Accurate tight-binding Hamiltonians for two-dimensional and layered materials

Luis A. Agapito,[1,2] Marco Fornari,[3,4] Davide Ceresoli,[5] Andrea Ferretti,[6] Stefano Curtarolo,[4,7] and Marco Buongiorno Nardelli[2,4,*]

[1]*Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708, USA*
[2]*Department of Physics, University of North Texas, Denton, Texas 76203, USA*
[3]*Department of Physics, Central Michigan University, Mount Pleasant, Michigan 48859, USA*
[4]*Center for Materials Genomics, Duke University, Durham, North Carolina 27708, USA*
[5]*Consiglio Nazionale delle Ricerche, ISTM, Istituto di Scienze e Tecnologie Molecolari, I-20133 Milano, Italy*
[6]*Consiglio Nazionale delle Ricerche, NANO S3 Center, Istituto Nanoscienze, I-41125 Modena, Italy*
[7]*Materials Science, Electrical Engineering, Physics and Chemistry, Duke University, Durham, North Carolina 27708, USA*

(Received 11 January 2016; revised manuscript received 4 March 2016; published 22 March 2016)

We present a scheme to controllably improve the accuracy of tight-binding Hamiltonian matrices derived by projecting the solutions of plane-wave *ab initio* calculations on atomic-orbital basis sets. By systematically increasing the completeness of the basis set of atomic orbitals, we are able to optimize the quality of the band-structure interpolation over wide energy ranges including unoccupied states. This methodology is applied to the case of interlayer and image states, which appear several eV above the Fermi level in materials with large interstitial regions or surfaces such as graphite and graphene. Due to their spatial localization in the empty regions inside or outside of the system, these states have been inaccessible to traditional tight-binding models and even to *ab initio* calculations with atom-centered basis functions.

DOI: 10.1103/PhysRevB.93.125137

I. INTRODUCTION

The generation of highly accurate tight-binding (TB) models for arbitrary systems is a long-lasting problem that has enormous implications in the development of efficient tools for the study of the electronic structure of molecules and solids [1,2], and for applications in accelerated materials development [3]. Tight-binding methods have evolved to include total-energy calculations [4–6] and self-consistent approaches [7–9] greatly extending the applicability. With the introduction of *ab initio* tight-binding Hamiltonians the accuracy of many of these methods has seen a substantial improvement. However, the best representations still rely on *ad hoc* basis sets that need to be iteratively optimized [10–12] and are computationally expensive. In recent papers, we have introduced an efficient scheme to construct optimal tight-binding Hamiltonians projecting the Bloch states obtained from plane-wave (PW) density functional theory (DFT) calculations onto atomic orbitals (AOs) derived directly from the generation of the atomic pseudopotentials [13–15]. In this scheme, the energy range in which the TB Hamiltonian reproduces the original states is limited by the finite number of pseudo-atomic-orbitals (PAOs) that comprise the minimal basis set. As such, only a few unoccupied bands are typically well represented and an accurate description of the conduction states is impossible beyond a few eV. If more conduction states are needed, the basis set needs to be systematically extended.

In this work, we propose a procedure that extends the validity of the TB representation of the band structure to electronic states far above the Fermi level.

This approach is based on the projector augmented wave (PAW) formalism and involves several AOs for each angular momentum that are directly computed from the all electron atomic potential.

The paper is organized as follows: in Sec. II.A we introduce the PAW formalism; since the PAW method requires smoother AO functions, i.e., the PAOs, the pseudization process is presented in Sec. II.B; in Sec. II.C we briefly summarize the projection, filtering and shifting procedure to generate accurate Hamiltonian matrices as originally discussed in Refs. [13,14]; in Sec. II.D we discuss the convergence properties of the unoccupied Kohn-Sham states and their dependence on the basis set representation; finally, in Sec. III we present two examples where we demonstrate the effectiveness of our enhanced scheme by reproducing the interlayer states of graphite and the image states of graphene in a wide energy range.

II. METHODOLOGY

A. The PAW method

DFT in combination with the plane-wave pseudopotential (PP) formalism is one of the most common methods to derive the electronic structure of molecules and solids. Pseudopotentials are constructed to remove core electrons from the Hamiltonian and to reduce drastically the number of plane waves that would be otherwise needed to represent the divergent Coulomb potential close to the nucleus. In addition, pseudo-wave-functions are smooth in the region close to the nuclei because they do not need to be orthogonalized to the core orbitals. The drawback of the PP method is that all the details on the wave functions within a specific distance from the nucleus (the atomic sphere) are lost. This is important when it comes to calculate, for instance, magnetic resonance parameters or core emission spectroscopies [16–18]. In that regard, the PAW method [19] allows reconstructing the full nodal structure of the wave functions near the ions. The basic

*mbn@unt.edu

idea is to "augment" the pseudo-wave-functions with a set of pseudo-partial-waves $|\tilde{\varphi}^n_\alpha\rangle$, which are localized functions centered at position $\mathbf{R}_\alpha$ for each atom α (n is a composite index for the quantum numbers n,l,m). This augmentation procedure is achieved through the application of the $\hat{\mathcal{T}}$ operator to the pseudo-wave-functions, $|\tilde{\psi}_i\rangle$, in order to obtain the all-electron wave functions, $|\psi_i\rangle = \hat{\mathcal{T}}|\tilde{\psi}_i\rangle$. The composite index i indicates the band state and wave vector $\mathbf{k}$ of the Bloch wave. Such an operator is defined as

$$\hat{\mathcal{T}} = \hat{1} + \sum_\alpha \hat{\mathcal{T}}_\alpha,$$
$$\hat{\mathcal{T}}_\alpha = \sum_n (|\varphi^n_\alpha\rangle - |\tilde{\varphi}^n_\alpha\rangle)\langle \tilde{p}^n_\alpha| \quad (1)$$

where $|\varphi^n_\alpha\rangle$ are the all-electron partial waves and $|\tilde{\varphi}^n_\alpha\rangle$ are the corresponding pseudo-partial-waves. The PAW method is grounded on the assumption of completeness of the basis of partial waves in Eq. (1): the wave function can be expanded in terms of partial waves inside the augmentation sphere. In practice, in order to enforce a high degree of completeness one has to include more than one partial wave per angular momentum (typically two or three), where the first corresponds to the bound energy state and the others correspond to unbound states of positive energy. The projectors $|\tilde{p}^n_\alpha\rangle$ are local functions centered at $\mathbf{R}_\alpha$ and vanishing beyond a certain cutoff radius r^c_α; they are determined such that

$$\langle \tilde{p}^n_\alpha | \tilde{\varphi}^{n'}_\alpha \rangle = \delta_{n,n'}, \quad \text{for } |\mathbf{r} - \mathbf{R}_\alpha| < r^c_\alpha. \quad (2)$$

The inclusion of multiple partial waves enables the high transferability and accuracy of the PAW potentials, which are defined by a given set of partial waves, projectors, and cutoff radii. Libraries of PAW datasets for almost all elements in the periodic table are available: PSlibrary [20], GBRV [21], JTH [22], GPAW [23], VASP [24], ATOMPAW [25], and GIPAW, used in the calculation of nuclear magnetic resonance shifts [26].

B. Generation of the PAO basis sets

For given energies ϵ_n, the all-electron atomic-orbital functions $|\phi^n_\alpha\rangle$ are obtained by solving the Schrödinger equation:

$$\left[-\frac{1}{2}\nabla^2 + V^{\mathrm{AE}}_\alpha \right] |\phi^n_\alpha\rangle = \epsilon_n |\phi^n_\alpha\rangle, \quad (3)$$

where V^{AE}_α is the screened all-electron potential found by self-consistently solving the Schrödinger equation for an isolated atom in a given reference electronic configuration. The corresponding smooth atomic orbitals $|\tilde{\phi}^n\rangle$ (dropping the atom index α for simplicity) are obtained by solving the implicit equation: $|\phi^n\rangle = \hat{\mathcal{T}}|\tilde{\phi}^n\rangle$, that is,

$$\tilde{\phi}^n(r) = \phi^n(r) - \sum_m [\varphi^m(r) - \tilde{\varphi}^m(r)] C_{m,n}, \quad (4)$$

where $C_{m,n} = \langle \tilde{p}^m | \tilde{\phi}^n \rangle = \int_0^{r^c} \tilde{p}^m m(r)^* \tilde{\phi}^n(r) r^2 dr$. Notice that we have included only the radial component of the functions, i.e., without the angular part, given by the spherical harmonics Y_{lm}.

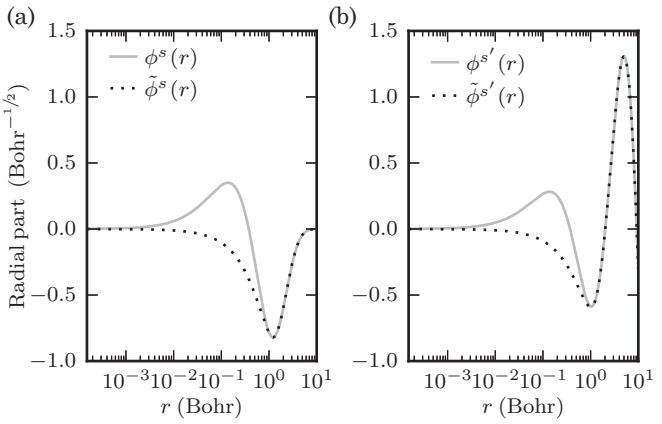

FIG. 1. All-electron and pseudoatomic functions for the $\{s,s'\}$ components of the PAO_3 set. The parameters used to compute the pseudo-atomic-orbitals are defined by the same PAW dataset used in the DFT calculation.

Equation (4) has the form of a Fredholm equation of the second kind [27] which can be reduced to a matrix equation defining $C = (I + A)^{-1}B$ where I, A, and B are identity, square, and rectangular matrices, respectively, with elements $A_{k,h} = \int_0^{r^c} \tilde{p}^k(r)^*[\varphi^h(r) - \tilde{\varphi}^h(r)]r^2dr$, $B_{k,m} = \int_0^{r^c} \tilde{p}^k(r)^*\phi^m(r)r^2dr$. The indices k and h run over the set of projectors and partial waves that define the PAW dataset, respectively, while the indices n and m run over the full set of radial wave functions from Eq. (3).

We have applied the procedure discussed above to derive sets of smooth AOs, i.e., the PAOs $\tilde{\phi}^n$. In a strict sense, while the PAOs are the auxiliary functions involved in the PAW method, the AOs ϕ^n are the functions that constitute the basis set for the tight-binding models. Figure 1 shows a comparison of the radial parts of the all-electron versus the pseudo-atomic-orbitals of carbon (the s and s' functions of PAO_3 given in Table I) using the PAW potential `C.pbe-n-kjpaw_psl.1.0.0.UPF` from Ref. [20]. The PAO functions are smoother than the all-electron functions inside the cutoff radius of 1.4 a_B, while, by construction, both are identical outside the cutoff radius.

One point of strength of this approach is that we can construct PAO sets of increasing size and completeness. The sets can include multiple functions for each $\{lm\}$ channel which correspond to different energy parameters ϵ_n (see Table I). The first choices for the values of ϵ_n are the eigenenergies of the bound states ($\epsilon_n < 0$) of the isolated atom. These orbitals form the minimal set PAO_1 which is equivalent

TABLE I. Energy parameters (in Ry) that define the functions used in the construction of the PAO sets $\{\tilde{\phi}^n(\mathbf{r})\}$ for carbon via Eq. (3).

Set	ϵ_s	ϵ_p	ϵ_d	$\epsilon_{s'}$	$\epsilon_{p'}$	$\epsilon_{s''}$	$\epsilon_{p''}$
PAO_1	−1.01	−0.39					
PAO_2	−1.01	−0.39	0.05				
PAO_3	−1.01	−0.39	0.05	0.05	0.05		
PAO_4	−1.01	−0.39	0.20	0.20	0.20	0.4	0.4

to the single-zeta basis used in quantum chemistry, e.g., $\{s,p\}$ functions for carbon, composed of $M = 4$ functions.

Hamann [28] showed that positive-energy (scattering) states can be employed to improve the accuracy of norm conserving pseudopotentials. Within our approach, we augment the minimal PAO_1 set with scattering states of energies $\epsilon_n > 0$ (these energy parameters can be chosen arbitrarily). The PAO_2 set is of single-zeta-polarized quality that includes polarization functions with higher angular momentum, $l_{\max}+1$, i.e., $\{s,p,d\}$, $M = 9$. PAO_3 is of double-zeta-polarized quality, i.e., $\{s,s',p,p',d\}$, $M = 13$, and PAO_4 triples the number of minimal basis functions and includes additional polarization functions, i.e., $\{s,s',s'',p,p',p'',d\}$, $M = 17$; thus, it is of triple-zeta-polarized quality.

C. Building the TB Hamiltonian matrices

Accurate TB Hamiltonian matrices can be built from the direct projection of the Kohn-Sham (KS) Bloch states $|\psi_{n\mathbf{k}}\rangle$ onto a chosen PAO set as discussed extensively in Refs. [13,14]. This procedure is satisfactory when Bloch states, that project well on the selected AO basis set, are kept and states that do not project well are eliminated, i.e., filtering. In this process the crucial quantities that quantify the accuracy of the basis are the projectabilities $p_{n\mathbf{k}} = \langle\psi_{n\mathbf{k}}|\hat{P}|\psi_{n\mathbf{k}}\rangle \geqslant 0$ ($\hat{P}$ is the operator that projects onto the space of the PAO basis set, as defined in Ref. [14]) which indicate the representability of a Bloch state $|\psi_{n\mathbf{k}}\rangle$ on the chosen PAO set. Maximum projectability, $p_{n\mathbf{k}} = 1$, indicates that the particular Bloch state can be perfectly represented in the chosen PAO set; contrarily, $p_{n\mathbf{k}} \approx 0$ indicates that the PAO set is insufficient and should be augmented. Once the Bloch states with good projectabilities have been identified, the TB Hamiltonian is constructed as

$$H(\mathbf{k}) = AEA^{\dagger} + \kappa[I - A(A^{\dagger}A)^{-1}A^{\dagger}] \qquad (5)$$

where E is the diagonal matrix of KS eigenenergies and A is the matrix of coefficients obtained from projecting the Bloch wave functions onto the PAO set (see Ref. [14].) Since the filtering procedure introduces a null space, the parameter κ is used to shift all the unphysical solutions outside a given energy range of interest.

The real-space TB matrix, $H(\mathbf{R})$, between the central and the neighboring unit cell at lattice vector $\mathbf{R}$ is obtained via Fourier transform:

$$H(\mathbf{R}) = \frac{1}{N_{\mathcal{V}}}\sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}}H(\mathbf{k})\,, \qquad (6)$$

where $N_{\mathcal{V}}$ is the number of $\mathbf{k}$ points in the reciprocal unit cell. Using these matrices, one can calculate the interpolated TB band structure, for any $\mathbf{k}$, using the inverse Fourier transform.

D. TB representation of the unoccupied bands

When a linear combination of AOs (LCAO) is used as basis sets in DFT based methods, the unoccupied bands tend to substantially depend upon basis set size. As an illustration of the above argument, we have computed the band structure of graphite and graphene using both approaches. All PW- and LCAO-DFT calculations presented in this work were performed using the software packages QUANTUM ESPRESSO [29]

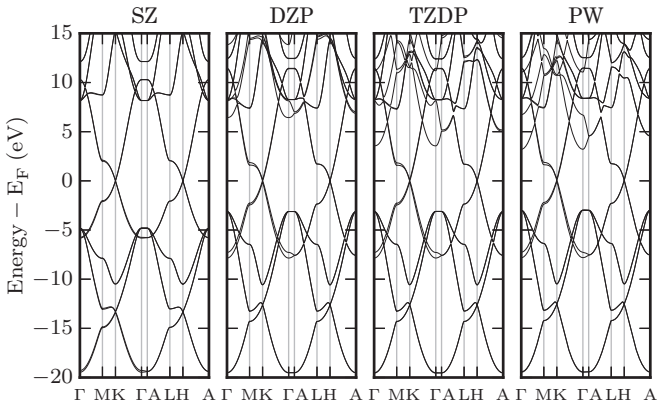

FIG. 2. Band structure of graphite using atomic-orbital (at the single-zeta, double-zeta polarized, and triple-zeta doubly polarized level) and well-converged plane-wave basis sets.

or OPENMX [30] using the Perdew-Burke-Ernzerhof [31] exchange and correlation functional. The PW calculations use the PAW pseudopotential `C.pbe-n-kjpaw_psl.1.0.0.UPF` from the PSlibrary [20] and a kinetic-energy cutoff energy of 40 Ry. The LCAO-DFT calculations use norm conserving pseudopotentials and the optimized AO basis sets from the OPENMX pseudopotential database [32], with a cutoff radius for all carbon basis functions of 7 a_B. The basis set used for the "empty atoms" contains two s, two p, two d, and one f function, all with a cutoff of 13 a_B.

The first three panels in Figs. 2 and 3 show the LCAO-DFT band structures for graphite and graphene, respectively, calculated using AO basis sets of increasing quality, taken from a public database [32]: single zeta (SZ), double zeta with polarization (DZP), and triple zeta with double polarization (TZDP). Under the same approximation to the exchange-correlation functional, calculations using a well-converged plane-wave (PW) basis set (fourth panel) reproduce the unoccupied states systematically better than the AO basis sets.

Not surprisingly, a minimal basis set such as SZ fails to reproduce not only the unoccupied but also the occupied

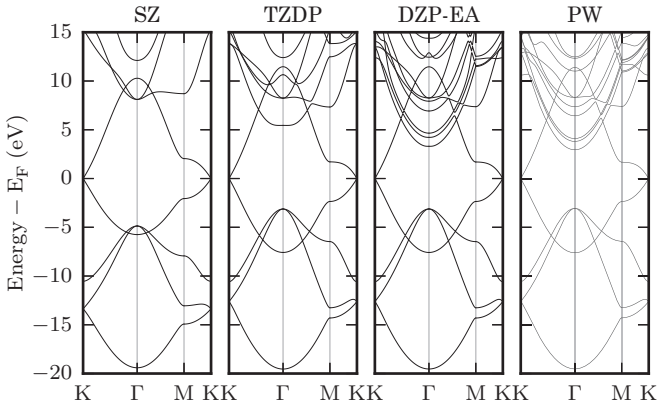

FIG. 3. Band structure of graphene using atomic-orbital (at the single-zeta, double-zeta polarized, and triple-zeta doubly polarized level) and well-converged plane-wave basis sets. The first three panels use atomic-orbital-like sets: SZ, TZDP, and DZP with empty atoms from Ref. [32].

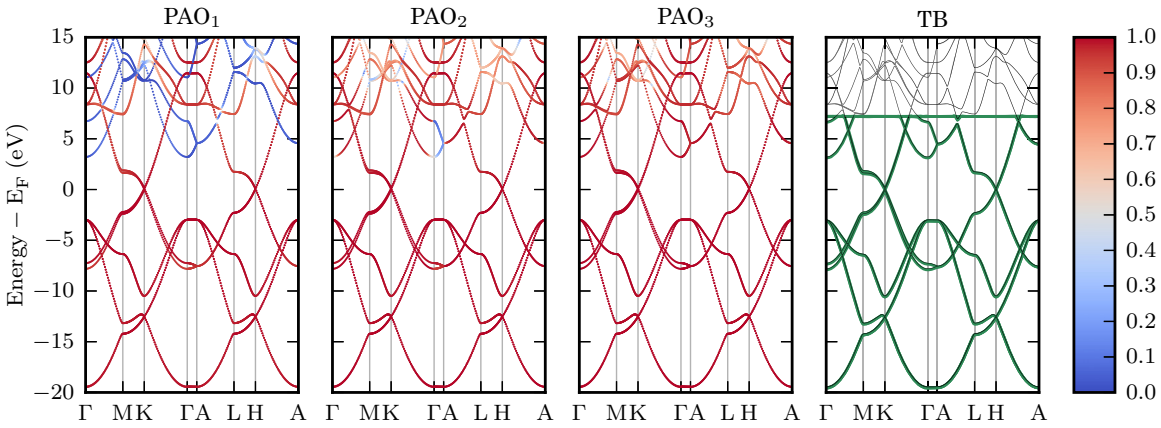

FIG. 4. Projectability $p_{n\mathbf{k}}$, in color scale, of Bloch states $\psi_{n\mathbf{k}}$ of graphite on the PAO_1, PAO_2, and PAO_3 sets. The fourth panel shows the interpolated band structure in green obtained from the TB Hamiltonians built on the PAO_3, superimposed to the plane-wave band structure. The color bar is shown in Fig. 6.

bands of graphite and graphene. The SZ calculation completely misses the lowest conduction band at Γ, as seen in the first panel in Figs. 2 and 3. DZP basis sets are generally considered satisfactory to reproduce ground-state properties, reaching close to chemical accuracy [33]. Indeed, we find that all occupied states are well converged at the DZP level; however, it offers little improvement to the unoccupied bands. Only when the much larger TZDP set is used, the unoccupied bands start to qualitatively match the fully converged PW results.

Despite the deficiencies of standard implementation of DFT, single-particle KS eigenstates (occupied and unoccupied) are often needed, for instance, as the starting point for more refined calculation of the excited states (time-dependent DFT [34], density-functional perturbation theory [35], many-body perturbation theory GW [36], coupled-cluster theory [37], etc.), thus TB Hamiltonians that are expressed in small AO basis sets but that can still deliver the accuracy of the converged PW DFT results, especially for the unoccupied states, are a valuable tool for the study of novel materials and further development of theoretical methods.

III. APPLICATIONS

For most crystalline materials, minimal PAO sets (of SZ or SZP quality) are sufficient for constructing TB Hamiltonians that are describing accurately the band structure up to ∼2 eV above the Fermi energy [13,14]. However, materials containing extended interstitial regions tend to exhibit "interlayer states" usually located several electron volts above the Fermi energy. Similarly, the description of "image states" in metallic surfaces also requires a larger energy range in the unoccupied bands. Here, we discuss two prototypical cases, graphite and graphene, to demonstrate the effectiveness of our augmented PAO basis sets and TB scheme for the treatment of interlayer and image states, which have been out of reach of traditional parameterized TB models so far.

A. Graphite

The first step toward quantifying the quality of LCAO basis sets involves a detailed analysis of the projectabilities. As mentioned before we construct PAO sets of increasing "completeness" and derive an appropriate TB model intended to interpolate the fully converged PW band structure. In Fig. 4 we discriminate between bands with moderate to high projectabilities (plotted in red) and bands with lower projectabilities (plotted in blue). (See the color scale in Fig. 6.) The first panel shows the projectabilities on the PAO_1 set. There are two discernible groups of bands: the states with predominant $\{s,p\}$ character (in red) and the parabolic bands near Γ (in blue). Expectedly, the high projectabilities bands qualitatively resemble the LCAO-DFT calculation with SZ basis in Fig. 2. Conversely, the low projectability bands (in blue), which can not be well represented on the minimal $\{s,p\}$ basis, are absent in Fig. 2 (SZ.) Interestingly, those bands correspond to the so-called interlayer states of graphite [38–40], which are strongly dispersed unoccupied states located in-between the graphitic planes [see Fig. 3(c) in Ref. [40]].

Interlayer states are characteristic of materials with interstitial hollow regions [41,42], layered structures [43], and/or reduced dimensionality such as carbon nanotubes [44,45], C_{60} [46], etc. Naturally, atom-centered basis functions are inappropriate to describe states that extend to the interstitial and/or vacuum regions, whereas PW bases are particularly well suited for this [47]. This is reflected in the pronounced discrepancy between unoccupied eigenenergies calculated using PW and AO basis sets [41,42]. The interlayer states of graphite are not captured with commonly used AO basis sets [48,49] such as DZP, as seen in Fig. 2.

Our PAW-based procedure allows a systematic extension of the PAO sets (see Table I) to distill TB Hamiltonians that capture the details of the band structure including "interlayer states." The performance of each PAO set is assessed by determining its energy range of good projectability for a particular material, which in turn leads to a TB model that is highly accurate within the same energy range.

The performances of the four PAO sets in Table I in terms of projectabilities are shown in Fig. 5. We choose the threshold of $p_{n\mathbf{k}} \geqslant 0.95$ to determine the target energy range of accuracy. The performance of the PAO_1 set (blue line) sharply declines above 3.3 eV due to the presence of parabolic bands that do not

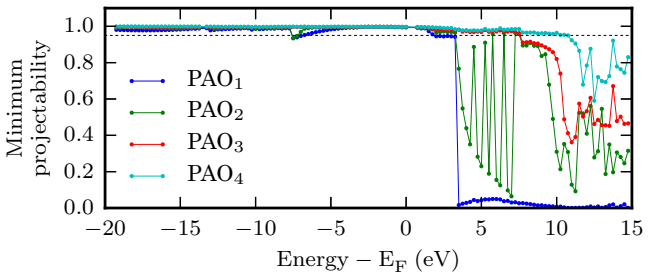

FIG. 5. Projectability of the Bloch states $\psi_{n\mathbf{k}}$ of graphite onto the four PAO sets defined in Table I. The plot shows the minimum value of projectability for each bin of a discretized energy grid. The dotted line indicates the projectability threshold of 0.95. All $\mathbf{k}$ points in the reciprocal unit cell are included in the calculation.

project well on the small LCAO basis. The TB model using this set is accurate only up to 3.3 eV. The larger spatial range of the $\{s', p'\}$ functions added in PAO_2 facilitates the representation of the interlayer states closer to the graphitic planes, yielding a noticeable increase in the projectabilities with respect to PAO_1 over the entire energy range. This is also observed in the second panel in Fig. 4 where most of the blue bands switched to red; the interlayer states with wave-vector component perpendicular to the graphitic planes, those in the Γ-A direction, are still of low projectability. The wave functions of perpendicular Bloch states are primarily localized at the center of the interstitial space, the farthest from the planes, and the $\{s', p'\}$ functions are still insufficient to fully capture them; this is reflected by the oscillations of the green line in the 3.3–7-eV range in Fig. 5. When we augment the basis set with d functions (PAO_3) the range of high projectability increases up to 7.2 eV above the Fermi level. PAO_4 (light blue line) further extends that range up to 10.8 eV. From the above results it is clear that the upper bound of the energy range of high projectability, and consequently the range of accuracy of the TB matrices, can be systematically increased in this way.

The interpolated TB band structure constructed using the PAO_3 set is shown in green in Fig. 4. An excellent agreement with the PW-DFT bands up to 7.2 eV above the Fermi level is observed, as expected from the energy range of high projectabilities deduced from Fig. 5.

B. Graphene

The high-energy electronic bands of graphene are characterized by the presence of image states. Image states give rise to superconductivity in metal-doped graphite [40,50]. Similarly, they play a critical role in several phenomena such as the functionalization of graphene, the adsorption of oxygenated moieties and hydrogen, the formation of defects [51,52], and the "finger-print" peaks in the x-ray-absorption spectra found in the same energy region [53–58]. Image states have also been shown to mediate electron tunneling in graphene [59]. The importance of TB models that reproduce well "interlayer" and "image states" cannot be underestimated especially when designing devices and interpreting experiments.

Image states in graphene follow [60] double Rydberg series $n^{\pm}$ and, expectedly, have low projectability on PAO_1 (blue lines

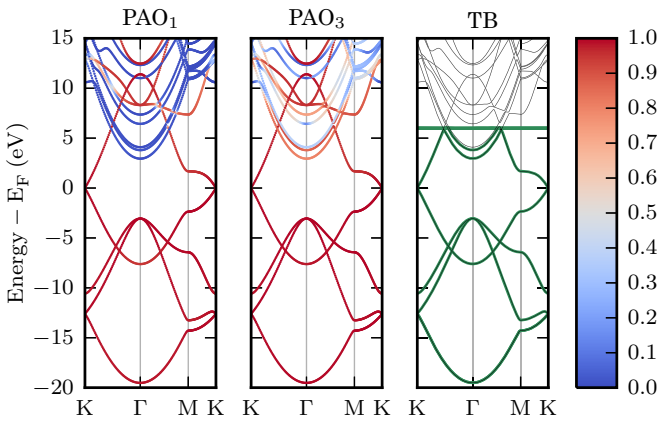

FIG. 6. Projectabilities of the Bloch states of graphene on the PAO_1 and PAO_3 sets. The interpolated TB band structure, computed using the PAO_3 basis set, is shown in green on the third panel superimposed to the PW bands.

in the first panel of Fig. 6). The two lowest unoccupied bands, parabolic at Γ, are the first states of the series, denoted as 1^+ and 1^-. The third lowest unoccupied band corresponds to the state 2^+. Increasing the projectability of the image-state bands by augmenting the basis set is difficult for the particular case of graphene. The projectabilities improve with higher PAO sets, but fail to reach the threshold of 0.95. As seen in the second panel, bands 1^+ and 1^- can reach moderately high projectability (∼0.8 at Γ) with PAO_3; in sharp contrast, however, band 2^+ (light blue) still exhibits low projectability. This behavior is due to the spatial distribution of the image states. Indeed, the wave functions of 1^+ and 1^- around Γ have a component that is localized in the graphene plane forming σ and π hybridizations, respectively [see also Fig. 3(b) in Ref. [60]]. This component can be partially accounted for with the inclusion of $\{s', p'\}$ (and to a lesser extent, d) functions in PAO_3, leading to the observed increase of projectabilities in the second panel with respect to the first in Fig. 6. The other component of the wave functions 1^+ and 1^- is more localized in the vacuum region [60] and, thus, is not captured by the s' and p' functions. Band 1^- loses its dispersion and becomes flat around K, where a marked reduction in projectability is also seen (blue segment at ∼10.6 eV in the second panel.) This happens because around K the wave function of 1^- loses its in-plane π-like component and consequently can no longer be expanded with the $\{s', p'\}$ functions [see also the charge-density plot in Fig. 3(a) in Ref. [61]]. On the other hand, the wave function 2^+ is fully localized in the vacuum region with the position of the maximum electron density away from the graphene plane [60]. Contrary to $1^{\pm}$, it has no in-plane component at Γ and, thus, cannot be represented just by adding basis functions that are centered in the plane; therefore, 2^+ exhibits low projectability on the PAO_3 set.

This is corroborated by examining the LCAO-DFT bands, shown in the first three panels in Fig. 3. Even with the large TZDP basis set (second panel), the band structure completely misses the parabolic image state bands obtained when using the PW basis. The image states are reached only after extending the DZP with empty-atom (EA) basis functions centered at 2.8 Å above and below the graphene sheet. Although the

DZP-EA set is too expensive for practical calculations, it is observed that the bands (third panel) reach qualitative agreement to the PW solution (fourth panel).

As seen in the third panel in Fig. 6, the TB Hamiltonian generated using the PAO_3 set is able to correctly reproduce the band structure (in green) up to the two lowest parabolic bands 1^+ and 1^-. Our method supersedes parameterized tight-binding schemes that are only suitable in the vicinity of the Dirac point. The TB band structure shown here achieves higher accuracy than even the LCAO-DFT result, and with a less expensive basis set. We expect that extending the PAO set with empty-atom functions located off plane will noticeably increase the projectabilities of the interlayer states $1^\pm$ and 2^+ above the threshold of 0.95. We leave this for future investigation.

IV. SUMMARY AND CONCLUSIONS

We presented a scheme to extend PAO basis sets to systematically increase the level of completeness of tight-binding representations obtained from plane-wave *ab initio* calculations. While minimal PAO sets (of SZ or SZP quality) can be sufficient for generating TB Hamiltonian matrices accurate up to ∼2 eV above the Fermi level for most materials, we have shown that enhanced PAO basis sets, containing both negative and positive energy atomic-orbital functions, can controllably increase the energy window in which the TB model is faithfully representing the details of the bands. Results for graphite and graphene, notably very difficult systems to represent in a TB scheme, demonstrate the accuracy and effectiveness of the method.

ACKNOWLEDGMENTS

We want to thank Dr. Derek Stewart for helpful discussions, the Texas Advanced Computing Center at the University of Texas Austin for providing computing facilities, and the funding provided by the ONR Multidisciplinary University Research Initiative under Contract No. N00014-13-1-0635. The authors acknowledge the Duke University Center for Materials Genomics and the CRAY Corporation for computational assistance.

[1] E. Hückel, Quantentheoretische Beiträge zum Benzolproblem, Z. Phys. **70**, 628 (1931).

[2] H. Jones, N. F. Mott, and H. W. B. Skinner, A Theory of the Form of the X-Ray Emission Bands of Metals, Phys. Rev. **45**, 379 (1934).

[3] S. Curtarolo, G. L. W. Hart, M. Buongiorno Nardelli, N. Mingo, S. Sanvito, and O. Levy, The high-throughput highway to computational materials design, Nat. Mater. **12**, 191 (2013).

[4] D. A. Papaconstantopoulos and M. J. Mehl, The Slater-Koster tight-binding method: a computationally efficient and accurate approach, J. Phys. Condens. Matter **15**, R413 (2003).

[5] L. Colombo, Tight-binding theory of native point defects in siicon, Annu. Rev. Mater. Res. **32**, 271 (2002).

[6] O. F. Sankey and D. J. Niklewski, *Ab initio* multicenter tight-binding model for molecular-dynamics simulations and other applications in covalent systems, Phys. Rev. B **40**, 3979 (1989).

[7] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties, Phys. Rev. B **58**, 7260 (1998).

[8] N. Bernstein, J. R. Kermode, and G. Csányi, Hybrid atomistic simulation methods for materials systems, Rep. Prog. Phys. **72**, 026501 (2009).

[9] D. Drabold, A. Demkov, J. P. Lewis, J. Ortega, W. Windl, and S. Lindsay, Large-scale simulations in materials science, Phys. Status Solidi B **248**, 1987 (2011).

[10] O. K. Andersen and T. Saha-Dasgupta, Muffin-tin orbitals of arbitrary order, Phys. Rev. B **62**, R16219 (2000).

[11] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, Maximally localized Wannier functions: Theory and applications, Rev. Mod. Phys. **84**, 1419 (2012).

[12] W. C. Lu, C. Z. Wang, T. L. Chan, K. Ruedenberg, and K. M. Ho, Representation of electronic structures in crystals in terms of highly localized quasiatomic minimal basis orbitals, Phys. Rev. B **70**, 041101 (2004).

[13] L. A. Agapito, A. Ferretti, A. Calzolari, S. Curtarolo, and M. Buongiorno Nardelli, Effective and accurate representation of extended Bloch states on finite Hilbert spaces, Phys. Rev. B **88**, 165127 (2013).

[14] L. A. Agapito, S. Ismail-Beigi, S. Curtarolo, M. Fornari, and M. Buongiorno Nardelli, Accurate tight-binding Hamiltonian matrices from ab-initio calculations: Minimal basis sets, Phys. Rev. B **93**, 035104 (2016).

[15] L. A. Agapito, S. Curtarolo, and M. Buongiorno Nardelli, Reformulation of DFT + U as a Pseudohybrid Hubbard Density Functional for Accelerated Materials Discovery, Phys. Rev. X **5**, 011006 (2015).

[16] B. Pavan, D. Ceresoli, M. M. Tecklenburg, and M. Fornari, First principles NMR study of fluorapatite under pressure, Solid State Nucl. Magn. Reson. **45–46**, 59 (2012).

[17] M. Taillefumier, D. Cabaret, A.-M. Flank, and F. Mauri, X-ray absorption near-edge structure calculations with the pseudopotentials: Application to the K edge in diamond and α-quartz, Phys. Rev. B **66**, 195107 (2002).

[18] A. Pasquarello, M. S. Hybertsen, and R. Car, Theory of Si 2 p core-level shifts at the Si(001)-SiO_2 interface, Phys. Rev. B **53**, 10942 (1996).

[19] P. E. Blöchl, Generalized separable potentials for electronic-structure calculations, Phys. Rev. B **41**, 5414 (1990).

[20] A. Dal Corso, Pseudopotentials periodic table: From H to Pu, Comp. Mat. Sci. **95**, 337 (2014).

[21] K. F. Garrity, J. W. Bennett, K. M. Rabe, and D. Vanderbilt, Pseudopotentials for high-throughput DFT calculations, Comp. Mat. Sci. **81**, 446 (2014).

[22] F. Jollet, M. Torrent, and N. Holzwarth, Generation of Projector Augmented-Wave atomic data: A 71 element validated table in the XML format, Comput. Phys. Commun. **185**, 1246 (2014).

[23] https://wiki.fysik.dtu.dk/gpaw/setups/setups.html.

[24] G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B **59**, 1758 (1999).

[25] http://users.wfu.edu/natalie/papers/pwpaw/newperiodictable/.

[26] D. Ceresoli, N. Marzari, M. G. Lopez, and T. Thonhauser, *Ab initio* converse NMR approach for pseudopotentials, Phys. Rev. B **81**, 184424 (2010).

[27] R. P. Kanwal, *Linear Integral Equations* (Springer, New York, 2013).

[28] D. R. Hamann, Generalized norm-conserving pseudopotentials, Phys. Rev. B **40**, 2980 (1989).

[29] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials, J. Phys. Conden. Matt. **21**, 395502 (2009).

[30] T. Ozaki, Variationally optimized atomic orbitals for large-scale electronic structures, Phys. Rev. B **67**, 155108 (2003).

[31] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. **77**, 3865 (1996).

[32] T. Ozaki and H. Kino, Numerical atomic basis orbitals from H to Kr, Phys. Rev. B **69**, 195113 (2004).

[33] M. Gusso, Study on the maximum accuracy of the pseudopotential density functional method with localized atomic orbitals versus plane-wave basis sets, J. Chem. Phys. **128**, 044102 (2008).

[34] M. Petersilka, U. J. Gossmann, and E. K. U. Gross, Excitation Energies from Time-Dependent Density-Functional Theory, Phys. Rev. Lett. **76**, 1212 (1996).

[35] A. Görling and M. Levy, DFT ionization formulas and a DFT perturbation theory for exchange and correlation, through adiabatic connection, Int. J. Quantum Chem. **56**, 93 (1995).

[36] M. S. Hybertsen and S. G. Louie, Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies, Phys. Rev. B **34**, 5390 (1986).

[37] A. G. Taube and R. J. Bartlett, Frozen natural orbital coupled-cluster theory: Forces and application to decomposition of nitroethane, J. Chem. Phys. **128**, 164101 (2008).

[38] M. Posternak, A. Baldereschi, A. J. Freeman, E. Wimmer, and M. Weinert, Prediction of Electronic Interlayer States in Graphite and Reinterpretation of Alkali Bands in Graphite Intercalation Compounds, Phys. Rev. Lett. **50**, 761 (1983).

[39] D. A. Fischer, R. M. Wentzcovitch, R. G. Carr, A. Continenza, and A. J. Freeman, Graphitic interlayer states: A carbon *K* near-edge x-ray-absorption fine-structure study, Phys. Rev. B **44**, 1427 (1991).

[40] G. Csányi, P. Littlewood, A. H. Nevidomskyy, C. J. Pickard, and B. Simons, The role of the interlayer state in the electronic structure of superconducting graphite intercalated compounds, Nat. Phys. **1**, 42 (2005).

[41] Y.-i. Matsushita, S. Furuya, and A. Oshiyama, Floating Electron States in Covalent Semiconductors, Phys. Rev. Lett. **108**, 246404 (2012).

[42] Y.-i. Matsushita and A. Oshiyama, Interstitial Channels that Control Band Gaps and Effective Masses in Tetrahedrally Bonded Semiconductors, Phys. Rev. Lett. **112**, 136403 (2014).

[43] X. Blase, A. Rubio, S. G. Louie, and M. L. Cohen, Quasiparticle band structure of bulk hexagonal boron nitride and related systems, Phys. Rev. B **51**, 6868 (1995).

[44] S. Okada, A. Oshiyama, and S. Saito, Nearly free electron states in carbon nanotube bundles, Phys. Rev. B **62**, 7634 (2000).

[45] J. W. Chiou, C. L. Yueh, J. C. Jan, H. M. Tsai, W. F. Pong, I.-H. Hong, R. Klauser, M.-H. Tsai, Y. K. Chang, Y. Y. Chen, C. T. Wu, K. H. Chen, S. L. Wei, C. Y. Wen, L. C. Chen, and T. J. Chuang, Electronic structure of the carbon nanotube tips studied by x-ray-absorption spectroscopy and scanning photoelectron microscopy, Appl. Phys. Lett. **81**, 4189 (2002).

[46] M. Feng, J. Zhao, and H. Petek, Atomlike, Hollow-Core–Bound Molecular Orbitals of C60, Science **320**, 359 (2008).

[47] D. Stewart, A Cautionary Tale of Two Basis Sets and Graphene, Comput. Sci. Eng. **14**, 55 (2012).

[48] L. Samuelson and I. P. Batra, Electronic properties of various stages of lithium intercalated graphite, J. Phys. C **13**, 5105 (1980).

[49] S. B. Trickey, F. Müller-Plathe, G. H. F. Diercksen, and J. C. Boettger, Interplanar binding and lattice relaxation in a graphite dilayer, Phys. Rev. B **45**, 4460 (1992).

[50] T. E. Weller, M. Ellerby, S. S. Saxena, R. P. Smith, and N. T. Skipper, Superconductivity in the intercalated graphite compounds C_6Yb and C_6Ca, Nat. Phys. **1**, 39 (2005).

[51] V. A. Coleman, R. Knut, O. Karis, H. Grennberg, U. Jansson, R. Quinlan, B. C. Holloway, B. Sanyal, and O. Eriksson, Defect formation in graphene nanosheets by acid treatment: an x-ray absorption spectroscopy and density functional theory study, J. Phys. Chem. **41**, 062001 (2008).

[52] C. Ehlert, W. E. S. Unger, and P. Saalfrank, C K-edge NEXAFS spectra of graphene with physical and chemical defects: a study based on density functional theory, Phys. Chem. Chem. Phys. **16**, 14083 (2014).

[53] D. Pacilé, M. Papagno, A. Fraile Rodríguez, M. Grioni, L. Papagno, Ç. Ö. Girit, J. C. Meyer, G. E. Begtrup, and A. Zettl *et al.*, *Pacilé Reply:*, Phys. Rev. Lett. **102**, 099702 (2009).

[54] D. Pacilé, M. Papagno, A. F. Rodríguez, M. Grioni, L. Papagno, Ç. Girit, J. C. Meyer, G. E. Begtrup, and A. Zettl, Near-Edge X-Ray Absorption Fine-Structure Investigation of Graphene, Phys. Rev. Lett. **101**, 066806 (2008).

[55] M. Papagno, A. F. Rodríguez, Ç. Girit, J. Meyer, A. Zettl, and D. Pacilé, Polarization-dependent C K near-edge X-ray absorption fine-structure of graphene, Chem. Phys. Lett. **475**, 269 (2009).

[56] H.-K. Jeong, H.-J. Noh, J.-Y. Kim, L. Colakerol, P.-A. Glans, M. H. Jin, K. E. Smith, and Y. H. Lee, Comment on Near-Edge X-Ray Absorption Fine-Structure Investigation of Graphene, Phys. Rev. Lett. **102**, 099701 (2009).

[57] B. J. Schultz, C. J. Patridge, V. Lee, C. Jaye, P. S. Lysaght, C. Smith, J. Barnett, D. A. Fischer, D. Prendergast, and S. Banerjee, Imaging local electronic corrugations and doped regions in graphene, Nat. Commun. **2**, 372 (2011).

[58] W. Hua, B. Gao, S. Li, H. Ågren, and Y. Luo, X-ray absorption spectra of graphene from first-principles simulations, Phys. Rev. B **82**, 155433 (2010).

[59] Y. Zhang, V. W. Brar, F. Wang, C. Girit, Y. Yayon, M. Panlasigui, A. Zettl, and M. F. Crommie, Giant phonon-induced conductance in scanning tunneling spectroscopy of gate-tunable graphene, Nat. Phys. **4**, 627 (2008).

[60] V. M. Silkin, J. Zhao, F. Guinea, E. V. Chulkov, P. M. Echenique, and H. Petek, Image potential states in graphene, Phys. Rev. B **80**, 121408 (2009).

[61] E. Kogan, V. U. Nazarov, V. M. Silkin, and M. Kaveh, Energy bands in graphene: Comparison between the tight-binding model and *ab initio* calculations, Phys. Rev. B **89**, 165430 (2014).