp\sigma}$, $s_{pp\sigma}$, and $s_{pp\pi}$, and hopping integrals $h_{ss\sigma}$, $h_{sp\sigma}$, $h_{pp\sigma}$, and $h_{pp\pi}$ are then extracted using the Slater-Koster geometrical factors [13]. The results for the overlap and hopping integrals as a function of interatomic distance in the three different crystal structures are plotted in Figs. 3 and 4, respectively.

Figure 3 shows the overlap parameters $s_{ss\sigma}$, $s_{sp\sigma}$, $s_{pp\sigma}$, and $s_{pp\pi}$ from different structures and different pairs of atoms, plotted as a function of interatomic distance. Note that the two-center nature of overlap integrals for fixed atomic minimal basis orbitals may not necessarily
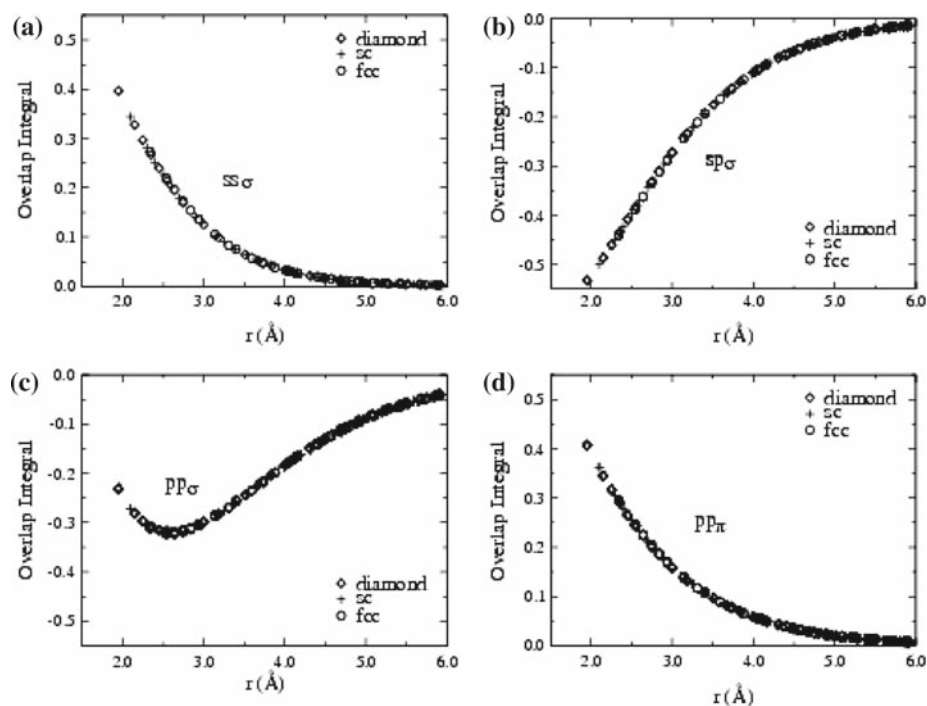
Fig. 3 Overlap integrals as a function of interatomic distance for Si in the diamond, sc, and fcc structures

hold for the QUAMBOs because QUAMBOs are deformed according to the bonding environments of the atoms. Nevertheless, the overlap parameters obtained from our calculations as plotted in Fig. 3 fall into smooth scaling curves nicely. These results suggest that the two-center approximation is adequate for overlap integrals.

By contrast, the hopping parameters as plotted in Fig. 4 are far from being transferable, especially for $h_{pp\sigma}$. Even for the best case of $h_{ss\sigma}$, the spread in the first neighbor interaction is about 1 eV. For a given pair of atoms, the hopping parameters $h_{pp\sigma}$ and $h_{pp\pi}$ obtained from the decompositions of different matrix elements can exhibit slightly different values, especially for the sc and fcc structures. The hopping parameters from different structures do not follow the same scaling curve. For a given crystal structure, although the bond-length dependence of hopping parameters for the first and second neighbor interactions can be fitted to separate smooth scaling curves respectively, these two scaling curves cannot be joined together to define an unique transferable scaling function for the structure. These results suggest that under the two-center approximation, it is not possible to describe the scaling of the tight-binding hopping parameters accurately.

It is interesting to note from Fig. 4 that the structure which has larger coordination number tends to have larger hopping parameter (in magnitude) as compared to the lower-coordinated structure at the same interatomic separation. It is also interesting to note that the scaling curve of the second neighbor interactions tends to be above that of the first neighbors at the same interatomic distance. These behaviors are indications of significant contributions from three-center integrals, because more contribution from the three-center integrals is expected for pair of atoms that have more neighbors which enhance the effective hopping between the two atoms.
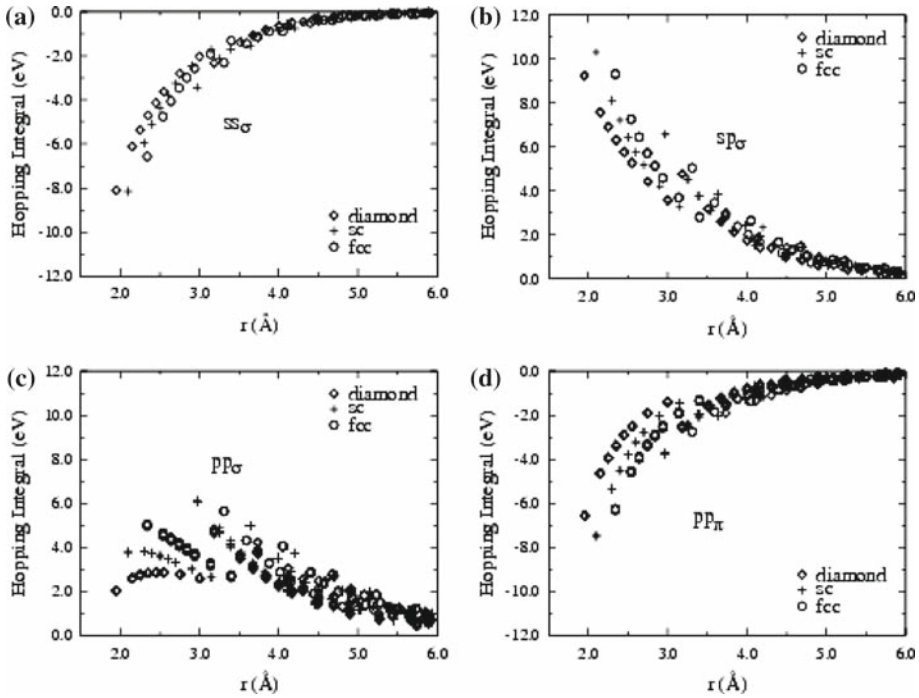
Fig. 4 Non-orthogonal tight-binding hopping integrals for Si as a function of interatomic distance in the diamond, sc, and fcc structures obtained by decomposing the QUAMBO-based effective one-electron Hamiltonian according to the Slater-Koster tight-binding scheme

To express the tight-binding Hamiltonian matrix in terms of QUAMBOs also allows us to address the issue of the effects of orthogonality on the transferability of tight-binding models from the first-principles perspective. We can construct orthogonal QUAMBOs from non-orthogonal ones using the symmetrical orthogonalization method of Löwdin [22]. Starting from the Bloch sum of non-orthogonal QUAMBOs

$$\tilde{A}_\alpha(\mathbf{k}, \mathbf{r}) = \frac{1}{\sqrt{N}} \sum_n \exp(i\mathbf{k} \cdot \mathbf{r}_n) A_\alpha(\mathbf{r} - \mathbf{r}_n) \tag{11}$$

the overlap matrix of $\tilde{A}_\alpha(\mathbf{k}, \mathbf{r})$ can be defined as

$$S_{\alpha\beta}(\mathbf{k}) = \sum_n \exp(i\mathbf{k} \cdot \mathbf{r}_n) \langle A_\alpha(\mathbf{r}) \mid A_\beta(\mathbf{r} - \mathbf{r}_n) \rangle \tag{12}$$

For each k-point, we perform the symmetrical orthogonalization method of Löwdin [22],

$$\tilde{A}_\alpha^{orthog}(\mathbf{k}, \mathbf{r}) = \sum_\beta S_{\beta\alpha}(\mathbf{k})^{-1/2} \tilde{A}_\beta(\mathbf{k}, \mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{n,\beta} \exp(i\mathbf{k} \cdot \mathbf{r}_n) S_{\beta\alpha}(\mathbf{k})^{-1/2} A_\beta(\mathbf{r} - \mathbf{r}_n)$$

$$= \frac{1}{\sqrt{N}} \sum_n \exp(i\mathbf{k} \cdot \mathbf{r}_n) \sum_{\mathbf{k'},\beta} S_{\beta\alpha}(\mathbf{k'})^{-1/2} A_\beta^{\mathbf{k'}}(\mathbf{r} - \mathbf{r}_n) \tag{13}$$

Then the orthogonal QUAMBOs can be expressed by

$$\tilde{A}_\alpha^{orthog}(\mathbf{r} - \mathbf{r}_n) = \sum_{\mathbf{k},\beta} S_{\beta\alpha}(\mathbf{k})^{-1/2} A_\beta^{\mathbf{k}}(\mathbf{r} - \mathbf{r}_n) \tag{14}$$

since

$$\begin{aligned}
&\left\langle A_\alpha^{orthog}(\mathbf{r} - \mathbf{r}_n) \mid A_\beta^{orthog}(\mathbf{r} - \mathbf{r}_{n'}) \right\rangle \\
&= \frac{1}{N} \sum_{\mathbf{k},\mathbf{k}'} \left\langle \tilde{A}_\alpha^{orthog}(\mathbf{k},\mathbf{r}) \mid \tilde{A}_\beta^{orthog}(\mathbf{k}', \mathbf{r}) \right\rangle \exp(i\mathbf{k} \cdot \mathbf{r}_n - i\mathbf{k}' \cdot \mathbf{r}_{n'}) \\
&= \frac{1}{N} \sum_{\mathbf{k}} \exp[i\mathbf{k} \cdot (\mathbf{r}_n - \mathbf{r}_{n'})]\delta_{\alpha\beta} \\
&= \delta_{nn'}\delta_{\alpha\beta}
\end{aligned} \tag{15}$$

Figure 5 shows the orthogonal *s*- and *p*- like QUAMBOs in Si diamond and fcc structures with three different bond lengths, respectively. In comparison with the non-orthogonal QUAMBOs as shown in Fig. 1, the orthogonal QUAMBOs are much more tightly localized on the center atoms, but some wavefuction components have been pushed out to the neighboring atoms in order to satisfy the orthogonal requirement. Using the orthogonal QUAMBOs, the effective one-electron Hamiltonian matrix in the orthogonal QUAMBO representation can be calculated and the orthogonal Slater-Koster hopping integrals can be extracted following the decomposition procedures discussed in the non-orthogonal tight-binding case. The results are plotted in Fig. 6. It is interesting to note that the orthogonal hopping parameters as a function of interatomic distance decay much faster than their non-orthogonal counterparts. Therefore, the interactions in the orthogonal tight-binding scheme are essentially dominated by the first neighbor interactions which depend not only on the interatomic separations but also on the coordination of the structures. In contrast to the non-orthogonal model, the magnitudes of the orthogonal hopping parameters decrease as the coordination number of the structure increases. These coordination-dependence of the hopping parameters and the short-range nature of the interactions are qualitatively similar to the environment-dependent tight-binding model of Wang et al. [23,24]. In their model, the coordination dependence of the hopping parameters is considered through a bond-length scaling function, and the short-ranged interactions is guaranteed by the screening function. However, though small, the contributions from the second and higher neighbor hopping parameters are not entirely negligible. In particular, some hopping parameters in the orthogonal TB scheme are found to change sign at the second and higher neighbors. The sign changes in the second and higher neighbor interactions can be attributed to the effects of the orthogonality which push some orbital wavefunctions to the nearby atomic sites in order to satisfy the orthogonal condition as one can see from Fig. 5. Such effects have not been noticed in previous tight-binding models.

4 Large-scale electronic calculations using the QUAMBO scheme

The above development in QUAMBO construction and "exact" tight-binding matrix elements calculation enables us to perform tight-binding electronic-structure calculations for large systems directly from the first-principles approach, without a fitting procedure to generate tight-binding parameters. A scheme based on this idea has been developed by Yao et al. [25]. In this scheme, an overlap or tight-binding Hamiltonian matrix of a big system is built
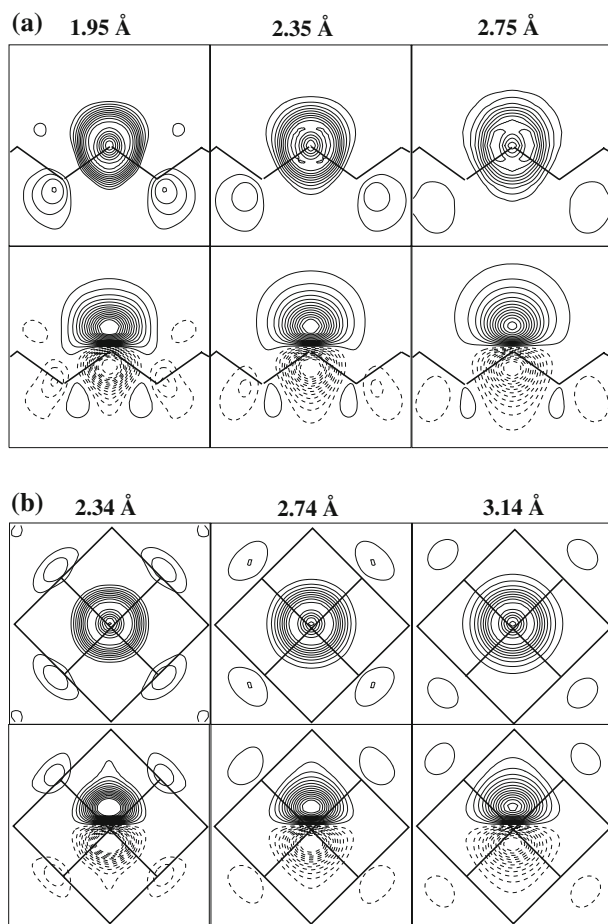
Fig. 5 Orthogonal *s*- and *p*- like QUAMBOs in Si (a) diamond structure in the (110) plane for three bond lengths 1.95 Å, 2.35 Å and 2.75 Å, and (b) fcc structure in the (100) plane for three bond lengths 2.34 Å, 2.74 Å, and 3.14 Å

by filling in the $n \times n$ "exact" sub-matrices (where n is the number of minimal basis orbitals for each atom) for every pair of atoms in the system. Note that the QUAMBOs and hence the $n \times n$ sub-matrices of tight-binding are dependent on the environment around the pair of atoms, the $n \times n$ "exact" sub-matrices has to be calculated for every pair of atoms in the system. This can be done by first performing first-principles calculations for a relatively small system with the same environment around the pair of atoms as if they are in the big systems, then the $n \times n$ tight-binding matrix for this pair of atoms can be constructed following the QUAMBO scheme. This approach will break the first-principles calculations of a big system into calculations for many much smaller sub-systems. In many cases of our interest (e.g., defects in crystals), the bonding environment of many different atom pairs in the big system may be essentially the same, therefore, first-principles calculations are needed only for a limited number of smaller systems and an accurate tight-binding overlap and Hamiltonian matrices for the big system can be constructed.
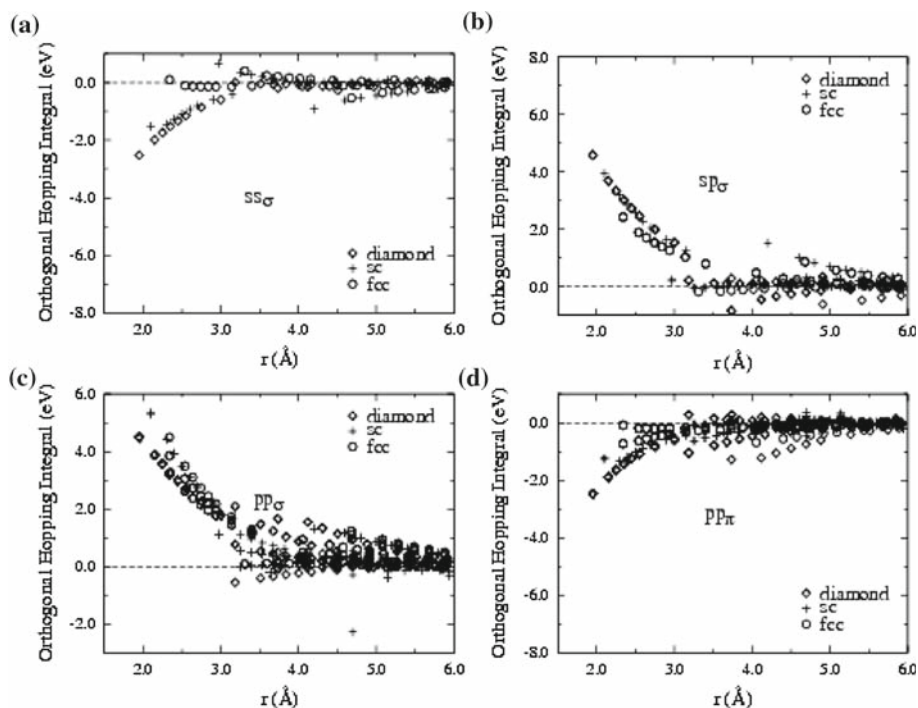
Fig. 6 Orthogonal tight-binding hopping integrals for Si as a function of interatomic distances in the diamond, sc, and fcc structures obtained by decomposing the QUAMBO-based effective one-electron Hamiltonian according to the Slater-Koster tight-binding scheme

The method was recently applied to studies the electronic structure of graphene nanoribbons [25]. For calculating the electronic structure of perfect armchair-grapheme nano-ribbons (A-GNRs) of different width, three different types of atoms in the nano-ribbons have been identified as illustrated in Fig. 7 where atom **a** represents a carbon atom inside the ribbon, atom **b** represents a carbon atom at the edges, and atom **c** is a hydrogen atom for passivation. Only one training sample of a $N_a = 7$ A-GNRs as shown in Fig. 7 and a single first-principles calculation are needed to extract all the necessary "exact" 4×4 or 4×1 tight-binding matrices for each pair of **a-a**, **a-b**, **b-b**, and **b-c** atoms from these three type of non-equivalent atoms, respectively. Fig. 8 shows the band structures and electronic density of states (DOS) for A-GNR with the width $N_a = 7$ and 13 (solid lines) from the QUAMBO-tight-binding scheme using the small 4×4 and 4×1 tight-binding matrices generated from the $N_a = 7$ training cell as described above. The results from full first-principles calculations (circle) were also shown for comparison. One can see that the QUAMBO-TB band structures agree very well with the full first-principles results up to 1eV above the Fermi-level. The electronic band gap variation of a perfect A-GNR as a function of the width of the nanoribbon has also been studied. Fig. 9 shows the oscillating behavior of band gap with a period of $N_a = 3$ obtained from our QUAMBO-TB scheme agree very well the results from first-principles calculations [26–28]. The efficiency of the QUAMBO-TB scheme enable us to calculate the electronic structure of much wider grapheme nano-ribbon, as one can also see from Fig. 9 where the band gap of a nanoribbon up to 100 Å in width has been calculated by our QUAMBO-TB method.

Fig. 7 (Color online) A-GNR with $N_a = 7$ was chosen to be a training cell. Dotted rectangle indicates the unit cell. The left arrow gives the periodical direction. Atom **a** and **b** are treated to be different due to different local environment
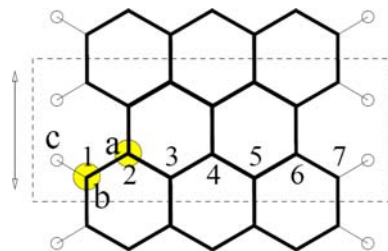
Fig. 8 TB band structures based on the QUAMBO-generated TB parameters (solid line) compared with first-principles DFT results (circle) for A-GNR with $N_a = 7$ and 13 respectively
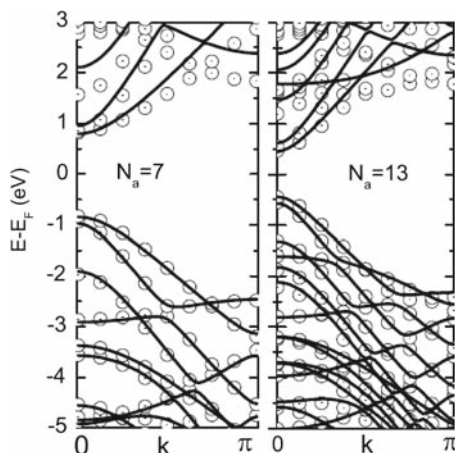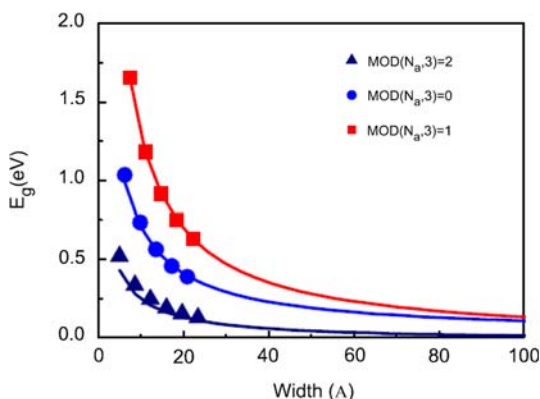
Fig. 9 (Color online) TB band gap (solid lines) of A-GNR with different size compared with first-principles DFT results (symbols)

The efficiency of the QUAMBO-TB scheme also enable us to study the electronic structure of grapheme nano-ribbons with defects, which usually require a much large unit cell and it is not easy to calculate using straightforward first-principles calculations. Yao et al. have studied the electronic structures of a $N_a = 6$ A-GNR with random edge defects on one edge of the ribbon at different defect ratio [25]. The supercell used in the calculation contains 1200 carbon atoms and about 200 hydrogen atoms. The edge defects were generated by randomly removing pairs of carbon atoms at one side of A-GNR as shown on Fig. 10(a). The carbon atoms at the defected edge were again passivated by hydrogen atoms. For this defect system, only some additional QUAMBO-TB matrix elements around the edge defects are needed to
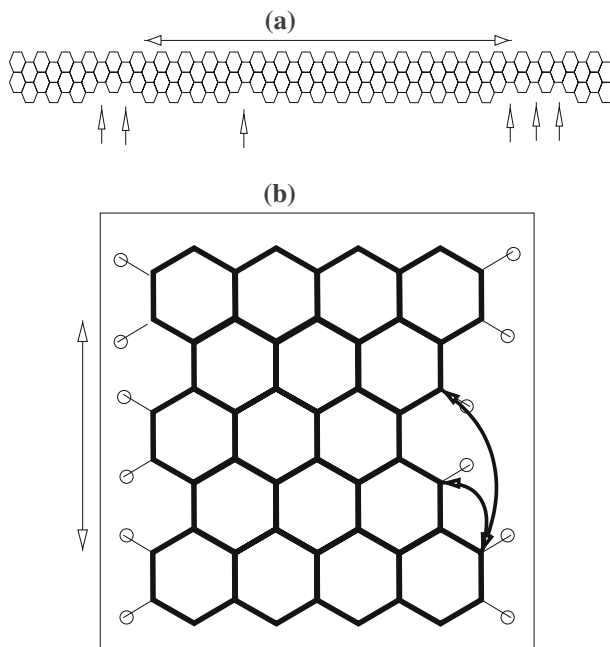
Fig. 10 (Color online) (a) Schematic view of a part of a supercell of $N_a = 6$ A-GNR containing more than one thousand atoms with edge defects randomly distributed on one side.. (b) The training cell used to generate the additional TB parameters for the A-GNR with edge defects

be generated using a training cell as shown in Fig. 10(b), where the curved arrows indicate the new matrix elements between these sites to be added to the existing QUAMBO-TB matrix elements database from the $N_a = 7$ training cell as discussed above. Based on this set of QUAMBO-TB matrix elements from first-principles calculations performed on two small unit cells, actuate tight-binding overlap and Hamiltonian matrices for the defected graphene nano-ribbons of various defect concentration can be constructed, and the electronic structure of A-GNRs with random edge defects can be studied. The results of band gap as the function of defect ratio in the $N_a = 6$ A-GNR are shown in Fig. 11. The random distribution of the edge defects gives some variation of the band gap at each defect concentration; however, there exists a general trend of the band gap with increasing defect concentration. The band gap reaches its minimum (which is quite small) at the edge defect ratio of 70%. This implies that edge defects have a significant effect on electronic structures of A-GNRs, which is consistent with the indications from experiments [29].

5 Concluding remarks

Using the recently developed quasi-atomic minimal-basis-set orbitals, we show that accurate tight-binding Hamiltonian and overlap matrix elements can be extracted from first-principles calculations. Based on the information from the QUAMBO-TB matrix elements, the transferability of two-center tight-binding models can be examined from a first-principles perspective. Our studies show that tight-binding models with two-center approximation are not adequate for describing the effective one-electron Hamiltonian matrix elements under
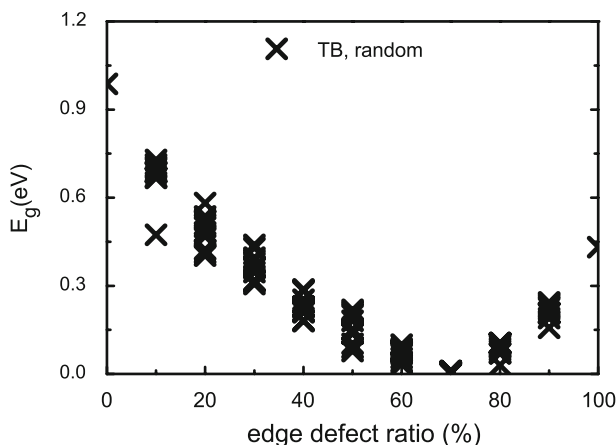
Fig. 11 Band gap as a function of edge defects ratio in an $N_a = 6$ A-GNR with random edge defects. A perfect $N_a = 6$ and $N_a = 5$ A-GNRs corresponds to 0% and 100% defect ratio in this plot respectively

different bonding environments. While we discuss about Si and C systems in this article, similar analyses have been carried out for other systems such as Al, Mo, Fe, SiC etc. [9,11,12]. Such analyses provide useful insights and guidance for generating accurate and transferable tight-binding models. In particular, we show that environment-dependence of the tight-binding parameters need to be adequately described, and it may also be necessary to include three-center integrals explicitly if we want to describe accurately the electronic structures of complex systems by tight-binding approach. Although the QUAMBO-TB scheme can help us gaining insight into how the tight-binding interactions are dependent on the environment, how to model and parameterize such environment-dependence of the tight-binding interaction still remains an open question and need much further investigation.

Another route to utilize the QUAMBO-TB scheme for calculating the electronic structure of a large system is to use a divide-and-conquered strategy which divides the Hamiltonian and overlap matrices of a big system into a set of much smaller $n \times n$ QUAMBO-TB matrices of pair of atoms with different bonding environment. First-principles calculations are needed for small number of atoms around the pairs, yet a QUAMBO-TB matrix for the whole large system can be constructed accurately. We have shown that such "QUAMBO-on-demand" approach has been quite successful for the studies of electronic structure in grapheme nanoribbons. One could construct a variety of training cells to generate a database of Hamiltonian parameters for a catalogue of local bonding environments. This opens a promising avenue to do electronic-structure simulations and total energy calculations for big systems directly from first principles. The computational savings thus achieved is analogous to savings obtained using Green's function boundary condition near infinite half space, but it can handle complex geometric arrangements. A sophisticated computational technology needs to be developed in the future to automate this process.

References

1. Andersen, O.K., Jepsen, O.: Phys. Rev Lett. **53**, 2571 (1984)
2. Andersen, O.K., Sha-Dasgupta, T.: Phys.Rev. B **62**, R16219 (2000)
3. Andersen, O.K., et al.: Electronic structure and physical properties of solid. The uses of the LMTO method. In: Dreysse, H. (ed.) Springer Lecture Notes in Physics. Springer, New York (2000)
4. Sankey, O.F., Niklewski, D.J.: Phys. Rev. B **40**, 3979 (1989)
5. Porezag, D., Frauenheim, Th., Köhler, Th., Seifert, G., Kaschner, R.: Phys. Rev. B **51**, 12947 (1995)
6. McMahan, A.K., Klepeis, J.E.: . Phys. Rev. B **56**, 12250 (1997)
7. Lu, W.C., Wang, C.Z., Schmidt, M.W., Bytautas, L., Ho, K.M., Ruedenberg, K.: J. Chem. Phys. **120**, 2629 (2004)
8. Lu, W.C., Wang, C.Z., Schmidt, M.W., Bytautas, L., Ho, K.M., Ruedenberg, K.: J. Chem. Phys. **120**, 2638 (2004)
9. Lu, W.C., Wang, C.Z., Chan, Z.L., Ruedenberg, K., Ho, K.M.: Phys Rev B **70**, 041101 (R) (2004)
10. Lu, W.C., Wang, C.Z., Ruedenberg, K., Ho, K.M.: Phys. Rev. B **72**, 205123 (2005)
11. Chan, T.-L., Yao, Y.X., Wang, C.Z., Lu, W.C., Li, J., Qian, X.F., Yip, S., Ho, K.M.: Phys. Rev. B **76**, 205119 (2007)
12. Qian, X.-F., Li, J., Wang, C.-Z., Qi, L., Chan, T.-L., Yao, Y.-X., Ho, K.-M., Yip, S.: Phys. Rev. B, (to be published)
13. Slater, J.C., Koster, G.F.: Phys. Rev. **94**, 1498 (1954)
14. Hamann, D.R., Schluter, M., Chiang, C.: Phys. Rev. Lett. **43**, 1494 (1979)
15. Vanderbilt, D.: Phys. Rev. B **41**, 7892 (1990)
16. Blöchl, P.E.: Phys. Rev. B **50**, 17953 (1994)
17. Kresse, G., Hafner, J.: Phys. Rev. B **47**, 558 (1993)
18. Kresse, G., Furthmller, J.: Phys. Rev. B **54**, 169 (1996)
19. Pulay, P.: Chem. Phys. Lett. **100**, 151 (1983)
20. Sæbø, S., Pulay, P.: Chem. Phys. Lett. **113**, 13 (1985)
21. Sæbø, S., Pulay, P.: Annu. Rev. Phys. Chem. **44**, 213 (1993)
22. Löwdin, P.: J. Chem. Phys. **18**, 365 (1950)
23. Tang, M.S., Wang, C.Z., Chan, C.T., Ho, K.M.: Phys. Rev. B **53**, 979 (1996)
24. Wang, C.Z., Pan, B.C., Ho, K.M.: J. Phys. Condens. Matter **11**, 2043 (1999)
25. Yao, Y.X., Wang, C.Z., Zhang, G.P., Ji, M., Ho, K. M.: (to be published)
26. Son, Y.-W., Cohen, M.L., Louie, S.G.: Nature (London) **444**, 347 (2006)
27. Son, Y.-W., Cohen, M.L., Louie, S.G.: Phys. Rev. Lett. **97**, 216803 (2006)
28. Yan, Q.M., Huang, B., Yu, J., Zheng, F.W., Zang, J., Wu, J., Gu, B.L., Liu, F., Duan, W.H.: Nano Lett. **7**, 1469 (2007)
29. Han, M.Y., Ozyilmaz, B., Zhang, Y., Kim, P.: Phys. Rev. Lett. **98**, 206805 (2007)