12-16-2008

Quasiatomic orbitals for ab initio tight-binding analysis

Xiaofeng Qian
Massachusetts Institute of Technology

Ju Li
University of Pennsylvania, liju@seas.upenn.edu

Liang Qi
University of Pennsylvania, liangqi@seas.upenn.edu

Cai-Zhuang Wang
Iowa State University

Tzu-Liang Chan
Iowa State University

See next page for additional authors

Quasiatomic orbitals for ab initio tight-binding analysis Xiaofeng Qian, Ju Li, Liang Qi, Cai-Zhuang Wang, Tzu-Liang Chan, Yong-Xin Yao, Kai-Ming Ho, and Sidney Yip, Phys. Rev. B 78, 245112 (2008), DOI:10.1103/PhysRevB.78.245112

Publisher URL: http://link.aps.org/doi/10.1103/PhysRevB.78.245112

Quasiatomic orbitals for ab initio tight-binding analysis

Abstract
Wave functions obtained from plane-wave density-functional theory (DFT) calculations using norm-conserving pseudopotential, ultrasoft pseudopotential, or projector augmented-wave method are efficiently and robustly transformed into a set of spatially localized nonorthogonal quasiatomic orbitals (QOs) with pseudoangular momentum quantum numbers. We demonstrate that these minimal-basis orbitals can exactly reproduce all the electronic structure information below an energy threshold represented in the form of environment-dependent tight-binding Hamiltonian and overlap matrices. Band structure, density of states, and the Fermi surface are calculated from this real-space tight-binding representation for various extended systems (Si, SiC, Fe, and Mo) and compared with plane-wave DFT results. The Mulliken charge and bond order analyses are performed under QO basis set, which satisfy sum rules. The present work validates the general applicability of Slater and Koster's scheme of linear combinations of atomic orbitals and points to future ab initio tight-binding parametrizations and linear-scaling DFT development.

Keywords
AUGMENTED-WAVE METHOD, PARRINELLO MOLECULAR-DYNAMICS, LOCALIZED WANNIER FUNCTIONS, COMPOSITE ENERGY-BANDS, MINIMAL BASIS-SETS, ULTRASOFT PSEUDOPOTENTIALS, ELECTRON CORRELATION, MODEL, CRYSTALS, DENSITY

Comments
Quasiatomic orbitals for ab initio tight-binding analysis Xiaofeng Qian, Ju Li, Liang Qi, Cai-Zhuang Wang, Tzu-Liang Chan, Yong-Xin Yao, Kai-Ming Ho, and Sidney Yip, Phys. Rev. B 78, 245112 (2008), DOI:10.1103/PhysRevB.78.245112

Publisher URL: http://link.aps.org/doi/10.1103/PhysRevB.78.245112

Author(s)
Xiaofeng Qian, Ju Li, Liang Qi, Cai-Zhuang Wang, Tzu-Liang Chan, Yong-Xin Yao, Kai-Ming Ho, and Sidney Yip

Quasiatomic orbitals for *ab initio* tight-binding analysis

Xiaofeng Qian,[1] Ju Li,[2,*] Liang Qi,[2] Cai-Zhuang Wang,[3] Tzu-Liang Chan,[3] Yong-Xin Yao,[3] Kai-Ming Ho,[3] and Sidney Yip[1]

[1]*Department of Nuclear Science and Engineering and Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

[2]*Department of Materials Science and Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*

[3]*Department of Physics and Ames Laboratory U.S. DOE, Iowa State University, Ames, Iowa 50011, USA*

(Received 25 March 2008; revised manuscript received 31 October 2008; published 16 December 2008)

Wave functions obtained from plane-wave density-functional theory (DFT) calculations using norm-conserving pseudopotential, ultrasoft pseudopotential, or projector augmented-wave method are efficiently and robustly transformed into a set of spatially localized nonorthogonal quasiatomic orbitals (QOs) with pseudoangular momentum quantum numbers. We demonstrate that these minimal-basis orbitals can exactly reproduce all the electronic structure information below an energy threshold represented in the form of environment-dependent tight-binding Hamiltonian and overlap matrices. Band structure, density of states, and the Fermi surface are calculated from this real-space tight-binding representation for various extended systems (Si, SiC, Fe, and Mo) and compared with plane-wave DFT results. The Mulliken charge and bond order analyses are performed under QO basis set, which satisfy sum rules. The present work validates the general applicability of Slater and Koster's scheme of linear combinations of atomic orbitals and points to future *ab initio* tight-binding parametrizations and linear-scaling DFT development.

DOI: 10.1103/PhysRevB.78.245112 PACS number(s): 71.15.Ap, 71.18.+y, 71.20.−b

I. INTRODUCTION

Density-functional theory (DFT) (Refs. 1 and 2) has been extensively developed in the past decades. For condensed-matter systems, efficient supercell calculations using plane-wave basis and ultrasoft pseudopotential (USPP) (Refs. 3–6) or projector augmented wave (PAW) (Ref. 7) are now the mainstream. Plane-wave basis is easy to implement. Its quality is continuously tunable and spatially homogeneous. The drawback is that this "rich basis" can sometimes mask the physical ingredients of a problem, making their detection and distillation difficult. This becomes particularly clear when one wants to develop a parametrized tight-binding (TB) potential[8–10] or classical empirical potential[11] based on plane-wave DFT results, often a crucial step in multiscale modeling.[12] For developing TB potentials, one usually fits to the DFT total energy, forces, and quasiparticle energies $\{\varepsilon_n\}$ (band diagram). However the plane-wave electronic-structure information is still vastly underutilized in this TB potential development process.

Modern TB approach assumes the existence of a minimal basis of dimension qN, where N is the number of atoms and q is a small prefactor (four for Si), without explicitly stating what these basis orbitals are. Under this minimal basis, the electronic Hamiltonian is represented by a small matrix $\mathbf{H}^{\mathrm{TB}}_{qN\times qN}$, which is parametrized[13] and then explicitly diagonalized at runtime to get $\{\varepsilon_n^{\mathrm{TB}}\}$. In contrast, under plane-wave basis the basis-space dimension is pN, where p is a large number, usually 10^2–10^3. The Kohn-Sham (KS) Hamiltonian represented under the plane-wave basis, $\mathbf{H}^{\mathrm{KS}}_{pN\times pN}$, is often so large that it cannot be stored in computer memory. So instead of direct diagonalization which yields the entire eigenspectrum, matrix-free algorithms that only call upon matrix-vector products are employed to find just a small portion of the eigenspectrum $\{\varepsilon_n\}$ at the low-energy end.[14] This is wise because the ground-state total energy and a great majority of the system's physical properties depend only on a small portion of the electronic eigenstates with ε_n below or near the Fermi energy ε_F.

Unlike many *ab initio* approaches that adopt explicit spatially localized basis sets such as Slater-type orbitals (STOs) and Gaussian-type orbitals (GTOs),[15] the defining characteristic of the *empirical* TB approach is the unavailability of the minimal-basis orbitals, which are declared to exist but never shown explicitly. This leads to the following conundrum. In constructing material-specific TB potentials,[8–10] the $\mathbf{H}^{\mathrm{TB}}_{qN\times qN}$ matrix is parametrized but the $qN(qN+1)/2$ matrix elements are not targets of fitting themselves because one does not have access to their values since one never knows the minimal-basis orbitals to start with. Instead, the fitting targets are the eigenvalues of $\mathbf{H}^{\mathrm{TB}}_{qN\times qN}$ and $\{\varepsilon_n^{\mathrm{TB}}\}$, which are demanded to match the occupied eigenvalues $\{\varepsilon_n\}_{\mathrm{occ}}$ of $\mathbf{H}^{\mathrm{KS}}_{pN\times pN}$ from plane-wave DFT calculation and perhaps a few unoccupied $\{\varepsilon_n\}$ as well. A transferable TB potential should have the correct physical ingredients; but a great difficulty arises here because $\{\varepsilon_n\}$ in fact contain much less information than the $\mathbf{H}^{\mathrm{TB}}_{qN\times qN}$ matrix elements. From $\mathbf{H}^{\mathrm{TB}}_{qN\times qN}$ matrix we can get $\{\varepsilon_n^{\mathrm{TB}}\}$ but not vice versa. As fitting targets, not only are the $\{\varepsilon_n^{\mathrm{TB}}\}$ much fewer in number than the matrix elements $[qN$ versus $qN(qN+1)/2]$ but they are also much less physically transparent. The TB matrix elements must convey clear spatial (both position and orientation) information, as is evident from the $pp\pi$, $pd\sigma$, $dd\delta$, etc. analytic angular functions in the original Slater-Koster linear combination of atomic orbitals (LCAO) (Ref. 16) scheme. Physichemical effects such as charge transfer, saturation, and screening[8–10] should manifest more directly in the matrix elements; but such information gets scrambled after diagonalization. For example, if the fifth eigenvalue $\varepsilon_{n=5}^{\mathrm{TB}}$ at $\mathbf{k}=[111]\pi/3a$ in β-SiC crystal is lower than that of plane-wave DFT by 0.2 eV, should one increase the screening term[8–10] in the TB model to get a better fit or not? The answer will not be at all obvious since (a) the **k**-space result masks the real-space physics and (b) the eigenvalue reflects nothing about the spatial features of the eigenfunction $|\psi_{n\mathbf{k}}\rangle$. The information necessary for answering the question is hidden in the wave functions $\{\psi_n\}$ (now expanded in plane waves) and the electronic Hamiltonian $\mathbf{H}^{\mathrm{KS}}_{pN\times pN}$ (now a huge matrix). But the clues are simply not sufficiently embedded in $\{\varepsilon_n\}$, which do not contain any spatial information.[17] Thus, the present empirical TB approach is similar to "shooting in the dark."

It is thus desirable to come up with a systematic and numerically robust method to distill information from plane-wave DFT calculation into a TB representation. Philosophically this is the same as the "downfolding" procedure of Andersen and Saha-Dasgupta.[18] Namely, can we construct the minimal-basis functions from $\{\psi_n\}$ explicitly? Can we get $\mathbf{H}^{\mathrm{TB}}_{qN\times qN}$ from $\mathbf{H}^{\mathrm{KS}}_{pN\times pN}$? This $\mathbf{H}^{\mathrm{KS}}_{pN\times pN}\rightarrow\mathbf{H}^{\mathrm{TB}}_{qN\times qN}$ mapping would work similar to a computer file compression because $\mathbf{H}^{\mathrm{KS}}_{pN\times pN}$ is a huge matrix and $\mathbf{H}^{\mathrm{TB}}_{qN\times qN}$ is small. Can then the compression be lossless? That is, can we retain exactly the occupied eigenspectrum $\{\varepsilon_n\}_{\mathrm{occ}}$ of $\mathbf{H}^{\mathrm{KS}}_{pN\times pN}$ and perhaps a few unoccupied $\{\varepsilon_n\}$ as well? For modeling the total energy of the system, only the occupied bands are important. But if one is interested in transport properties,[19] the low-energy portion of the unoccupied bands will be important as well.

In this paper we present an explicit *ab initio* TB matrix construction scheme based on plane-wave DFT calculations. The present scheme is significantly improved over our previous developments[20–24] in efficiency and stability and now extended to work with USPP/PAW formalisms and popular DFT programs such as VASP (Refs. 6 and 25) and DACAPO.[26] The improved scheme no longer requires the computation and storage of the wave functions of hundreds of unoccupied DFT bands, reducing disk, memory, and CPU time requirements by orders of magnitude. But one also obtains converged quasiatomic orbitals (QOs) of the previous scheme[20–24] as if infinite number of unoccupied bands were taken—the "infiniband" limit that eliminates the so-called unoccupied bands truncation error (UBTE). The source code of our method and input conditions for all examples in this paper are put on the web.[27] We will demonstrate through a large number of examples that an "atomic orbital (AO)-like" minimal basis can generally be constructed and are sufficiently localized for both insulators and metals. These examples[27] demonstrate the physical soundness underlying the environment-dependent TB approach.[8] While we stop short of giving material-specific parametrizations for the $\mathbf{H}^{\mathrm{TB}}_{qN\times qN}$ matrix elements, their physical properties will be discussed with a view toward explicit parametrizations[8–10] later.

Our method follows the general approach of the Wannier function (WF),[28–40] which combines Bloch eigenstates obtained from periodic cell calculation in **k** space to achieve good localization in real space. Other than chemical analysis, linear-scaling (order-N) methods,[41–44] transport,[45–47] modern theory of polarization[17] and magnetization,[48] LDA+U (Refs. 49–51) and self-interaction correction,[52] etc., also rely on high-quality localized basis set. The WF approach guarantees exact reproducibility of the occupied subspace and exponential localization in the case of a single band[53] and isolated bands in insulators.[54]

There is some indeterminacy ("gauge" freedom[55,56]) in the WF approach. One could multiply a smooth phase function on the Bloch band states, and they would still be smooth Bloch bands. One could also mix different band branches and still maintain unitarity of the WF transform. Marzari and Vanderbilt[32] proposed the concept of maximally localized Wannier functions (MLWFs) for an isolated group of bands using the quadratic spread localization measure originally proposed by Foster and Boys[57] for molecular systems. Later Souza *et al.*[34] extended this scheme for entangled bands by optimizing a subspace from a larger Hilbert space within a certain energy window. Choosing the MLWF gauge for a given energy window removes most indeterminacy in the WF transform. Unfortunately, there is no closed-form solution for MLWF; so iterative numerical procedures must be adopted, associated with which is the problem of finding global minima. Despite the tremendous success of the MLWF approach,[32,34] there are still something to be desired of in the way of a robust and physically transparent algorithm, resulting in a great deal of recent activities.[20–24,35–40]

Here we take a different strategy.[20–24] While maximal localization is a worthy goal, if there is no analytical solution its attainment is sometimes uncertain. The question is, does one really need *maximal* localization? May one be satisfied if a set of WF orbitals can be constructed robustly, and they are "localized enough"? The quasiatomic minimal-basis orbitals (QUAMBOs) (Refs. 20–24) are constructed based on the projection operation where one demands maximal similarity between the minimal-basis orbitals with preselected atomic orbitals with angular momentum quantum numbers. Since "maximal similarity" is a quadratic problem, it has exact solution and the numerical procedure is noniterative and relatively straightforward. On the other hand, whether these maximally similar WF orbitals are localized enough for the practical purpose of *ab initio* TB analysis and constructing *ab initio* TB potentials needs to be demonstrated, through a large number of examples. Early results are encouraging. We note that philosophically these minimal-basis orbitals "maximally similar" to atomic orbitals are probably closest to the original idea of Slater and Koster[16] of linear combinations of atomic orbitals since using true atomic orbitals as minimal basis leads to very poor accuracy compared to present-day empirical TB potentials.[8]

This paper is organized as follows. In Sec. II we review USPP and PAW formalisms required for properly defining projection. In Sec. III nonorthogonal QOs within USPP and PAW formalisms are derived for extended systems. In Sec. IV *ab initio* tight-binding Hamiltonian and overlap matrices are derived under the QO basis set. The Mulliken charge and bond order (BO) analyses are also formulated for QO. To demonstrate the efficiency and robustness of this method, in Sec. V band structure, total density of states (DOS), QO-projected band structure, QO-projected density of states (PDOS), and the Fermi surface are calculated and compared with plane-wave DFT results for various extended systems (Si, β-SiC, Fe, and Mo). In Sec. VI we discuss the similarity and difference between QO and other localized orbitals. In Sec. VII we summarize our work and discuss some future applications of quasiatomic orbitals. Finally, in the Appendix we mathematically prove that QO is equivalent to the infinite band limit of the quasiatomic minimal-basis orbital by Lu and co-workers.[20–24]

II. PROJECTION OPERATION IN USPP/PAW

The computational cost of plane-wave DFT calculations is strongly dependent on the selected type of pseudopotentials. Compared to more traditional norm-conserving pseudopotentials (NCPPs), Vanderbilt's USPP,[3–5] and Blöchl's PAW method[7] achieve dramatic savings for $2p$ and $3d$ elements with minimal loss of accuracy. In this paper we implement QO method with NCPP, USPP, and PAW method, which are used in popular DFT codes such as VASP,[6,25] DACAPO,[26] PWSCF,[58] CPMD,[59] CP-PAW,[60] and ABINIT.[61] Currently we have implemented interfaces to VASP and DACAPO.[27] The formalisms of USPP/PAW method were reviewed in Ref. 19. Here we just highlight the part important to quasiatomic orbitals, which is the metric operator $\hat{S}$.

The key idea behind USPP and PAW method is a mapping of the true valence electron wave function $\tilde{\psi}(\mathbf{x})$ to a pseudo wave function $\psi(\mathbf{x}):\tilde{\psi}\leftrightarrow\psi$, just as in any pseudopotential scheme. However, by discarding the requirement that $\psi(\mathbf{x})$ must be norm conserved ($\langle\psi|\psi\rangle=1$) while matching $\tilde{\psi}(\mathbf{x})$ outside the pseudopotential cutoff, a greater smoothness of $\psi(\mathbf{x})$ in the core region can be achieved; and therefore less plane waves are required to represent $\psi(\mathbf{x})$. In order for the physics to still work, in USPP and PAW schemes one must define augmentation charges in the core region and solve a generalized eigenvalue problem,

$$\hat{H}|\psi_n\rangle = \varepsilon_n\hat{S}|\psi_n\rangle, \tag{1}$$

where $\hat{S}$ is a Hermitian and positive definite operator. $\hat{S}$ defines the fundamental metric of the linear Hilbert space of pseudo wave functions. Since in USPP and PAW methods the pseudo wave functions do not satisfy the norm-conserving property, the inner product (ψ,ψ') between two pseudo wave functions is always $\langle\psi|\hat{S}|\psi'\rangle$ instead of $\langle\psi|\psi'\rangle$. The $\hat{S}$ operator is given by

$$\hat{S} = 1 + \sum_{i,j,I} q_{ij}^I|\beta_i^I\rangle\langle\beta_j^I|, \tag{2}$$

where $i\equiv\{\tau lm\sigma\}$ and I labels the ions. τ and lm are the orbital radial and angular quantum numbers.[4] σ is the spin. In this paper, all "orbitals" are meant to be spin orbitals although in the case of spin-unpolarized calculations, there is a degeneracy of 2 in the orbital wave function and eigenenergy. In above, the projector wave function $\beta_i^I(\mathbf{x})\equiv\langle\mathbf{x}\sigma|\beta_i^I\rangle$ of atom I's channel i is

$$\beta_i^I(\mathbf{x}) = \beta_i(\mathbf{x}-\mathbf{X}_I), \tag{3}$$

where $\mathbf{X}_I$ is the ion position, and $\beta_i(\mathbf{x})$ vanishes outside the pseudopotential cutoff.

Just like $\hat{H}$, $\hat{S}$ contains contributions from all ions. Consider a parallelepiped computational supercell of volume Ω, with N ions inside. One usually performs $L_1\times L_2\times L_3$ $\mathbf{k}$ sampling in the supercell's first Brillouin zone. For the sake of clarity, let us define a Born–von Kármán (Bv) universe, which is an $L_1\times L_2\times L_3$ replica of the computational supercell, periodically wrapped around. So the Bv universe has finite volume $L_1L_2L_3\Omega$, with a total of $L_1L_2L_3N$ ions. Using Bloch's theorem, it is easy to show that all the eigenstates in the Bv universe can be labeled by $L_1L_2L_3$ $\mathbf{k}$'s of the Monkhorst-Pack $\mathbf{k}$ mesh.[62] The basic metric of function length and inner product should be defined in the Bv universe,

$$(\psi,\psi') \equiv \langle\psi|\hat{S}|\psi'\rangle = \delta_{\sigma\sigma'}\int_{\mathrm{Bv}} d^3\mathbf{x}\,\psi^*(\mathbf{x})(\hat{S}|\psi'\rangle)(\mathbf{x}). \tag{4}$$

$\hat{S}$ above contains contributions from all $L_1L_2L_3N$ ions. With the inner product defined in Eq. (4), the projection of any state $|\phi\rangle$ on $|\psi\rangle$ is straightforward;

$$\hat{P}_\psi|\phi\rangle \equiv \frac{(\psi,\phi)}{(\psi,\psi)}|\psi\rangle = \frac{\langle\psi|\hat{S}|\phi\rangle}{\langle\psi|\hat{S}|\psi\rangle}|\psi\rangle. \tag{5}$$

Note that all functions discussed in this paper must be nominally periodic in the Bv universe. $|\phi\rangle$ could be AO-like. Even though real AOs are represented in infinite space, this is not a problem numerically so long as the AO extent is much smaller than the size of the Bv universe. (The AO extent does not need to be smaller than the computational supercell Ω.)

It is easy to show that if

$$\psi(\mathbf{x}-\mathbf{a}) = \psi(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{a}}, \quad \psi'(\mathbf{x}-\mathbf{a}) = \psi'(\mathbf{x})e^{-i\mathbf{k}'\cdot\mathbf{a}}, \tag{6}$$

where $\mathbf{k},\mathbf{k}'\in L_1\times L_2\times L_3$ $\mathbf{k}$ mesh and $\mathbf{a}=l_1\mathbf{a}_1+l_2\mathbf{a}_2+l_3\mathbf{a}_3$ is an arbitrary integer combination of supercell edge vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$, then ψ and ψ' will be orthogonal in the sense of Eq. (4) unless $\mathbf{k}=\mathbf{k}'$. Consequently we can label ψ by $\mathbf{k}$, e.g., $\psi_{\mathbf{k}}(\mathbf{x})$ and $\psi'_{\mathbf{k}'}(\mathbf{x})$. $\psi_{\mathbf{k}}(\mathbf{x})$ can be expressed as the product of a phase modulation $e^{i\mathbf{k}\cdot\mathbf{x}}$ and a periodic function $u_{\mathbf{k}}(\mathbf{x})$ within Ω. It is always advantageous to "think" in the Bv universe; but employing Bloch's theorem we often only need to "compute" in the Ω supercell.

III. QUASIATOMIC ORBITAL CONSTRUCTION

From a plane-wave calculation using USPP or PAW method, we obtain Bloch eigenstates labeled by supercell $\mathbf{k}$ and band index n (occupied) or $\bar{n}$ (unoccupied; we use index with bar on top to label unoccupied states). n labels both the wave function and spin of the eigenstates although there is often an energy degeneracy of 2. These supercell Bloch states $\{\psi_{n\mathbf{k}}\}$, $\{\psi_{\bar{n}\mathbf{k}}\}$ are often delocalized making them hard to visualize and interpret. An alternative representation of electronic wave function and bonding is often needed in the flavor of the LCAO (Ref. 16) or tight-binding[8–10] approach. Ideally, this representation should have features such as exponential localization of the basis orbitals,[53] should be "AO-like," and should retain all the information of the original Bloch eigenstates expressed in plane waves, at least of all

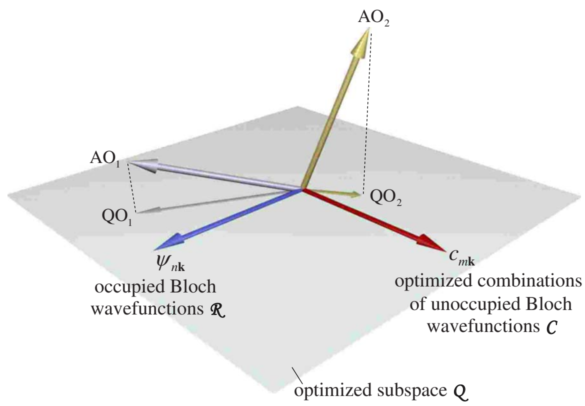

FIG. 1. (Color online) Illustration of QO construction. We seek a reduced optimized subspace $\mathcal{Q}$ spanned by the desired Bloch wave functions $\{\psi_{n\mathbf{k}}\}$ plus a limited number of $\{c_{m\mathbf{k}}\}$ wave functions to be determined, such that the AOs have maximal sum of their projection squares onto the subspace $\mathcal{Q}$. Once this optimized subspace is determined, the QOs, which are the shadows (projections) of the AOs onto the subspace, form a nonorthogonal but complete basis for subspace $\mathcal{Q}$. The QOs can then be used to reconstruct all the desired Bloch wave functions $\{\psi_{n\mathbf{k}}\}$ without loss. This means that in a variational calculation using the QO basis for this particular configuration would achieve the same total energy minimum as the full plane-wave basis. Furthermore since the QOs are maximally similar to the AOs, they inherit most of the AO characters.

the *occupied* Bloch states $\{\psi_{n\mathbf{k}}\}$ so they can be losslessly reconstructed.

QO is a projection-based noniterative approach. It was first implemented by Lu and co-workers,[20–24] called QUAMBO, after the work of Ruedenberg *et al.*[63] on molecular systems. The basic idea is illustrated in Fig. 1. The objective is to seek an optimized subspace $\mathcal{Q}$ containing the occupied $\{\psi_{n\mathbf{k}}\}$ in its entirety plus a limited set of combined unoccupied $\{c_{m\mathbf{k}}\}$ wave functions to be determined, such that the atomic orbitals have maximal sum of their projection squares onto this subspace. The dimension of this "optimized Bloch subspace" is constrained to be that of the minimal (tight-binding) basis, and $\{\psi_{n\mathbf{k}}\}$ and $\{c_{m\mathbf{k}}\}$ form an orthonormal basis for it. But the "shadows" of the AOs projected onto this subspace, which are the QOs, can represent the subspace equally well, forming a nonorthogonal but also complete basis for the subspace. Furthermore, since the QOs are maximally similar to the AOs (under the constraint that they contain $\{\psi_{n\mathbf{k}}\}$ exactly), their localization properties should be "good."

It is important to realize that here we are doing dimension reduction, and the optimized subspace is but a small part of the entire function space, which is infinite dimensional. Since each AO makes one shadow and we use all shadows collected on the plane as nonorthogonal complete basis for the subspace, the total dimension of the subspace has to be $qL_1L_2L_3N$, where q is the average number of AOs per atom. With the minimal-basis scheme, q should be eight for Si and C, and the AOs are $\{s\uparrow, p_x\uparrow, p_y\uparrow, p_z\uparrow; s\downarrow, p_x\downarrow, p_y\downarrow, p_z\downarrow\}$. If we take the smallest supercell admissible for diamond cubic Si, for instance, then $N=2$ and the dimension of the optimized subspace has to be $16L_1L_2L_3$, which is equal to the total number of AOs in the Bv universe. Since we have $L_1L_2L_3$ $\mathbf{k}$ points, this comes down to 16 $\psi_{n\mathbf{k}}, c_{m\mathbf{k}}$'s per $\mathbf{k}$. Because there are eight occupied $\psi_{n\mathbf{k}}$'s at each $\mathbf{k}$ point (doubly degenerate in wave function and energy though), we need to choose eight complementary $c_{m\mathbf{k}}$'s per $\mathbf{k}$. These eight $c_{m\mathbf{k}}$'s will be chosen from the unoccupied $\{\psi_{\bar{n}\mathbf{k}}\}$ subspace, which is infinite dimensional. The whole process can be visualized as rotating the plane around the $\psi_{n\mathbf{k}}$ axis in Fig. 1 and seeking the orientation where the longest shadows fall onto the plane (subspace $\mathcal{Q}$).

Two remarks are in order. First, the label occupied can be replaced by "desired" Bloch wave functions in Fig. 1. While many problems such as fitting TB potentials are mainly concerned with reproducing the occupied bands and the total energy using a minimal basis, problems such as excited-state calculations require more bands to be reproduced. Then, one just needs to generalize the meaning of band index n in Fig. 1 from occupied to desired bands. To be able to do this and still retain AO-like characters, the size of the subspace may necessarily be expanded, for example, from $\{3s,3p\}$ ($q=8$) to $\{3s,3p,4s,3d\}$ ($q=20$) for Si, and then the "minimal basis" is taken to mean the minimal set of AO-like orbitals to reproduce the desired bands, whatever they may be, instead of just the occupied bands. Indeed, a utility of the present QO scheme is to quantitatively guide the user in deciding (a) when to expand, (b) how to expand, and (c) the effectiveness of representing the *desired* part of the electronic structure in AO-like orbitals with pseudoangular momentum quantum numbers. Formally, denote the subspace we want to reproduce at each $\mathbf{k}$ by $\mathcal{R}(\mathbf{k})\equiv\{\psi_{n\mathbf{k}}\}$. Then, the wave functions we do not desire to reproduce at each $\mathbf{k}$ form a complementary subspace $\overline{\mathcal{R}}(\mathbf{k})\equiv\{\psi_{\bar{n}\mathbf{k}}\}$, which is infinite dimensional. We note that $\langle\dim\mathcal{R}(\mathbf{k})\rangle=rN$, but $\dim\mathcal{R}(\mathbf{k})$ or $\mathcal{R}(\mathbf{k})$ generally may not be a continuous function of $\mathbf{k}$. For instance in metals, the Fermi energy ε_F cuts across continuous bands, and the set of occupied bands is not a continuous function of $\mathbf{k}$. We shall call any mathematical or numerical feature caused by a discontinuity in the to-be-reproduced $\mathcal{R}(\mathbf{k})$ as being caused by "type-I" discontinuity.

Second, note that the subspace $\mathcal{Q}$ we seek in Fig. 1 in the Bv universe can be decomposed into smaller subspaces labeled by the Bloch $\mathbf{k}$'s that are mutually orthogonal;

$$\mathcal{Q}=\mathcal{Q}(\mathbf{k}_1)\cup\mathcal{Q}(\mathbf{k}_2)\cup\cdots\cup\mathcal{Q}(\mathbf{k}_{L_1L_2L_3}). \tag{7}$$

Therefore, the length squared of an AO's shadow in $\mathcal{Q}$ is exactly the sum of the projected length squared onto every smaller plane $\mathcal{Q}(\mathbf{k})$. If without any other considerations, the choice of the best rotation can be made independently for each $\mathbf{k}$;

$$\mathcal{Q}(\mathbf{k})=\mathcal{R}(\mathbf{k})\cup\mathcal{C}(\mathbf{k}),\quad \mathcal{C}(\mathbf{k})\subset\overline{\mathcal{R}}(\mathbf{k}), \tag{8}$$

with

$$\dim\mathcal{Q}(\mathbf{k})=qN,$$

$$\dim\mathcal{C}(\mathbf{k})=qN-\dim\mathcal{R}(\mathbf{k}),$$

$$\langle \dim \mathcal{R}(\mathbf{k}) \rangle = rN.$$

Note that all $\mathcal{Q}(\mathbf{k})$ planes are of equal dimension qN. For each AO, one picks up a distinct shadow $|\mathrm{QO}(\mathbf{k})\rangle = \hat{P}_{\mathcal{Q}(\mathbf{k})}|\mathrm{AO}\rangle$ on each $\mathcal{Q}(\mathbf{k})$ plane, then simply adds these $|\mathrm{QO}(\mathbf{k})\rangle$'s together to get the corresponding QO.

$\mathcal{C}(\mathbf{k}) \equiv \{c_{m\mathbf{k}}\}$ is the *choice* of $\psi_{\bar{n}\mathbf{k}}$ combinations,

$$c_{m\mathbf{k}} = \sum_{\bar{n}} C_{m\bar{n}}(\mathbf{k}) \psi_{\bar{n}\mathbf{k}}. \tag{9}$$

Here, $\mathbf{C}(\mathbf{k}) \equiv \{C_{m\bar{n}}(\mathbf{k})\}$ is theoretically a $\dim \mathcal{C}(\mathbf{k}) \times \infty$ matrix. We note that in Eq. (8), only the total function content belonging to subspace $\mathcal{C}(\mathbf{k})$ is important so any unitary transformation $\mathbf{UC}(\mathbf{k})$ is equivalent to the original choice $\mathbf{C}(\mathbf{k})$, where $\mathbf{U}$ is $\dim \mathcal{C}(\mathbf{k}) \times \dim \mathcal{C}(\mathbf{k})$ matrix and $\mathbf{U}^\dagger \mathbf{U} = \mathbf{I}$. Also, even if $\mathcal{R}(\mathbf{k})$ and $\overline{\mathcal{R}}(\mathbf{k})$ are continuous, $\mathcal{C}(\mathbf{k})$ does not have to be continuous in $\mathbf{k}$, in the same way that the minimum eigenvalue of a continuous matrix function $\mathbf{A}(\mathbf{k})$ may not be continuous in $\mathbf{k}$ due to eigenvalue crossings. We call such discontinuity in function content of $\mathcal{Q}(\mathbf{k})$ (not its dimension), which is not caused by discontinuity in $\mathcal{R}(\mathbf{k})$, "type-II" discontinuity. Both type-I and type-II discontinuities could negatively influence the localization properties of QOs, in the same way that the Fourier transform of a step function or functions containing higher-order discontinuities causes algebraic tails in the transformed function.[53] Algebraic decay, however, is not necessarily a show stopper.

In our previous development,[20–24] the "rotation" in Fig. 1 was formulated as a matrix problem with explicit $\{\psi_{\bar{n}\mathbf{k}}\}$ wave functions as the basis. While formally exact, in practice it requires the computation and storage of a large number of $\psi_{\bar{n}\mathbf{k}}$'s, which are then loaded into the postprocessing program to be taken in the inner product with the AOs. The disk space required to store the $\psi_{\bar{n}\mathbf{k}}$'s can run up to tens of gigabytes. Still, one has finite UBTE, which can severely impact the stability of the program. For instance, it was found that when $\{s,p,d\}$ AOs ($q=18$) are used for each Mo atom in bcc Mo, the condition number of the constructed QO overlap matrix is so bad that the numerically calculated TB bands turn singular at some $\mathbf{k}$ points unless exorbitant numbers of unoccupied bands are kept. The bad condition number problem can be somewhat alleviated if $\{s,d\}$ AOs ($q=12$) are used instead of $\{s,p,d\}$.[24] But such solutions are fundamentally unsatisfactory because it is the user's prerogative to decide what is the proper "minimal" basis for the physics one wants to represent and be able to use a richer QO basis if one desires.

It was found recently that a great majority of the bad condition number problems of the previous scheme[20–24] were associated with UBTE. In this work, by resorting to the resolution-of-identity property of the unoccupied subspace $\overline{\mathcal{R}}(\mathbf{k})$, we avoid Eq. (9) representation all together. This not only eliminates the requirement to save a large number of $\psi_{\bar{n}\mathbf{k}}$'s, reducing disk, memory, and memory time requirements by orders of magnitude but also eliminates UBTE as a source of bad condition number. This allows one to construct arbitrarily rich QO basis for bcc Mo such as $\{s,d\}$ and $\{s,p,d\}$ within reasonable computational cost without suffering the UBTE problem (shown in Sec. V D).

Before we move onto the algorithmic details, it is instructive to define qualitatively what we expect at the end. Let us use

$$\langle \mathbf{x}\sigma | A_{Ii} \rangle = A_{Ii}(\mathbf{x}) = A_i(\mathbf{x} - \mathbf{X}_I) \tag{10}$$

to denote the AOs, where I labels the ion and $i \equiv \{\tau l m\}$ is the radial and angular quantum numbers. The AO themselves (e.g., s, p_x, p_y, and p_z) are highly distinct from each other. Indeed, if there were just one isolated atom in a big supercell, AOs of different angular momentum are orthogonal to each other. When there are multiple atoms in the supercell and the metric $\hat{S}$ contains projectors from all ion centers, this orthogonality between AO pseudo wave functions on the same site is no longer rigorously true since two orbitals both centered at $\mathbf{X}_I$ could still overlap in regions covered by other projectors $|\beta_i^I\rangle\langle\beta_j^I|$. (The AO pseudo wave functions are spherical harmonics representing full rotation group, whereas $\hat{S}$ has crystal group symmetry.) Nonetheless, AOs of different angular momentum should be nearly orthogonal and should be highly distinguishable from each other. The same can be said for two AOs, $A_i(\mathbf{x}-\mathbf{X}_I)$ and $A_j(\mathbf{x}-\mathbf{X}_J)$, centered on two different ions. While this is obviously not true if $|\mathbf{X}_J - \mathbf{X}_I| \to 0$, in most systems $\mathbf{X}_I$ and $\mathbf{X}_J$ are well separated by 1 Å or more between nonhydrogen elements.[64] The full rankness of the AO basis in Bv universe guarantees the well behaving (not the same as accuracy) of the numerical LCAO energy bands in the entire Brillouin zone. If this is not the case, in particular if the AO overlap matrix is rank deficient when projected onto some $\mathbf{k}$ point, then the band eigenvalues cannot be obtained in a well-posed manner, and it would manifest as numerical singularities at the $\mathbf{k}$ point in the LCAO energy-band diagram due to bad condition number.

Corresponding to each AO, there is a shadow in the optimized subspace, the QO,

$$\langle \mathbf{x}\sigma | Q_{Ii} \rangle = Q_{Ii}(\mathbf{x}). \tag{11}$$

Even though $Q_{Ii}(\mathbf{x})$ is no longer rigorously spherical harmonic, in the spirit of LCAO $\{Q_{Ii}\}$ should inherit the main characters of $\{A_{Ii}\}$, and therefore should also be *highly distinct*. In other words, when presented with three-dimensional (3D) rendering of the QO orbitals, one should be able to recognize instantly that this is a "p_x-like" QO on atom I, that this is a "$d_{x^2-y^2}$-like" QO on atom J, etc. If this is impossible, the results would not be considered satisfactory even if these orbitals are localized. Also, since the AOs have identical or similar lengths, their shadows on $\mathcal{Q}$ should do too. It is not good news if one shadow is too short, as in the extreme limit of a zero-length shadow if one of the AOs is perpendicular to $\mathcal{Q}$ in Fig. 1. In fact, this needs to hold true for each subplane $\mathcal{Q}(\mathbf{k})$: if for whatever reason, a particular AO is nearly perpendicular to $\mathcal{Q}(\mathbf{k})$, it inevitably spells numerical trouble around that $\mathbf{k}$.

Mathematically the above translates to the following. If $\{\psi_{n\mathbf{k}}, c_{m\mathbf{k}}\}$ are individually normalized (they are orthogonal by construction), then the linear transformation matrix $\mathbf{\Omega_k}$ connecting $\{Q_{Ii,\mathbf{k}}\}$ to $\{\psi_{n\mathbf{k}}, c_{m\mathbf{k}}\}$ must have a reasonable condition number κ, defined here as the ratio of the maximal to minimal eigenvalues of $\mathbf{\Omega_k}^\dagger \mathbf{\Omega_k}$. The following pathology can be identified by a large κ, which is that one QO($\mathbf{k}$) orbital can be expressed as or well approximated by a linear combination of other QO($\mathbf{k}$) orbitals. The QO($\mathbf{k}$)'s are supposedly highly distinct from each other and linearly independent and have reasonable norms [the AO($\mathbf{k}$)'s are, otherwise there will not be well posed, let alone accurate, LCAO bands[16] near that $\mathbf{k}$ if the AOs are literally inserted into real-space DFT codes such as FIREBALL (Ref. 65) or SIESTA (Ref. 43)]. A large condition number would mean this is close to becoming false. This pathology happened in reality, for example, when we attempted to use $\{s,p,d\}$ AOs for each Mo atom ($q=18$) in extracting QOs for bcc Mo with the previous scheme.[24] The bad condition number (due to UBTE) corresponds to nearly linearly dependent QO orbitals when projected onto some $\mathbf{k}$ point, which means that some of the QO($\mathbf{k}$)'s have lost their distinct character or have become very small.

This good condition number criterion provides a quantitative measure of what constitutes a good minimal basis for solid-state systems. While it has not been proved that AO-like minimal basis can be found for all molecular[63] and solid-state systems, experiences with QO show that for the vast majority of systems, a very satisfactory minimal basis can be found (good condition number and good localization). Indeed, by changing the AOs "as little as possible" while maintaining the $\{\psi_{n\mathbf{k}}\}$ band structure, we believe QO fulfills the true spirit of LCAO.[16]

A. Optimized combination subspace

From a plane-wave calculation we obtain the occupied or the to-be-reproduced Bloch eigenstates,

$$\hat{H}|\psi_{n\mathbf{k}}\rangle = \varepsilon_{n\mathbf{k}}\hat{S}|\psi_{n\mathbf{k}}\rangle, \quad n=1,2,\ldots,R_{\mathbf{k}}, \tag{12}$$

as well as some other Bloch eigenstates that belong to the infinite-dimensional subspace $\overline{\mathcal{R}}(\mathbf{k})$;

$$\hat{H}|\psi_{\bar{n}\mathbf{k}}\rangle = \varepsilon_{\bar{n}\mathbf{k}}\hat{S}|\psi_{\bar{n}\mathbf{k}}\rangle. \tag{13}$$

When averaged over the Brillouin zone, we have $\langle R_{\mathbf{k}}\rangle = Nr$, but $R_{\mathbf{k}}$ can vary with $\mathbf{k}$. Different Bloch states are orthogonal to each other. Let us choose normalization

$$\|\psi_{n\mathbf{k}}\|^2 \equiv (\psi_{n\mathbf{k}},\psi_{n\mathbf{k}}) = \langle\psi_{n\mathbf{k}}|\hat{S}|\psi_{n\mathbf{k}}\rangle = 1, \tag{14}$$

$$\|\psi_{\bar{n}\mathbf{k}}\|^2 \equiv (\psi_{\bar{n}\mathbf{k}},\psi_{\bar{n}\mathbf{k}}) = \langle\psi_{\bar{n}\mathbf{k}}|\hat{S}|\psi_{\bar{n}\mathbf{k}}\rangle = 1. \tag{15}$$

We seek an optimized *combination* subspace $\mathcal{C}(\mathbf{k})\subset\overline{\mathcal{R}}(\mathbf{k})$, consisting of *mutually orthonormal states* $\{c_{m\mathbf{k}}\}$, $m=1,2,\ldots,C_{\mathbf{k}}$, to maximize the "sum-over-square" similarity measure $\mathcal{L}$ or the total sum of AO projection squares onto the subspace defined by $\{\psi_{n\mathbf{k}}\}$ and $\{c_{m\mathbf{k}}\}$,

$$\max\mathcal{L} \equiv \max\sum_{Ii}\left\|\left(\sum_{n\mathbf{k}}\hat{P}_{\psi_{n\mathbf{k}}}+\sum_{m\mathbf{k}}\hat{P}_{c_{m\mathbf{k}}}\right)|A_{Ii}\rangle\right\|^2. \tag{16}$$

The $c_{m\mathbf{k}}$ themselves are linear combinations of $\psi_{\bar{n}\mathbf{k}}$. $C_{\mathbf{k}}=qN-R_{\mathbf{k}}$. One may raise two questions. First, why shall we choose $\{c_{m\mathbf{k}}\}$ to be orthonormal? Actually one could choose a set of nonorthonormal states $\{\bar{c}_{m\mathbf{k}}\}$ as long as they span the same subspace as $\{c_{m\mathbf{k}}\}$. Correspondingly, the projection operator, $\hat{P}_{\{c_{m\mathbf{k}}\}}\equiv\Sigma_{m\mathbf{k}}\hat{P}_{c_{m\mathbf{k}}}$, for orthonormal states $\{c_{m\mathbf{k}}\}$ in Eq. (16) should be replaced by the generalized projection operator for nonorthonormal states $\{\bar{c}_{m\mathbf{k}}\}$, which is defined as

$$\hat{P}_{\{\bar{c}_{m\mathbf{k}}\}} \equiv \sum_{ll',\mathbf{k}}|\bar{c}_{l\mathbf{k}}\rangle(\mathbf{O}_{\mathbf{k}}^{-1})_{ll'}\langle\bar{c}_{l\mathbf{k}}|, \tag{17}$$

where $\mathbf{O}_{\mathbf{k}}$ is the overlap matrix between $\bar{c}_{l\mathbf{k}}$'s. Here,

$$(\mathbf{O}_{\mathbf{k}})_{ll'} = \langle\bar{c}_{l\mathbf{k}}|\hat{S}|\bar{c}_{l'\mathbf{k}}\rangle. \tag{18}$$

However, one could easily show that the projection operators for both cases are exactly equivalent to each other,

$$\hat{P}_{\{c_{m\mathbf{k}}\}} = \hat{P}_{\{\bar{c}_{m\mathbf{k}}\}}. \tag{19}$$

This is because both length and direction of the projection of any vector onto a hyperplane (or a subspace) do not depend on how we choose the relative angle and length of basis vectors to represent this hyperplane. Therefore, purely for later convenience we would like to choose a set of orthonormal states $\{c_{m\mathbf{k}}\}$. The second question is: why shall we separate $\Sigma_{n\mathbf{k}}\hat{P}_{\psi_{n\mathbf{k}}}$ from $\Sigma_{m\mathbf{k}}\hat{P}_{c_{m\mathbf{k}}}$? That is because our main goal is to preserve the subspace $\{\psi_{n\mathbf{k}}\}$ and then choose $\{c_{m\mathbf{k}}\}$ to maximize the sum-over-square projection. From the discussion on the first question we can see that once the occupied Bloch subspace $\{\psi_{n\mathbf{k}}\}$ is chosen, the total sum-over-square projection, $\Sigma_{Ii}\|\Sigma_{n\mathbf{k}}\hat{P}_{\psi_{n\mathbf{k}}}|A_{Ii}\rangle\|^2$, of all the AOs onto the occupied Bloch subspace defined in Eq. (16) is fixed. One then only needs to focus on how to choose the hyperplane (or subspace) defined by $\{c_{m\mathbf{k}}\}$ from the unoccupied Bloch subspace $\overline{\mathcal{R}}(\mathbf{k})$ to maximize the total sum of projection squares of AOs on this hyperplane.

The QUAMBO construction of Lu and co-workers[20–24] obtains $\{c_{m\mathbf{k}}\}$ by explicitly rotating a number of $\psi_{\bar{n}\mathbf{k}}$'s. This scheme often suffers from bad condition number problem numerically due to UBTE. It is often worse in metals and confined systems, where the AO's corresponding antibonding Bloch states, especially at Γ point, exist at very high energies. Therefore the original QUAMBO construction often requires obtaining hundreds of $\psi_{\bar{n}\mathbf{k}}$'s at each $\mathbf{k}$ point to include the antibonding states and be able to form the bonding-antibonding closure;[37] otherwise bad condition number would result. A simple example is to consider a hydrogen molecule far away from a metallic substrate with no physical interaction between them. The bonding state $|s_1\rangle+|s_2\rangle$ belongs to the $\psi_{n\mathbf{k}}$'s. The antibonding state $|s_1\rangle-|s_2\rangle$ belongs to the $\psi_{\bar{n}\mathbf{k}}$'s, but it could be higher in energy than many metallic $\psi_{\bar{n}\mathbf{k}}$ states and may not be selected as basis for rotation. $|s_1\rangle$ is AO_1 and $|s_2\rangle$ is AO_2. We can see from Fig. 1 that if $|s_1\rangle-|s_2\rangle$ is not included in the explicit $\psi_{\bar{n}\mathbf{k}}$ basis in which the plane could rotate in, then both AO_1 and AO_2 will always have the same "shadow" ($|s_1\rangle+|s_2\rangle$) on the plane no matter how the plane rotates. This then results in bad condition number. UBTE-caused closure failure can also happen in MLWF construction. However, in the case of MLWF, one may get bond-centered instead of atom-centered MLWFs after localization extremization. But QUAMBO similarity maximization would just fail numerically.

A closer inspection of Fig. 1 reveals that a better method may be found. First, we note that the combination subspace $\mathcal{C}(\mathbf{k})$ we seek is a subspace of $\overline{\mathcal{R}}(\mathbf{k})$. $\overline{\mathcal{R}}(\mathbf{k})$ itself is infinite dimensional, but much of $\overline{\mathcal{R}}(\mathbf{k})$ has *no overlap* with the designated AOs since there is only a *finite* number of these AOs. Those parts of $\overline{\mathcal{R}}(\mathbf{k})$ with no overlap to the AOs would *not improve* similarity with the designated AOs even if included. And thus they can be *excluded* from the basis for rotation. In other words, $\overline{\mathcal{R}}(\mathbf{k})$ can be further decomposed as

$$\overline{\mathcal{R}}(\mathbf{k}) = \overline{\mathcal{A}}(\mathbf{k}) \cup \overline{\mathcal{N}}(\mathbf{k}), \tag{20}$$

where $\overline{\mathcal{A}}(\mathbf{k})$ has overlap with the AOs but $\overline{\mathcal{N}}(\mathbf{k})$ has none. The assertion here is that choosing $\mathcal{C}(\mathbf{k})\subset\overline{\mathcal{A}}(\mathbf{k})\subset\overline{\mathcal{R}}(\mathbf{k})$ will just give identical result (same similarity measure and shadow wave functions) as $\mathcal{C}(\mathbf{k})\subset\overline{\mathcal{R}}(\mathbf{k})$. Because $\overline{\mathcal{A}}(\mathbf{k})$ is supposedly finite dimensional [in fact $\dim\overline{\mathcal{A}}(\mathbf{k})=qN$], one just needs to find basis functions for $\overline{\mathcal{A}}(\mathbf{k})$ and perform rotation in this finite subspace rather than finding the infinite $\{\psi_{\bar{n}\mathbf{k}}\}$ basis functions for $\overline{\mathcal{R}}(\mathbf{k})$ and rotating in $\overline{\mathcal{R}}(\mathbf{k})$.

To proceed, let us first define

$$A_{Ii,\mathbf{k}}(\mathbf{x}) \equiv \sum_{L=1}^{L_1L_2L_3} A_{Ii}(\mathbf{x}-\mathbf{X}_L)e^{i\mathbf{k}\cdot\mathbf{X}_L}, \tag{21}$$

which is a linear superposition of translated AOs in the Bv universe with Bloch phase factors. $\mathbf{X}_L=l_1\mathbf{a}_1+l_2\mathbf{a}_2+l_3\mathbf{a}_3$ is an integer combination of supercell edge vectors. $A_{Ii,\mathbf{k}}(\mathbf{x})$ is clearly a Bloch state,

$$A_{Ii,\mathbf{k}}(\mathbf{x}-\mathbf{a}) = A_{Ii,\mathbf{k}}(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{a}}, \tag{22}$$

and is just the projection (un-normalized) of $|A_{Ii}\rangle$ onto Bloch subspace $\mathcal{B}(\mathbf{k})\equiv\mathcal{R}(\mathbf{k})\cup\overline{\mathcal{R}}(\mathbf{k})$. Because of this, $|A_{Ii,\mathbf{k}}\rangle$ can be further decomposed into a component that belongs to $\mathcal{R}(\mathbf{k})$ and a component that belongs to $\overline{\mathcal{R}}(\mathbf{k})$;

$$|A_{Ii,\mathbf{k}}\rangle = |A^{\parallel}_{Ii,\mathbf{k}}\rangle + |A^{\perp}_{Ii,\mathbf{k}}\rangle, \tag{23}$$

where

$$|A^{\parallel}_{Ii,\mathbf{k}}\rangle \equiv \sum_n \hat{P}_{\psi_{n\mathbf{k}}}|A_{Ii,\mathbf{k}}\rangle \in \mathcal{R}(\mathbf{k}), \tag{24}$$

and

$$|A^{\perp}_{Ii,\mathbf{k}}\rangle = |A_{Ii,\mathbf{k}}\rangle - \sum_n \hat{P}_{\psi_{n\mathbf{k}}}|A_{Ii,\mathbf{k}}\rangle \in \overline{\mathcal{R}}(\mathbf{k}). \tag{25}$$

$|A^{\parallel}_{Ii,\mathbf{k}}\rangle$ and $|A^{\perp}_{Ii,\mathbf{k}}\rangle$ can be calculated straightforwardly in plane-wave basis according to Eqs. (21), (24), and (25) without knowing the $\{\psi_{\bar{n}\mathbf{k}}\}$'s explicitly. Similarly we can decompose QOs which are the projections of AOs into parallel and perpendicular part,

$$|Q_{Ii,\mathbf{k}}\rangle = |Q^{\parallel}_{Ii,\mathbf{k}}\rangle + |Q^{\perp}_{Ii,\mathbf{k}}\rangle, \tag{26}$$

$$|Q^{\parallel}_{Ii,\mathbf{k}}\rangle = \sum_n \hat{P}_{\psi_{n\mathbf{k}}}|A_{Ii,\mathbf{k}}\rangle, \tag{27}$$

$$|Q^{\perp}_{Ii,\mathbf{k}}\rangle = \sum_m \hat{P}_{c_{m\mathbf{k}}}|A_{Ii,\mathbf{k}}\rangle. \tag{28}$$

It is clearly $|Q^{\parallel}_{Ii,\mathbf{k}}\rangle=|A^{\parallel}_{Ii,\mathbf{k}}\rangle$. Therefore, the sum-over-square similarity measure $\mathcal{L}$ which we want to maximize in Eq. (16) can be simply rewritten as

$$\mathcal{L} = \sum_{Ii}\left(\left\|\sum_{\mathbf{k}}|Q^{\parallel}_{Ii,\mathbf{k}}\rangle\right\|^2 + \left\|\sum_{\mathbf{k}}|Q^{\perp}_{Ii,\mathbf{k}}\rangle\right\|^2\right). \tag{29}$$

This can be seen clearly in Fig. 1 from geometrical view, where the $\mathcal{L}$ measure is the sum of AO projection squares or the sum of length squares of QOs on the subspace $\mathcal{Q}$ formed by occupied Bloch subspace $\mathcal{R}$ and the combination subspace $\mathcal{C}$. QOs as the shadow of AOs have different lengths and directions if different $\mathcal{C}$ is chosen. Therefore, one is trying to "hold" (preserve) subspace $\mathcal{R}$ and "rotate" (search) $\mathcal{C}$ in unoccupied Bloch subspace $\overline{\mathcal{R}}$ to maximize the sum of QO length squares.

Furthermore, any Bloch state $|b_{\mathbf{k}}\rangle\in\mathcal{B}(\mathbf{k})$ orthogonal to $|A^{\parallel}_{Ii,\mathbf{k}}\rangle$ and $|A^{\perp}_{Ii,\mathbf{k}}\rangle$ must be orthogonal to $|A_{Ii,\mathbf{k}}\rangle$ and vice versa. Including such $|b_{\mathbf{k}}\rangle$ in the basis for $\mathcal{Q}(\mathbf{k})$ optimization in Fig. 1 will not improve similarity with this AO and thus can be excluded. So we only need to optimize $\mathcal{Q}(\mathbf{k})$ within $\{A^{\parallel}_{Ii,\mathbf{k}}\}$ and $\{A^{\perp}_{Ii,\mathbf{k}}\}$. Because $\{A^{\parallel}_{Ii,\mathbf{k}}\}\subset\mathcal{R}(\mathbf{k})$ and $\mathcal{R}(\mathbf{k})$ will anyhow be included in $\mathcal{Q}(\mathbf{k})=\mathcal{R}(\mathbf{k})\cup\mathcal{C}(\mathbf{k})$, it is thus only necessary to optimize $\mathcal{C}(\mathbf{k})$ within the subspace $\{A^{\perp}_{Ii,\mathbf{k}}\}$, which we identify to be $\overline{\mathcal{A}}(\mathbf{k})$. Clearly, $\dim\overline{\mathcal{A}}(\mathbf{k})=qN$. All we need to do then is to find a $C_{\mathbf{k}}=qN-R_{\mathbf{k}}$ dimensional optimized combination subspace $\mathcal{C}_{\mathbf{k}}\subset\overline{\mathcal{A}}(\mathbf{k})$.

The QO approach proposed here is similar to the projected atomic orbitals (PAO) approach of Sæbø and Pulay[66–68] for molecular systems. By combining Eqs. (23)–(25) and (29) we have

$$\begin{aligned}\max\mathcal{L} &= \sum_{Ii}\left\|\sum_{\mathbf{k}}|Q^{\parallel}_{Ii,\mathbf{k}}\rangle\right\|^2 + \max\sum_{Ii}\left\|\sum_{m\mathbf{k}}\hat{P}_{c_{m\mathbf{k}}}|A_{Ii,\mathbf{k}}\rangle\right\|^2\\ &= \sum_{Ii}\left\|\sum_{\mathbf{k}}|Q^{\parallel}_{Ii,\mathbf{k}}\rangle\right\|^2 + \max\sum_{Ii}\left\|\sum_{m\mathbf{k}}\hat{P}_{c_{m\mathbf{k}}}|A^{\perp}_{Ii,\mathbf{k}}\rangle\right\|^2.\end{aligned} \tag{30}$$

In the above equation we have used the fact that $\hat{P}_{c_{m\mathbf{k}}}|A^{\parallel}_{Ii,\mathbf{k}}\rangle=0$ due to the orthogonality between $\psi_{n\mathbf{k}}$ and $c_{m\mathbf{k}}$. Moreover, as we have argued above, optimized combination subspace $\mathcal{C}_{\mathbf{k}}$ formed by $\{c_{m\mathbf{k}}\}$ is a subset of $\overline{\mathcal{A}}(\mathbf{k})$ formed by $\{A^{\perp}_{Ii,\mathbf{k}}\}$. This means that we are seeking a transformation matrix $\overline{\mathbf{V}}_k$ such that

$$|c_{m\mathbf{k}}\rangle = \sum_{Ii}(\overline{\mathbf{V}}_{\mathbf{k}})_{Ii,m}|A^{\perp}_{Ii,\mathbf{k}}\rangle. \tag{31}$$

As we have mentioned earlier, due to Eq. (19) we can force ourselves to search a set of orthonormal states for the sake of convenience. Thus combined with Eq. (31), it immediately leads to

$$\langle c_{m\mathbf{k}}|\hat{S}|c_{m'\mathbf{k}}\rangle = \sum_{Ii,Jj}(\overline{\mathbf{V}}_{\mathbf{k}})^*_{Ii,m}(\overline{\mathbf{V}}_{\mathbf{k}})_{Jj,m'}\langle A^{\perp}_{Ii,\mathbf{k}}|\hat{S}|A^{\perp}_{Jj,\mathbf{k}}\rangle = \delta_{mm'}. \tag{32}$$

We denote the overlap matrix between $\{A^{\perp}_{Ii,\mathbf{k}}\}$ as $\mathbf{W}_{\mathbf{k}}$,

$$(\mathbf{W}_{\mathbf{k}})_{Ii,Jj} \equiv (A^{\perp}_{Ii,\mathbf{k}}, A^{\perp}_{Jj,\mathbf{k}}) = \langle A^{\perp}_{Ii,\mathbf{k}}|\hat{S}|A^{\perp}_{Jj,\mathbf{k}}\rangle. \tag{33}$$

Then the orthonormal condition of $\{c_{m\mathbf{k}}\}$ in Eq. (32) basically states that

$$\overline{\mathbf{V}}^{\dagger}_{\mathbf{k}}\mathbf{W}_{\mathbf{k}}\overline{\mathbf{V}}_{\mathbf{k}} = \mathbf{I}_{C_{\mathbf{k}}\times C_{\mathbf{k}}}. \tag{34}$$

We notice that the overlap matrix $\mathbf{W}_{\mathbf{k}}$ is a Gramian matrix which is positive semidefinite as we show in the Appendix. Meanwhile, it can be diagonalized by a unitary matrix $\mathbf{V}_{\mathbf{k}}$ such that $\mathbf{W}_{\mathbf{k}}=\mathbf{V}_{\mathbf{k}}\mathbf{Y}_{\mathbf{k}}\mathbf{V}^{\dagger}_{\mathbf{k}}$, where $\mathbf{V}_{\mathbf{k}}\mathbf{V}^{\dagger}_{\mathbf{k}}=\mathbf{I}_{qN\times qN}$, and the diagonal matrix $\mathbf{Y}_{\mathbf{k}}$ contains all the non-negative real eigenvalues. Therefore, Eq. (32) suggests $\overline{\mathbf{V}}^{\dagger}_{\mathbf{k}}\mathbf{V}_{\mathbf{k}}\mathbf{Y}_{\mathbf{k}}\mathbf{V}^{\dagger}_{\mathbf{k}}\overline{\mathbf{V}}_{\mathbf{k}}=\mathbf{I}_{C_{\mathbf{k}}\times C_{\mathbf{k}}}$. The solution for $\overline{\mathbf{V}}_{\mathbf{k}}$ is

$$(\overline{\mathbf{V}}_{\mathbf{k}})_{Ii,m} = (\mathbf{V}_{\mathbf{k}})_{Ii,m}(\mathbf{Y}_{\mathbf{k}})^{-1/2}_{mm}, \tag{35}$$

where $Ii=1,2,\ldots,qN$. Obviously any $C_{\mathbf{k}}$ positive eigenvalues $(\mathbf{Y}_{\mathbf{k}})_{mm}$ of $\mathbf{W}_{\mathbf{k}}$ matrix (as we have mentioned above, all the eigenvalues of $\mathbf{W}_{\mathbf{k}}$ matrix are non-negative real values) and their corresponding eigenvectors $(\mathbf{V}_{\mathbf{k}})_{Ii,m}$ will give a proper $\overline{\mathbf{V}}_{\mathbf{k}}$ matrix which satisfies the orthonormal condition for $\{c_{m\mathbf{k}}\}$ in Eq. (32). We then come back to the problem of maximizing the sum of projection squares $\mathcal{L}$ by choosing the "best" set of $(\mathbf{Y}_{\mathbf{k}})_{mm}$ and their eigenvectors. From Eq. (30) we only need to maximize the sum of projection squares on the subspace $\{c_{m\mathbf{k}}\}$ since the sum of projection squares on $\{\psi_{n\mathbf{k}}\}$ is fixed. Therefore, using Eqs. (30)–(33) and (35) we have

$$\begin{aligned}
&\max \sum_{Ii}\left\|\sum_{m\mathbf{k}}\hat{P}_{c_{m\mathbf{k}}}|A^{\perp}_{Ii,\mathbf{k}}\rangle\right\|^2 \\
&= \max \sum_{Ii}\sum_{m\mathbf{k}}\langle A^{\perp}_{Ii,\mathbf{k}}|\hat{S}|c_{m\mathbf{k}}\rangle\langle c_{m\mathbf{k}}|\hat{S}|A^{\perp}_{Ii,\mathbf{k}}\rangle \\
&= \max \sum_{Ii,m\mathbf{k}}\sum_{Jj,J'j'}(\mathbf{W}_{\mathbf{k}})_{Ii,Jj}(\overline{\mathbf{V}}_{\mathbf{k}})_{Jj,m}(\overline{\mathbf{V}}_{\mathbf{k}})^{\dagger}_{J'j',m}(\mathbf{W}_{\mathbf{k}})_{J'j',Ii} \\
&= \max \sum_{\mathbf{k}}\mathrm{Tr}(\mathbf{W}_{\mathbf{k}}\overline{\mathbf{V}}_{\mathbf{k}}\overline{\mathbf{V}}^{\dagger}_{\mathbf{k}}\mathbf{W}_{\mathbf{k}}),
\end{aligned} \tag{36}$$

where "Tr" means the trace. Thus, Eq. (16) for maximizing the total sum of projection squares is rewritten in the following simple form:

$$\max \mathcal{L} = \sum_{Ii}\left\|\sum_{\mathbf{k}}|Q^{\parallel}_{Ii,\mathbf{k}}\rangle\right\|^2 + \max\sum_{\mathbf{k}}\mathrm{Tr}(\mathbf{W}_{\mathbf{k}}\overline{\mathbf{V}}_{\mathbf{k}}\overline{\mathbf{V}}^{\dagger}_{\mathbf{k}}\mathbf{W}_{\mathbf{k}}) \tag{37}$$

$$= \sum_{Ii}\left\|\sum_{\mathbf{k}}|Q^{\parallel}_{Ii,\mathbf{k}}\rangle\right\|^2 + \max\sum_{m\mathbf{k}}(\mathbf{Y}_{\mathbf{k}})_{mm}, \tag{38}$$

where $\Sigma_m(\mathbf{Y}_{\mathbf{k}})_{mm}$ basically sums all the $C_{\mathbf{k}}$ eigenvalues arbitrarily chosen from the total qN non-negative real eigenvalues of $\mathbf{W}_{\mathbf{k}}$ matrix. Therefore, the equation above suggests that by choosing the largest $C_{\mathbf{k}}$ eigenvalues and their corresponding eigenvectors we will maximize the total sum of projection squares $\mathcal{L}$. Consequently $\{c_{m\mathbf{k}}\}$ are obtained from Eqs. (31) and (35).

To use $\{\psi_{n\mathbf{k}}\}$ and $\{c_{m\mathbf{k}}\}$ in band-structure and Fermi-surface calculations we have to construct Hamiltonian matrix $\boldsymbol{\epsilon}_{\mathbf{k}}$ between any two functions in $\{\psi_{n\mathbf{k}}, c_{m\mathbf{k}}\}$. Since $\{\psi_{n\mathbf{k}}\}$ are eigenfunctions of the Kohn-Sham Hamiltonian,

$$(\boldsymbol{\epsilon}_{\mathbf{k}})_{n,n'} \equiv \langle\psi_{n\mathbf{k}}|\hat{H}|\psi_{n'\mathbf{k}}\rangle = \varepsilon_{n\mathbf{k}}\delta_{nn'}, \tag{39}$$

with $n,n'=1,2,\ldots,R_{\mathbf{k}}$. It is also obvious that the matrix element of $\hat{H}$ between $\psi_{n\mathbf{k}}$ and $c_{m\mathbf{k}}$ is always zero since they are from two different Bloch eigensubspaces,

$$(\boldsymbol{\epsilon}_{\mathbf{k}})_{n,m+R_{\mathbf{k}}} \equiv \langle\psi_{n\mathbf{k}}|\hat{H}|c_{m\mathbf{k}}\rangle = 0,$$

$$(\boldsymbol{\epsilon}_{\mathbf{k}})_{m+R_{\mathbf{k}},n} \equiv \langle c_{m\mathbf{k}}|\hat{H}|\psi_{n\mathbf{k}}\rangle = 0, \tag{40}$$

where $n=1,2,\ldots,R_{\mathbf{k}}$ and $m=1,2,\ldots,C_{\mathbf{k}}$. Although $|c_{m\mathbf{k}}\rangle$ comes from diagonalization of $\mathbf{W}_{\mathbf{k}}$, it is *not* an eigenfunction of the Kohn-Sham Hamiltonian. Thus the matrix elements of $\boldsymbol{\epsilon}_{\mathbf{k}}$ between two different $c_{m\mathbf{k}}$'s at the same $\mathbf{k}$ may not be zero and we have to use the Kohn-Sham Hamiltonian $\hat{H}$ to calculate this part of $\boldsymbol{\epsilon}_{\mathbf{k}}$ explicitly,

$$(\boldsymbol{\epsilon}_{\mathbf{k}})_{m+R_{\mathbf{k}},m'+R_{\mathbf{k}}} \equiv \langle c_{m\mathbf{k}}|\hat{H}|c_{m'\mathbf{k}}\rangle, \tag{41}$$

with $m,m'=1,2,\ldots,C_{\mathbf{k}}$. In the end, the matrix $\boldsymbol{\epsilon}_{\mathbf{k}}$ consists of a diagonal submatrix for the occupied Bloch subspace $\mathcal{R}(\mathbf{k})$, a nondiagonal square submatrix for the optimized combination subspace $\mathcal{C}(\mathbf{k})$, and two rectangular zero matrices between $\mathcal{R}(\mathbf{k})$ and $\mathcal{C}(\mathbf{k})$.

We can now merge the basis functions for $\mathcal{R}(\mathbf{k})$ and $\mathcal{C}(\mathbf{k})$,

$$\{\phi_{n\mathbf{k}}\} = \{\psi_{n\mathbf{k}}\} \cup \{c_{m\mathbf{k}}\}, \tag{42}$$

where $\{\phi_{n\mathbf{k}}\}$ then constitutes a qN-dimensional basis for $\mathcal{Q}(\mathbf{k})$, which is orthonormal,

$$\langle\phi_{n\mathbf{k}}|\hat{S}|\phi_{n'\mathbf{k}}\rangle = \delta_{nn'}, \quad n,n' = 1,\ldots,qN, \tag{43}$$

in the sense of the Bv universe [Eq. (4)]. According to Fig. 1, the QO is just

$$|Q_{Ii}\rangle = \sum_{n\mathbf{k}}\hat{P}_{\phi_{n\mathbf{k}}}|A_{Ii}\rangle = \sum_{n\mathbf{k}}(\boldsymbol{\Omega}_{\mathbf{k}})_{n,Ii}|\phi_{n\mathbf{k}}\rangle, \tag{44}$$

where $n=1,\ldots,qN$, $\mathbf{k}$ runs over $1,\ldots,L_1L_2L_3$ Monkhorst-Pack grid, and

$$(\boldsymbol{\Omega}_{\mathbf{k}})_{n,Ii} \equiv \langle\phi_{n\mathbf{k}}|\hat{S}|A_{Ii}\rangle \tag{45}$$

is a $qN\times qN$ matrix. Actually one could further rescale $|Q_{Ii}\rangle$ by a constant such that Q_{Ii} satisfies the normalization condition, $\langle Q_{Ii}|\hat{S}|Q_{Ii}\rangle=1$, while this simple rescaling procedure will not affect the Mulliken charge and bond order analysis. Furthermore, one could perform Löwdin transformation to obtain a set of orthonormal QOs. Both transformations will not affect the band-structure calculations.

QO procedures (21)–(45) maximize the overall similarity measure [Eq. (16)] and in fact give identical results as the original QUAMBO scheme[20–24] in the infinite band limit. The proof is given in the Appendix.

B. Choosing reproduced subspace $\mathcal{R}(\mathbf{k})$

QO procedures (21)–(45) rely on a preselection of to-be-reproduced Bloch subspace $\mathcal{R}(\mathbf{k})$. It is necessary to give the user this freedom because it is up to the user to define which parts of the electronic structure are important and need to be preserved. For properties related to ground-state total energies, obviously the occupied bands are important. Therefore a quasiparticle energy-based selection criterion can be adopted, where all eigenstates whose energies are below a threshold energy ε_{th} several eV above the Fermi energy ε_F are included in $\mathcal{R}(\mathbf{k})$. On the other hand, a particular energy window of the unoccupied bands may be important for optical absorption at certain frequency or electronic transport at a certain bias voltage,[19,45–47] and they may need to be included in $\mathcal{R}(\mathbf{k})$. One may even choose to include in $\mathcal{R}(\mathbf{k})$ a certain continuous band at all **k** points irrespective of its eigenenergies if that band is deemed important for transport or chemical properties.

In the present QO scheme, say with an energy-based selection criterion, the distinction between selected and unselected is "sharp." That is, a Bloch eigenfunction is either chosen (1) or not chosen (0) to be in $\mathcal{R}(\mathbf{k})$. There is no grayscale in between, and depending on 1 or 0 the eigenfunction will be treated differently in the algorithm. A certain $\psi_{n\mathbf{k}}$ may be in $\mathcal{R}(\mathbf{k})$, but with just infinitesimal change in **k** and wave function character, and could be excluded in $\mathcal{R}(\mathbf{k}+d\mathbf{k})$. Such sharp type-I discontinuities in the Brillouin zone always lead to "long-ranged" interactions in real space (meaning algebraic instead of exponential decay with distance[53]). For example, in metals sharp type-I discontinuities in the occupation number at low temperature give rise to physical effects such as the Kohn anomaly (long-ranged interatomic force constants leading to weak singularities in the phonon-dispersion relation)[69] that can be measured by neutron scattering.[70]

Therefore type-I discontinuity is not just a numerical and/or algorithmic problem specific to QO but is also a physical and quite inherent issue in metals. Numerical techniques such as Fermi-Dirac smearing or Methfessel-Paxton smearing[71] with artificially chosen smearing widths have been used to regularize type-I discontinuity in total-energy calculations. In fact, without such artificial smearing it is quite challenging to obtain well-behaving (smooth) total energy and forces numerically in traditional DFT calculations. One thus wonders whether a similar approach can be applied to $\mathcal{R}(\mathbf{k})$ selection. We think this can be done by assigning weighting function $f(\varepsilon_{n\mathbf{k}})$ to Eqs. (24) and (25) projections that smoothly varies from 1 to 0 around ε_{th}. In such case, $\{A^{\parallel}_{Ii,\mathbf{k}}\}$ and $\{A^{\perp}_{Ii,\mathbf{k}}\}$ will no longer be rigorously orthogonal, and a weighted joint $2qN\times 2qN$ overlap matrix will be set up and diagonalized. This "grayscale QO" method can be shown to be identical to the present "sharp QO" method in the limit when $f(\varepsilon_{n\mathbf{k}})$ is a sharp step function but remove type-I discontinuities when $f(\varepsilon_{n\mathbf{k}})$ is not sharp. We will postpone full evaluation of this grayscale QO method to a later paper.

C. Choosing atomic orbitals

Another freedom the user has is choosing the atomic orbitals $A_{Ii}(\mathbf{x})$. While it is operationally straightforward to just use the pseudoatomic orbitals $\overline{\mathrm{A}}_{Ii}(\mathbf{x})$ of an isolated atom that come with the pseudopotential, we find that the pseudoatomic orbitals of some elements have very long tails, extending to 10 Å away from the ion. Then to use these long-tailed orbitals as similarity objects in Fig. 1 is not very good for localization. Also, it is not fundamentally obvious that the eigenorbitals of isolated atoms with unfilled electronic shells maximally reflect the electronic structure of bonded systems with filled shells. Although Slater and Koster[16] named their method linear combinations of atomic orbitals, which gave rise to the empirical tight-binding method, the term "atomic orbitals" may be taken with a grain of salt. The Slater-Koster paper[16] tabulated the angular interactions, implying that the atomic orbitals have Y_{lm} angular dependencies, but the radial functions were not specified.

Indeed, Slater[64] himself later defined the so-called empirial *atomic* radius R for many elements by regressing to an experimental database of 1200 bond lengths in crystals and molecules and demanding that the bond length $(A-B)\approx R(A)+R(B)$ between elements A and B. He found that these 1200 bond lengths can be regressed to an average error of 0.12 Å using empirial atomic radii. So the concept of Slater[64] of atomic radius and atomic orbital may be tied more to natural bonding environments than isolated atoms. It is also known that if one insists on using pseudoatomic orbitals $\overline{\mathrm{A}}_{Ii}(\mathbf{x})$ as the literal minimal basis in a local-basis DFT calculation,[43,65] one gets accuracy far worse than what empirical tight-binding methods can do nowadays *without* explicit statement of the radial functions.

The considerations above suggest a heuristic approach for choosing the radial part of the AOs. One simple strategy is to rescale the pseudoatomic radial function by multiplying an exponentially decaying function,

$$A_{Ii}(\mathbf{x}) = \xi_i \overline{\mathrm{A}}_{Ii}(\mathbf{x}) e^{-\eta_i|\mathbf{x}|}, \tag{46}$$

where η_i is a positive real number and ξ_i is the normalization factor to make $\langle A_{Ii}|\hat{S}|A_{Ii}\rangle=1$. The rationale behind Eq. (46) squeezing could be a screening effect[8] since the pseudoatomic orbitals now need to penetrate neighboring electron clouds, and a more localized AO may be a better descriptor of the electronic structure and chemistry.

We find that Eq. (46) indeed improves localization of QOs and subsequently that of the TB Hamiltonian. While η_i needs to be empirially chosen or even optimized systematically, we believe that this is not work in vain but is actually physichemically illuminating. In fact, it may eventually lead to generalization of the empirial atomic radius concept of Slater[64] to construction of empirial atomic orbitals. We envision a database of thousands of bonded molecules and solids, and one is constrained to choose just one η_i value for each element that will give the best overall QO description (localization and similarity) for a multitude of bonding environments. The hypothesis is that empirical atomic orbitals indeed exist for each element that robustly describe electronic structure in a wide range of molecular and solid bonding environments via the QO approach.

IV. *AB INITIO* TIGHT-BINDING ANALYSIS

Ab initio tight-binding approach differs from empirial tight-binding approach in explicitly specifying the minimal-basis functions used. Once the QOs are obtained via Eqs. (21)–(45), we can evaluate—and later parametrize—the tight-binding Hamiltonian $\mathbf{H}$ and overlap matrix $\mathbf{O}$, which are small matrices with real-space indices in contrast to the Kohn-Sham Hamiltonian in plane-wave basis that nonetheless reproduce *all* electronic structure information in $\mathcal{R}(\mathbf{k})$. In fact, if $\mathcal{R}(\mathbf{k})$ includes the occupied bands, the QOs can be used as literal basis to perform total-energy calculation in real-space DFT codes such as FIREBALL (Ref. 65) or SIESTA,[43] which will yield the same total-energy variational minimum as using full plane-wave basis.

Once the TB $\mathbf{H}$ and $\mathbf{O}$ matrices are constructed, they can be easily applied to calculate band structure, density of states, QO-projected band structure and density of states, the high-resolution Fermi surface, and Mulliken charge and bond order that satisfy exact sum rules. These calculations are much more efficient than direct plane-wave DFT calculations due to the small size of TB matrices and furthermore will carry valuable real-space information.

A. Tight-binding representation

Under QO basis, TB Hamiltonian $H_{Ii,Jj}(\mathbf{X}_L)$ between Q_{Ii}^0 and Q_{Jj}^L in two supercells is defined as

$$H_{Ii,Jj}(\mathbf{X}_L) \equiv \langle Q_{Ii}^0|\hat{H}|Q_{Jj}^L\rangle,$$

where $\mathbf{X}_L=l_1\mathbf{a}_1+l_2\mathbf{a}_2+l_3\mathbf{a}_3$ is an integer combination of supercell edge vectors. However we do not need to evaluate the above matrix element explicitly since we can obtain the Hamiltonian submatrices [Eqs. (39)–(41)] between optimized Bloch states $\{\phi_{m\mathbf{k}}\}$ and transformation matrix $\mathbf{\Omega}_\mathbf{k}$ from QO to $\{\phi_{m\mathbf{k}}\}$. From Eq. (44), we have the expression of QO of atom J in supercell $\mathbf{X}_L$,

$$\begin{aligned} Q_{Jj}^L(\mathbf{x}) &= Q_{Jj}(\mathbf{x}-\mathbf{X}_L) \\ &= \sum_{m\mathbf{k}} (\mathbf{\Omega}_\mathbf{k})_{m,Jj}\phi_{m\mathbf{k}}(\mathbf{x}-\mathbf{X}_L) \\ &= \sum_{m\mathbf{k}} (\mathbf{\Omega}_\mathbf{k})_{m,Jj}e^{-i\mathbf{k}\cdot\mathbf{X}_L}\phi_{m\mathbf{k}}(\mathbf{x}), \end{aligned}$$

then the above real-space TB Hamiltonian $H_{Ii,Jj}(\mathbf{X}_L)$ is

$$H_{Ii,Jj}(\mathbf{X}_L) = \sum_{m,m',\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{X}_L}(\mathbf{\Omega}_\mathbf{k})^*_{m,Ii}(\boldsymbol{\epsilon}_\mathbf{k})_{mm'}(\mathbf{\Omega}_\mathbf{k})_{m',Jj}.$$

Following the same procedure we can easily calculate the real-space TB overlap matrix $O_{Ii,Jj}(\mathbf{X}_L)$,

$$O_{Ii,Jj}(\mathbf{X}_L) = \langle Q_{Ii}^0|\hat{S}|Q_{Jj}^L\rangle = \sum_{m\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{X}_L}(\mathbf{\Omega}_\mathbf{k})^*_{m,Ii}(\mathbf{\Omega}_\mathbf{k})_{m,Jj}.$$

Clearly $H_{Ii,Jj}(\mathbf{X}_L)$ and $O_{Ii,Jj}(\mathbf{X}_L)$ should decay to zero as $\mathbf{X}_L \to \infty$ and have similar localization property as the QOs. Using them, we can efficiently compute the eigenvalues at an arbitrary $\mathbf{k}$ point (not necessarily one of the $L_1L_2L_3$ $\mathbf{k}$ points) by forming

$$H_{Ii,Jj}(\mathbf{k}) = \sum_{|\mathbf{X}_L|<R_\text{cut}} e^{i\mathbf{k}\cdot\mathbf{X}_L}H_{Ii,Jj}(\mathbf{X}_L),$$

$$O_{Ii,Jj}(\mathbf{k}) = \sum_{|\mathbf{X}_L|<R_\text{cut}} e^{i\mathbf{k}\cdot\mathbf{X}_L}O_{Ii,Jj}(\mathbf{X}_L), \tag{47}$$

where $\mathbf{X}_L$ runs over shells of neighboring supercells with significant $H_{Ii,Jj}(\mathbf{X}_L)$ and $O_{Ii,Jj}(\mathbf{X}_L)$. Typically we determine a radial cutoff distance R_cut and sum only those L's with $|\mathbf{X}_L| < R_\text{cut}$ in Eq. (47). Then, by solving the generalized eigenvalue matrix problem

$$\mathbf{H}(\mathbf{k})\Pi(\mathbf{k}) = \mathbf{O}(\mathbf{k})\Pi(\mathbf{k})\mathbf{E}(\mathbf{k}), \tag{48}$$

we obtain total $m=1,\ldots,qN$ eigenenergies $e_{m\mathbf{k}}$ from the diagonal matrix $\mathbf{E}(\mathbf{k})$ at each $\mathbf{k}$ point, with $qN=R_\mathbf{k}+C_\mathbf{k}$. It is expected that all the $R_\mathbf{k}$ energies lower than ε_th are the same as the eigenenergies from DFT calculation: $e_{n\mathbf{k}}=\varepsilon_{n\mathbf{k}}$, with $n=1,\ldots,R_\mathbf{k}$. The remaining $C_\mathbf{k}$ eigenenergies belong to the optimized combination Bloch states $\{c_{m\mathbf{k}}\}$. The physical interpretation of $e_{m\mathbf{k}}$, $m=1,\ldots,qN$ is that it is the variational minimum of Rayleigh quotient,

$$e_{m\mathbf{k}} = \min_{\pi_{m\mathbf{k}}} \frac{\langle \pi_{m\mathbf{k}}|\hat{H}|\pi_{m\mathbf{k}}\rangle}{\langle \pi_{m\mathbf{k}}|\hat{S}|\pi_{m\mathbf{k}}\rangle}, \tag{49}$$

subject to the constraint that $|\pi_{m\mathbf{k}}\rangle \in (\mathcal{R}_\mathbf{k}\cup\mathcal{C}_\mathbf{k}) \subset \mathcal{B}(\mathbf{k})$ and is furthermore orthonormal to $|\pi_{m'\mathbf{k}}\rangle$'s with $m'<m$;

$$\langle \pi_{m\mathbf{k}}|\hat{S}|\pi_{m'\mathbf{k}}\rangle = \delta_{mm'}. \tag{50}$$

It is clear that $|\pi_{m\mathbf{k}}\rangle$ is a linear combination of $|Q_{Ii}\rangle$ through the above transformation matrix $\Pi(\mathbf{k})$ at $\mathbf{k}$ point;

$$|\pi_{m\mathbf{k}}\rangle = \sum_{Ii} \Pi_{Ii,m\mathbf{k}}|Q_{Ii,\mathbf{k}}\rangle, \tag{51}$$

where $|Q_{Ii,\mathbf{k}}\rangle$ is defined as the Bloch sum of $|Q_{Ii}\rangle$,

$$Q_{Ii,\mathbf{k}}(\mathbf{x}) \equiv \sum_L e^{i\mathbf{k}\cdot\mathbf{X}_L}Q_{Ii}(\mathbf{x}-\mathbf{X}_L) = L_1L_2L_3\sum_m \phi_{m\mathbf{k}}(\mathbf{x})(\mathbf{\Omega}_\mathbf{k})_{m,Ii}. \tag{52}$$

By replacing $|\pi_{m\mathbf{k}}\rangle$ in the normalization condition shown in Eq. (50) with its expression in Eq. (52), we immediately obtain the following normalization condition for $\Pi_{Ii,m\mathbf{k}}$:

$$\sum_{Ii,Jj} \Pi^*_{Ii,m\mathbf{k}}O_{Ii,Jj}(\mathbf{k})\Pi_{Jj,m'\mathbf{k}} = \frac{\delta_{mm'}}{L_1L_2L_3}, \tag{53}$$

or in the matrix form

$$\Pi^\dagger(\mathbf{k})\mathbf{O}(\mathbf{k})\Pi(\mathbf{k}) = \frac{\mathbf{I}}{L_1L_2L_3}, \tag{54}$$

where $\mathbf{I}$ is a $qN\times qN$ identity matrix.

B. Mulliken charge and bond order

The Mulliken charge[72] is one popular definition of electronic charge associated with each atom. Here we give a derivation of the Mulliken charge analysis using the density-matrix formalism. We know that the trace of density operator

$\hat{\rho}$ defined under the basis of orthonormal Bloch states $\{\pi_{m\mathbf{k}}\}$ is equal to the total number of valence electrons since $\pi_{n\mathbf{k}} = \phi_{n\mathbf{k}} \equiv \psi_{n\mathbf{k}}$ for $n=1,\ldots,R_{\mathbf{k}}$. In addition, $\{\pi_{m\mathbf{k}}\}$ can be expressed as linear combinations of QOs $\{Q_{Ii,\mathbf{k}}\}$ as shown in Eq. (51). Thus the trace of density matrix can be represented in QO basis if the basis set is complete for the occupied Bloch subspace. If $\mathcal{R}(\mathbf{k})$ contains the occupied Bloch subspace and since $\mathcal{R}(\mathbf{k}) \subset \mathcal{Q}(\mathbf{k})$, this requirement is fulfilled. Therefore, by simply representing $\hat{\rho}$ in QO basis, we obtain atom-specific charge decomposition that satisfies the exact sum rule. Taking PAW formulation as an example, the density operator is defined as

$$\hat{\rho} \equiv \sum_{m\mathbf{k}} f_{m\mathbf{k}} |\tilde{\pi}_{m\mathbf{k}}\rangle\langle\tilde{\pi}_{m\mathbf{k}}|, \tag{55}$$

where $f_{m\mathbf{k}}$ is electron occupation number in the corresponding Bloch state $|\tilde{\pi}_{m\mathbf{k}}\rangle$ that is either 1 or 0 when m includes both band and spin index. In the PAW formalism,[7] true Bloch wave function $|\tilde{\pi}_{m\mathbf{k}}\rangle$ and pseudo-Bloch wave function $|\pi_{m\mathbf{k}}\rangle$ are related through transformation operator $\hat{T}$,

$$|\tilde{\pi}_{m\mathbf{k}}\rangle = \hat{T}|\pi_{m\mathbf{k}}\rangle,$$

while $\hat{S}$ and $\hat{T}$ are related by $\hat{S}=\hat{T}^{\dagger}\hat{T}$. Then

$$\hat{\rho} = \sum_{m\mathbf{k}} f_{m\mathbf{k}} \hat{T}|\pi_{m\mathbf{k}}\rangle\langle\pi_{m\mathbf{k}}|\hat{T}^{\dagger}. \tag{56}$$

Clearly,

$$\mathrm{Tr}(\hat{\rho}) = L_1L_2L_3N_e, \tag{57}$$

where $N_e=rN$ is the number of valence electrons within one supercell. There is also an idempotent property,

$$\hat{\rho}^2 = \hat{\rho}. \tag{58}$$

To split charge onto different orbitals on each atom, we represent the density operator $\hat{\rho}$ in Eq. (56) in terms of QO using Eq. (51),

$$\hat{\rho} = \sum_{m\mathbf{k}} f_{m\mathbf{k}} \sum_{Jj,Ii} \Pi_{Jj,m\mathbf{k}}\Pi^*_{Ii,m\mathbf{k}} \hat{T}|Q_{Jj,\mathbf{k}}\rangle\langle Q_{Ii,\mathbf{k}}|\hat{T}^{\dagger}. \tag{59}$$

Then

$$\begin{aligned}
\mathrm{Tr}(\hat{\rho}) &= \sum_{\sigma} \int d^3\mathbf{x}\langle\mathbf{x}|\hat{\rho}|\mathbf{x}\rangle \\
&= \sum_{m\mathbf{k}} f_{m\mathbf{k}} \sum_{Jj,Ii} \Pi_{Jj,m\mathbf{k}}\Pi^*_{Ii,m\mathbf{k}} \langle Q_{Ii,\mathbf{k}}|\hat{T}^{\dagger}\hat{T}|Q_{Jj,\mathbf{k}}\rangle \\
&= \sum_{Jj,Ii}\sum_{m\mathbf{k}} f_{m\mathbf{k}}\Pi_{Jj,m\mathbf{k}}\Pi^*_{Ii,m\mathbf{k}} \times \sum_{L,L'} e^{i\mathbf{k}\cdot(\mathbf{X}_L-\mathbf{X}_{L'})}\langle Q^{L'}_{Ii}|\hat{S}|Q^{L}_{Jj}\rangle \\
&= \sum_{Jj,Ii}\sum_{m\mathbf{k}} f_{m\mathbf{k}}\Pi_{Jj,m\mathbf{k}}\Pi^*_{Ii,m\mathbf{k}} \times L_1L_2L_3\sum_{L} e^{i\mathbf{k}\cdot\mathbf{X}_L}\langle Q^{0}_{Ii}|\hat{S}|Q^{L}_{Jj}\rangle \\
&= L_1L_2L_3\sum_{\mathbf{k}}\sum_{Jj,Ii} D_{Jj,Ii}(\mathbf{k})O_{Ii,Jj}(\mathbf{k}),
\end{aligned} \tag{60}$$

where $\mathbf{D}(\mathbf{k})$ and $\mathbf{O}(\mathbf{k})$ matrices are defined as the following:

$$D_{Jj,Ii}(\mathbf{k}) \equiv \sum_{m} f_{m\mathbf{k}}\Pi_{Jj,m\mathbf{k}}\Pi^*_{Ii,m\mathbf{k}}, \tag{61}$$

$$O_{Ii,Jj}(\mathbf{k}) \equiv \sum_{L} e^{i\mathbf{k}\cdot\mathbf{X}_L}\langle Q^{0}_{Ii}|\hat{S}|Q^{L}_{Jj}\rangle. \tag{62}$$

Clearly $D_{Jj,Ii}(\mathbf{k})$ represents the element of density matrix $\mathbf{D}(\mathbf{k})$ between $|Q_{Ii,\mathbf{k}}\rangle$ and $|Q_{Jj,\mathbf{k}}\rangle$, while $O_{Ii,Jj}(\mathbf{k})$ represents the element of overlap matrix $\mathbf{O}(\mathbf{k})$ between two QOs at the same $\mathbf{k}$ point. Both $\mathbf{D}(\mathbf{k})$ and $\mathbf{O}(\mathbf{k})$ are the Hermitian matrices.

Thus we can straightforwardly define the Mulliken charge on a particular QO as

$$\rho_{Ii} \equiv \sum_{\mathbf{k}}\sum_{I'i'} D_{Ii,I'i'}(\mathbf{k})O_{I'i',Ii}(\mathbf{k}), \tag{63}$$

and the Mulliken charge on atom I as

$$\rho_I \equiv \sum_{i} \rho_{Ii}, \tag{64}$$

resulting in a simple sum rule from Eqs. (57) and (60);

$$\sum_{I} \rho_I = N_e. \tag{65}$$

Similarly, bond order between any two atoms can be derived using $\hat{\rho}^2$. We note from Eqs. (59) and (61) that

$$\hat{\rho} = \sum_{\mathbf{k},Jj,Ii} D_{Jj,Ii}(\mathbf{k})|\tilde{Q}_{Jj,\mathbf{k}}\rangle\langle\tilde{Q}_{Ii,\mathbf{k}}|, \tag{66}$$

where $|\tilde{Q}_{Ii,\mathbf{k}}\rangle \equiv \hat{T}|Q_{Ii,\mathbf{k}}\rangle$ so

$$\hat{\rho}^2 = \sum\nolimits_{\mathbf{k},Jj,Ii,I'i',J'j'} D_{Jj,Ii}(\mathbf{k})D_{I'i',J'j'}(\mathbf{k})|\tilde{Q}_{Jj,\mathbf{k}}\rangle\langle\tilde{Q}_{Ii,\mathbf{k}}|\tilde{Q}_{I'i',\mathbf{k}}\rangle \times\langle\tilde{Q}_{J'j',\mathbf{k}}|,$$

and

$$\mathrm{Tr}(\hat{\rho}^2) = \sum\nolimits_{\mathbf{k},Jj,Ii,I'i',J'j'} D_{Jj,Ii}(\mathbf{k})D_{I'i',J'j'}(\mathbf{k})\langle\tilde{Q}_{Ii,\mathbf{k}}|\tilde{Q}_{I'i',\mathbf{k}}\rangle \times\langle\tilde{Q}_{J'j',\mathbf{k}}|\tilde{Q}_{Jj,\mathbf{k}}\rangle.$$

We note from Eq. (62) that

$$\langle\tilde{Q}_{Ii,\mathbf{k}}|\tilde{Q}_{I'i',\mathbf{k}}\rangle = L_1L_2L_3O_{Ii,I'i'}(\mathbf{k}). \tag{67}$$

So we get

$$\mathrm{Tr}(\hat{\rho}^2) = (L_1L_2L_3)^2\sum_{\mathbf{k}} \mathrm{Tr}[\mathbf{D}(\mathbf{k})\mathbf{O}(\mathbf{k})\mathbf{D}(\mathbf{k})\mathbf{O}(\mathbf{k})], \tag{68}$$

where Tr[ ] is the matrix trace. Indeed the derivations above can be easily generalized into

$$\mathrm{Tr}(\hat{\rho}^n) = (L_1L_2L_3)^n\sum_{\mathbf{k}} \mathrm{Tr}\{[\mathbf{D}(\mathbf{k})\mathbf{O}(\mathbf{k})]^n\}, \quad n=1,\ldots,\infty.$$

Let us define $\mathbf{P}(\mathbf{k}) \equiv \mathbf{D}(\mathbf{k})\mathbf{O}(\mathbf{k})$, with

$$P_{Ii,Jj}(\mathbf{k}) \equiv \sum_{I'i'} D_{Ii,I'i'}(\mathbf{k})O_{I'i',Jj}(\mathbf{k}). \tag{69}$$

The discrete Fourier transform of $P_{Ii,Jj}(\mathbf{k})$ is

$$P_{Ii,Jj}(\mathbf{X}_L) \equiv \sum_{\mathbf{k}} P_{Ii,Jj}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{X}_L}, \tag{70}$$

and

$$P_{Ii,Jj}(\mathbf{k}) = \frac{1}{L_1L_2L_3}\sum_L P_{Ii,Jj}(\mathbf{X}_L)e^{-i\mathbf{k}\cdot\mathbf{X}_L}. \tag{71}$$

It can then be easily shown that

$$\mathrm{Tr}(\hat{\rho}^n) = L_1L_2L_3\sum\nolimits_{\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_{n-1}} \mathrm{Tr}[\mathbf{P}(\mathbf{X}_1)\mathbf{P}(\mathbf{X}_2)\cdots\mathbf{P}(\mathbf{X}_{n-1})\mathbf{P}(-\mathbf{X}_1-\mathbf{X}_2-\cdots-\mathbf{X}_{n-1})]. \tag{72}$$

Thus the real-space matrix $\mathbf{P}(\mathbf{X}_L)$ in Eq. (70) completely characterizes bonding in the system.

So we may define a *pair-specific* quantity between Ii in supercell 0 and Jj in supercell $\mathbf{X}_L$ as

$$B^L_{Ii,Jj} \equiv P_{Ii,Jj}(\mathbf{X}_L)P_{Jj,Ii}(-\mathbf{X}_L) \tag{73}$$

and that between atom I in supercell 0 and atom J in supercell $\mathbf{X}_L$ as

$$B^L_{I,J} \equiv \sum_{ij} B^L_{Ii,Jj}, \tag{74}$$

which satisfy the sum rule

$$\sum_{I,J,L} B^L_{I,J} = \frac{1}{L_1L_2L_3}\mathrm{Tr}(\hat{\rho}^2) = N_e. \tag{75}$$

According to convention, H–H is a single bond and should have bond order 1, while C=C is a double bond and should have bond order 2. Let us calibrate against this convention for hydrogen molecule. Suppose we have the bonding states $(|s_1\rangle+|s_2\rangle)\uparrow/\sqrt{2}$ and $(|s_1\rangle+|s_2\rangle)\downarrow/\sqrt{2}$, and the antibonding states $(|s_1\rangle-|s_2\rangle)\uparrow/\sqrt{2}$ and $(|s_1\rangle-|s_2\rangle)\downarrow/\sqrt{2}$, where for simplicity we assume $|s_1\rangle$ and $|s_2\rangle$ are orthogonal to each other. Then the overlap matrix $\mathbf{O}$ is a 4×4 identity matrix, and the density matrix $\mathbf{D}$ has two block 2×2 submatrices with all submatrix elements equal to 0.5. The population matrix $\mathbf{P}=\mathbf{DO}=\mathbf{D}$. Then from Eq. (73) we obtain $B^0_{Ii,Jj}$ as having two 2×2 submatrices with all submatrix elements equal to 0.25. By summing over all matrix elements we have $\Sigma_{Ii,Jj}B^0_{Ii,Jj}=N_e=2$. Thus we have $B^0_{1,2}=B^0_{1,2}(\uparrow)+B^0_{1,2}(\downarrow)=0.5$, and we see that the bond order defined in literature is twice as much as $B^0_{I,J}$. We will therefore always use $2B^L_{I,J}$ or $2\Sigma_{ij}B^L_{Ii,Jj}$ for bond order between two atoms in real systems, as shown in Table III.

Note that in sum rule (75), there are contributions from terms such as

$$B^0_{Ii,Ii} = [P_{Ii,Ii}(\mathbf{X}_L=0)]^2, \tag{76}$$

as well as

$$B^0_{Ii,Ij} = P_{Ii,Ij}(\mathbf{X}_L=0)P_{Ij,Ii}(\mathbf{X}_L=0). \tag{77}$$

According to Eq. (70),

$$P_{Ii,Ii}(\mathbf{X}_L=0) = \sum_{\mathbf{k}} P_{Ii,Ii}(\mathbf{k}) = \rho_{Ii}. \tag{78}$$

So the Mulliken charge squared ρ^2_{Ii} and same-site-different-orbital couplings $P_{Ii,Ij}(\mathbf{X}_L=0)P_{Ij,Ii}(\mathbf{X}_L=0)$ appear in sum rule (75), which means the sum of different site $B^L_{I,J}$'s should be less than the total number of electrons N_e. This is consistent with the practice of using $2B^L_{I,J}$ to denote bond order. Note also that there can be lone pairs in the system and not all electrons need to be engaged in bonding. Indeed, as we separate H–H to distance infinity, we see that it is not reasonable to demand the bond order to stay at integer 1.

The definition above assumes all N_e electrons reside in *bonding states*. The more general definition of bond order in chemical literature is bond order $\equiv$(number of bonding electrons $-$number of antibonding electrons)/2. The subtraction occurs when some eigenstates $|\pi_{m\mathbf{k}}\rangle$ are occupied but are deemed *antibonding*, for instance with eigenenergies above an internal gap that varies sensitively with atomic distance. In such a case, the total density operator needs to be split into bond and antibonding parts;

$$\hat{\rho}_{\mathrm{bond}} \equiv \sum_{m\mathbf{k}} f^{\mathrm{bond}}_{m\mathbf{k}}|\tilde{\pi}_{m\mathbf{k}}\rangle\langle\tilde{\pi}_{m\mathbf{k}}|, \tag{79}$$

$$\hat{\rho}_{\mathrm{anti}} \equiv \sum_{m\mathbf{k}} f^{\mathrm{anti}}_{m\mathbf{k}}|\tilde{\pi}_{m\mathbf{k}}\rangle\langle\tilde{\pi}_{m\mathbf{k}}|, \tag{80}$$

where $f^{\mathrm{bond}}_{m\mathbf{k}}=1$ for occupied bonding states and 0 otherwise and $f^{\mathrm{anti}}_{m\mathbf{k}}=1$ for occupied antibonding states and 0 otherwise with $f^{\mathrm{bond}}_{m\mathbf{k}}f^{\mathrm{anti}}_{m\mathbf{k}}=0$. The following sum rules hold:

$$\mathrm{Tr}(\hat{\rho}_{\mathrm{bond}}) = N_{\mathrm{bond}}, \quad \mathrm{Tr}(\hat{\rho}_{\mathrm{anti}}) = N_{\mathrm{anti}}, \tag{81}$$

where N_{bond} is the total number of bonding electrons and N_{anti} is the total number of antibonding electrons. All derivations of Eqs. (58)–(75) apply to $\hat{\rho}_{\mathrm{bond}}$ and $\hat{\rho}_{\mathrm{anti}}$ individually. We can therefore compute $B^{\mathrm{bond}L}_{I,J}$ and $B^{\mathrm{anti}L}_{I,J}$ individually and then subtract

$$B^L_{I,J} \equiv B^{\mathrm{bond}L}_{I,J} - B^{\mathrm{anti}L}_{I,J} \tag{82}$$

to get the net bond order. QO analysis would work so long as $\mathcal{R}(\mathbf{k})$ includes both the deemed bonding and antibonding eigenstates.

C. Projected density of states

Projected density of states (PDOS) is a powerful tool for analyzing energy- and site-resolved electronic structure. Let us define the total density of states (DOS) of our *ab initio* tight-binding system to be

$$\rho(\varepsilon) \equiv \frac{1}{L_1L_2L_3}\sum_{m\mathbf{k}} \delta(\varepsilon - e_{m\mathbf{k}}), \tag{83}$$

where $e_{m\mathbf{k}}$ has the interpretation of constrained variational Rayleigh quotient [Eq. (49)]. $\rho(\varepsilon)$ clearly satisfies the total sum rules,

$$\int_{-\infty}^{\varepsilon_F} d\varepsilon\rho(\varepsilon) = N_e = rN, \tag{84}$$

and

$$\int_{-\infty}^{\infty} d\varepsilon\rho(\varepsilon) = qN. \tag{85}$$

In real numerical calculations, $\delta(\varepsilon-e_{m\mathbf{k}})$ is often replaced by normalized Gaussian centered around $e_{m\mathbf{k}}$.

TABLE I. Parameters used in plane-wave DFT calculation and QO construction for various systems. ε_{th} is the energy threshold for $\mathcal{R}(\mathbf{k})$ selection (the Fermi energy ε_F is set to 0). R_{cut} is the radial cutoff of tight-binding Hamiltonian and overlap matrices in Eq. (47).

Material	No. of atoms	Structure	a_0 and c_0 (Å)	E_{cut} (eV)	No. of **k** points	No. of bands	XC	R_{cut} (Å)	ε_{th} (eV)
CH_4	5		1.1	350	Γ point	60	PW91	8.0	0
SiH_4	5		1.48	350	Γ point	40	PW91	8.0	0
Si	2	Diamond	5.430	300	7×7×7	60	PW91	12.0	0
β-SiC	2	fcc	4.32	350	7×7×7	40	PW91	12.0	0
Al	1	fcc	4.030	300	9×9×9	60	PW91	8.0	1.0
Fe[a]	1	bcc	2.843	400	9×9×9	40	PW91	10.0	3.0
Mo	1	bcc	3.183	400	13×13×13	20	PW91	10.0	0.0/8.0[b]
MgB_2	3	hcp	3.067, 3.515	300	7×7×7	40	PW91	10.0	3.0

[a]Ferromagnetic.
[b]We use $\varepsilon_{\text{th}}=0$ eV for $\{s,d\}$ basis and 8.0 eV for $\{s,p,d\}$ basis.

Our goal is to decompose $\rho(\varepsilon)$ into a sum of site, angular momentum, and spin-specific PDOS functions;

$$\rho(\varepsilon) = \sum_{Ii} \rho_{Ii}(\varepsilon). \tag{86}$$

Because the QOs are nonorthogonal, the decomposition cannot be done by a simple projection.[73]

The solution is very simple. Replacing $f_{m\mathbf{k}}$ by $\delta(\varepsilon - e_{m\mathbf{k}})/(L_1L_2L_3)$ in Eq. (55), we can define energy-resolved density operator,

$$\hat{\rho}(\varepsilon) \equiv \frac{1}{L_1L_2L_3}\sum_{m\mathbf{k}} \delta(\varepsilon - e_{m\mathbf{k}})|\tilde{\pi}_{m\mathbf{k}}\rangle\langle\tilde{\pi}_{m\mathbf{k}}|. \tag{87}$$

Clearly

$$\text{Tr}[\hat{\rho}(\varepsilon)] = \rho(\varepsilon). \tag{88}$$

Thus, if we just replace $f_{m\mathbf{k}}$ by $\delta(\varepsilon - e_{m\mathbf{k}})/(L_1L_2L_3)$ everywhere in Eqs. (55)–(65), the entire decomposition scheme would work for $\rho(\varepsilon)$. We will have energy-resolved density matrix,

$$D_{Jj,Ii}(\mathbf{k},\varepsilon) \equiv \sum_m \delta(\varepsilon - e_{m\mathbf{k}})\Pi_{Jj,m\mathbf{k}}\Pi^*_{Ii,m\mathbf{k}} \tag{89}$$

and

$$\rho(\varepsilon) = \text{Tr}[\hat{\rho}(\varepsilon)] = \sum_{\mathbf{k}}\sum_{Jj,Ii} D_{Jj,Ii}(\mathbf{k},\varepsilon)O_{Ii,Jj}(\mathbf{k}). \tag{90}$$

All we need to do is therefore to define the projected density of states as

$$\rho_{Jj}(\varepsilon) \equiv \sum_{\mathbf{k}}\sum_{Ii} D_{Jj,Ii}(\mathbf{k},\varepsilon)O_{Ii,Jj}(\mathbf{k}), \tag{91}$$

and the PDOS sum rule [Eq. (86)] would be satisfied for every ε. A rigorous connection between $\rho_{Ii}(\varepsilon)$ and the QO-based Mulliken charge,

$$\int_{-\infty}^{\varepsilon_F} d\varepsilon \rho_{Ii}(\varepsilon) = \rho_{Ii}, \tag{92}$$

exists, with ρ_{Ii} was defined in Eq. (63). Thus $\rho_{Ii}(\varepsilon)$ can be regarded as the energy-resolved Mulliken charge.

Following the same procedure, we can define energy-resolved bond order $2B^L_{Ii,Jj}(\varepsilon)$ and its integral,

$$2B^L_{Ii,Jj}(\varepsilon_1,\varepsilon_2) \equiv 2\int_{\varepsilon_1}^{\varepsilon_2} d\varepsilon B^L_{Ii,Jj}(\varepsilon). \tag{93}$$

For example, it is valid to say that among the total 1.2 bond order between atom I in supercell 0 and atom J in supercell L, energy bands in $[\varepsilon_F - 5, \varepsilon_F - 2]$ contribute 0.7.

V. QO APPLICATIONS

We have constructed QO for various materials, including semiconductors, simple metals, ferromagnetic materials, transition metals and their oxides, high-temperature superconductors, and quasi-one-dimensional materials such as carbon nanotubes. These QOs are then used for *ab initio* tight-binding calculations, including band structure, density of states, QO-projected band structure and density of states, and the high-resolution Fermi surface. We have also combined QO with the Green's function method to efficiently calculate electrical conductance of molecular and nanoscale junctions using the Landauer formalism.[74] Currently we have implemented QO interfaces[27] to VASP and DACAPO; the source codes of our method and input conditions for all examples in this section are put on the web.[27]

In this paper, the ground-state electronic configurations are calculated using DACAPO DFT package[26,75,76] with Vanderbilt USPP (Refs. 3–5) and PW91 generalized gradient approximation (GGA) of the exchange-correlation functional.[77] Parameters for the DFT calculations are included in Table I. Due to page limitation, we demonstrate only four materials in detail: diamond cubic silicon, β-silicon carbide, bcc ferromagnetic iron, and bcc molybdenum.

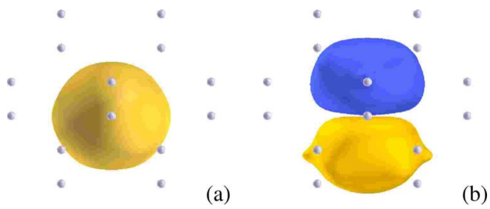

FIG. 2. (Color online) QO in Si crystal. (a) s-like and (b) p_z-like. [Absolute isosurface value: 0.03 Å$^{-3/2}$. Yellow or light gray for positive values and blue or dark gray for negative values. The same color scheme is used in all the other isosurface plots of QOs in this paper. They are plotted with XCRYSDEN (Refs. 78–80).]

A. Semiconductor: Diamond cubic Si crystal

The diamond cubic Si crystal has an indirect band gap of 1.17 eV at 0 K. In Fig. 2 we show two of total eight QOs: s-like and p_z-like QOs. Since in this case we use the unpolarized spin configuration, we have the same s-like and p-like QOs for both spins. As shown in the figure these QOs are slightly deformed due to the interaction with nearest-neighbor atoms, but the overall shape of s and p_z is largely maintained. Figure 3 compares the band structure between plane-wave DFT and *ab initio* TB calculations. It is seen that among the total eight TB bands, four valence bands below ε_F are exactly reproduced with each band doubly occupied.

The indirect band gap from DFT calculation is about 0.7 eV, smaller than 1.17 eV from experiments, which is a common problem of DFT due to the ground-state nature of DFT and inaccurate exchange-correlation functional. However QO-based TB calculation gives a band gap of around 2.0 eV. In general the conduction bands from *ab initio* TB calculation using QO basis set are higher than those from plane-wave DFT calculation due to the constrained variation interpretation of the TB eigenvalues [Eq. (49)]. They are higher because the optimized combination Bloch states $\{c_{m\mathbf{k}}\}$ are manually constructed and they are not true unoccupied low-lying Bloch eigenstates. In other words, these *optimized* combination states in $\mathcal{C}(\mathbf{k})$ can be represented by a linear combination of the infinite *true* unoccupied Bloch states in

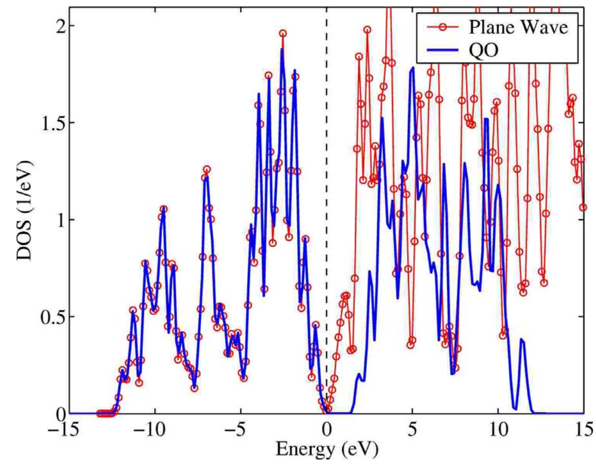

FIG. 4. (Color online) Density of states of Si crystal. (Circle-dot line: plane-wave DFT calculation; solid line: TB calculation; and dashed line: Fermi energy.)

$\overline{\mathcal{R}}(\mathbf{k})$. Therefore the eigenenergies (Rayleigh quotients) above the energy threshold ε_{th} ($\varepsilon_{\text{th}}=\varepsilon_F$ in this case) from QO-based TB calculation are always higher than the Kohn-Sham eigenenergies. DOS in Fig. 4 also shows this energy shift in the conduction bands, while DOS below ε_F is exactly reproduced.

B. Covalent compound: β-SiC crystal

Silicon carbide is a typical covalent compound and it has two well-known polymorphs: α-SiC and β-SiC. The former is an intrinsic semiconductor in hexagonal structures and the latter has an indirect band gap of 2.2 eV in zinc-blende-type structure. From DFT calculation of β-SiC, a band gap of around 1.0 eV is found, while from our *ab initio* TB calculation it is around 3.0 eV. Band structure (Fig. 5) and density of states (Fig. 6) in conduction bands from TB calculation change a lot and shift up due to the same reason as in the Si crystal case. It is seen from Fig. 7 that both s-like and p_z-like QOs of Si atom are relatively more delocalized than those of

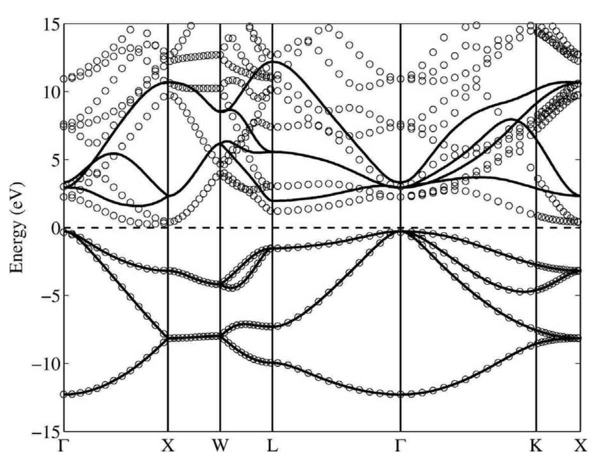

FIG. 3. Band structure of Si crystal. (Circle dot: plane-wave DFT calculation; solid line: TB calculation based on eight QOs; and dashed line: Fermi energy.)

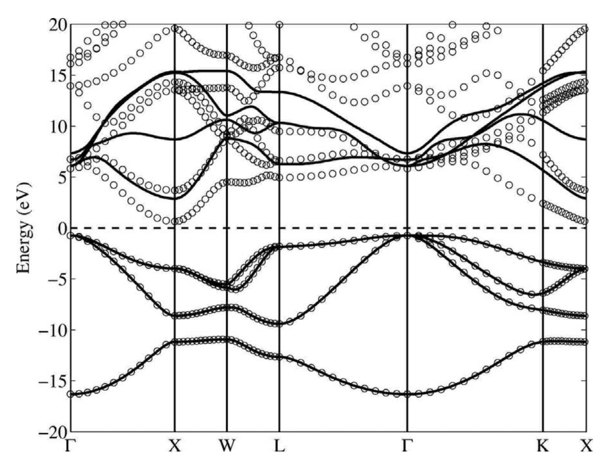

FIG. 5. Band structure of β-SiC. (Circle dot: plane-wave DFT calculation; solid line: TB calculation based on eight QOs; and dashed line: Fermi energy.)

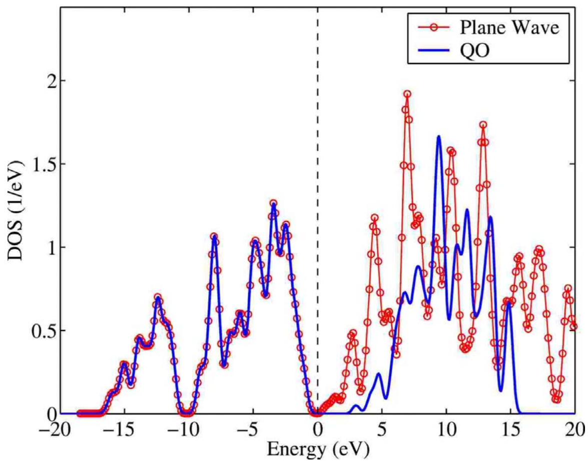

FIG. 6. (Color online) Density of states of β-SiC. (Circle-dot line: plane-wave DFT calculation; solid line: TB calculation; and dashed line: Fermi energy.)

C atom, which suggests Si has less ability to attract electron than C in β-SiC crystal. This intuition is confirmed by the QO-projected density of states plot in Fig. 8 where the total density of states on C atom below ε_F is much more than that on Si atom and it further indicates that more charges are localized at C atom. The total area of Fig. 8 below ε_F for each atom is exactly equal to the total Mulliken charge associated with each atom. Note that the sum of QO-projected density of states [Eq. (91)] is exactly equal to the total density of states, while this is not true for standard atomic-orbital-projected density of states widely used in analyzing plane-wave DFT results.

Compared to Fig. 3 in the Si crystal case, there is a large splitting between two bottom bands along the X-W line in Fig. 5 in the SiC crystal. Four higher peaks of DOS, shown in Fig. 8, are useful for explaining this splitting. Two peaks around -12.0 eV (C's s peak in the bottom panel and Si's p peak in the top panel) and another two peaks around -8.0 eV (C's p peak in the bottom panel and Si's s peak in

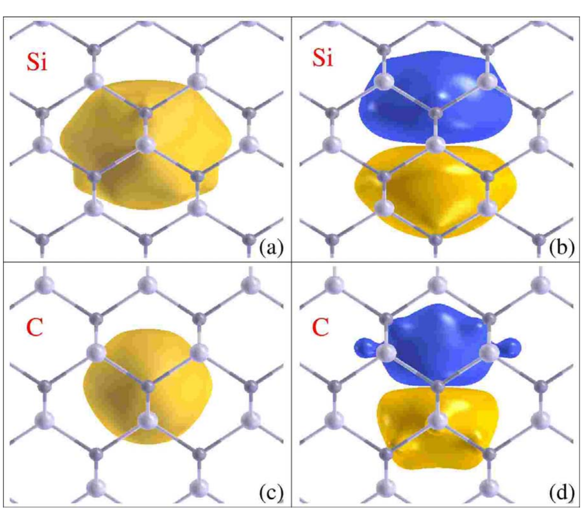

FIG. 7. (Color online) QO in β-SiC crystal. (a) Si: s-like; (b) Si: p_z-like; (c) C: s-like; and (d) C: p_z-like. (Absolute isosurface value: 0.03 Å$^{-3/2}$.)

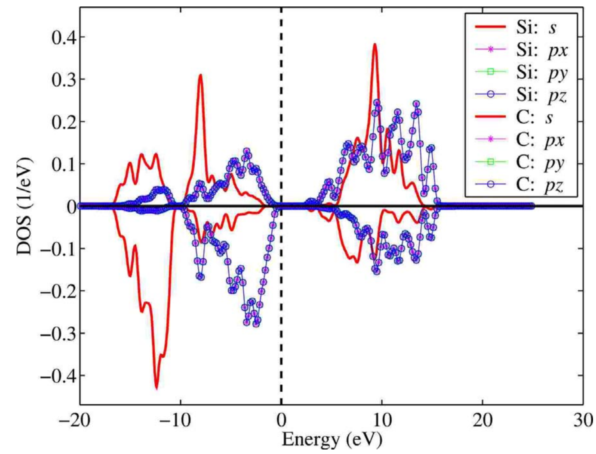

FIG. 8. (Color online) QO-projected density of states of β-SiC. (Top panel: Si; bottom panel: C; and dashed line: Fermi energy.)

the top panel) lead to two nonsymmetric types of s-p bonding. One is the bond between Si's s-like QO and C's p-like QOs and the other is the bond between C's s-like QO and Si's p-like QOs. In Si crystal the above two types are degenerate bonds, which give two degenerate bands at the bottom of band structure between X and W. This splitting is much more clearly reflected in QO-projected band structure shown in Fig. 9, where the bonding between silicon's s-like QO and carbon's three p-like QOs is dominant in the higher-energy band while the bonding between carbon's s-like QO and silicon's three p-like QOs is dominant in the lower-energy band.

To further study electron transfer we investigate the Mulliken charges in three different compounds shown in Table II, including methane (CH_4), silane (SiH_4), and β-SiC. It is seen that the capability of three different elements to attract electrons is in the following order: C$>$H$>$Si. Table III shows bond order between atoms and their first-nearest and second-nearest neighbors in various systems. It is not surprising that in covalent systems bond order between the atom and its second-nearest neighbor is almost zero and it is much less than the bond order between the atom and its first-nearest neighbor. However, unlike covalent systems, fcc aluminum, bcc molybdenum, and bcc iron have smaller bond orders for

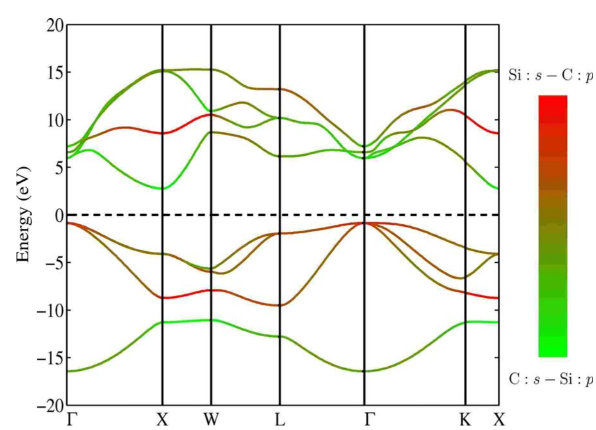

FIG. 9. (Color online) QO-projected band structure of SiC crystal with red (dark gray) for Si s and C p and green (light gray) for C s and Si p.

TABLE II. The Mulliken charges for CH_4, SiH_4, and β-SiC.

Material	Mulliken Charge		Total charge
CH_4	C: 5.160	H: 0.710	8.0
SiH_4	Si: 3.300	H: 1.175	8.0
β-SiC	Si: 2.729	C: 5.271	8.0

both the first-nearest and second-nearest neighbors as shown in the table, indicating metallic bonding. In the case of MgB_2 crystal, it shows strong covalent bonding on the boron plane and relative large bond order between boron and magnesium but very small bond order between magnesium atoms. The latter is due to large distance between magnesium atomic layers and the ionic nature of magnesium in MgB_2 crystal. It should be emphasized that the QO-based Mulliken charge and bond order satisfy the sum rules very well, which is not the case for the traditional charge analysis, widely used for analyzing plane-wave DFT calculations, by setting a radial cutoff and integrating electron density within that radius around each atom.

C. Ferromagnetic bcc Fe crystal

Ferromagnetic bcc iron is investigated, in which we expect some differences between the QOs with majority spin and those with minority spin. Here the energy threshold is 3 eV above ε_F to keep electronic structure near the Fermi energy to be exact. Pseudoatomic orbitals 3d, 4s, and 4p are rescaled by $e^{-\eta|\mathbf{x}|}$, with η=1.0 Å^{-1} and then renormalized.

Figure 10 displays 10 of the total 18 QOs. The QOs with majority spin and minority spin, on the left and middle columns, respectively, look quite similar. Their differences are shown in the right column, having the same symmetry as the corresponding QOs. Figures 11 and 12 present two different band structures with majority spin and minority spin, respectively. Similar to the above two band structures, DOS plotted in Fig. 13 displays the dramatic difference of electronic structure information between majority spin and minority spin in bcc Fe. As expected, Figs. 11–13 demonstrate that all the electronic structure below the energy threshold is well reproduced by QO.

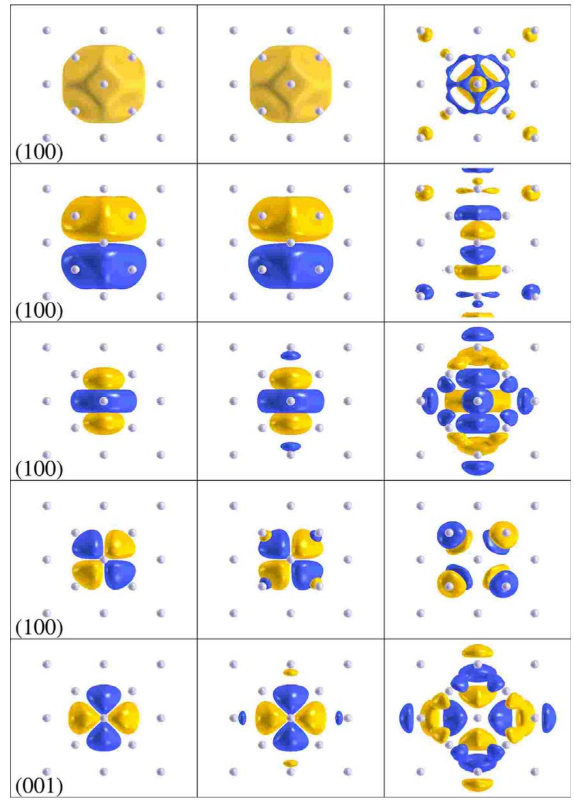

FIG. 10. (Color online) QO in bcc Fe crystal. From top to bottom they are s-like, p_z-like, d_{z^2}-like, d_{yz}-like, and $d_{x^2-y^2}$-like QOs. Left column: QO with majority spin (absolute isosurface value: 0.03 Å$^{-3/2}$). Middle column: QO with minority spin (absolute isosurface value: 0.03 Å$^{-3/2}$). Right column: difference between QO with majority spin and QO with minority spin (absolute isosurface value: 0.003 Å$^{-3/2}$).

TABLE III. Bond orders for various systems.

Material	Bond order $(2\Sigma_{ij}B^L_{Ii,Jj})$		Total BO/sum rule
CH_4	C–H: 0.882	H–H: 0.012	8.0/8.0
SiH_4	Si–H: 0.866	H–H: 0.033	8.0/8.0
β-SiC	Si–C: 0.823	Si–Si: 0.009	8.0/8.0
	C–C: 0.015		
Si-cubic	1st: 0.874[a]	2nd: 0.009	8.0/8.0
Al-fcc	1st: 0.213	2nd: 0.015	2.898/2.896
Fe-bcc (↑) [b]	1st: 0.184	2nd: 0.070	4.967/4.967
Fe-bcc (↓)	1st: 0.328	2nd: 0.114	2.842/2.843
Mo-bcc[c]	1st: 0.589	2nd: 0.193	5.876/5.876
MgB_2	B–B: 0.698	Mg–B: 0.206	13.868/13.868
	Mg–Mg: 0.085		

[a]"1st" and "2nd" stand for the first-nearest and second-nearest neighbors.
[b]↑ for majority spin; ↓ for minority spin.
[c]The calculation is based on $\{s,p,d\}$-QO basis with ε_{th}=8.0 eV.

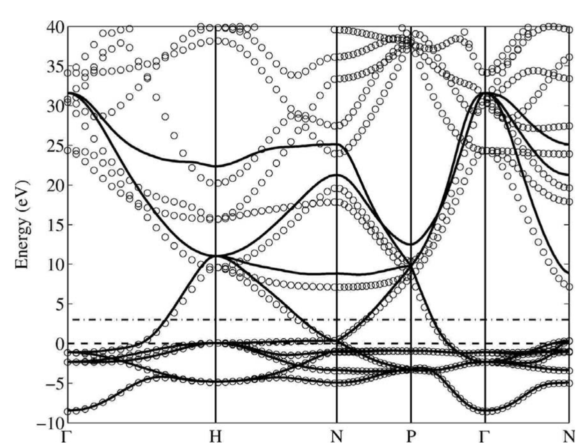

FIG. 11. Band structure of bcc Fe with majority spin. (Circle dot: plane-wave DFT calculation; solid line: TB calculation based on nine QOs for majority spin; dashed line: Fermi energy; and dash-dot line: energy threshold with ε_{th}=3 eV.)

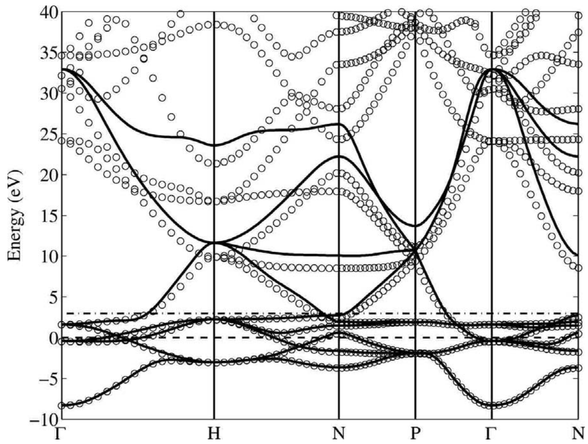

FIG. 12. Band structure of bcc Fe with minority spin. (Circle dot: plane-wave DFT calculation; solid line: TB calculation based on nine QOs for minority spin; dashed line: Fermi energy; and dash-dot line: energy threshold with $\varepsilon_{\rm th}$=3 eV.)

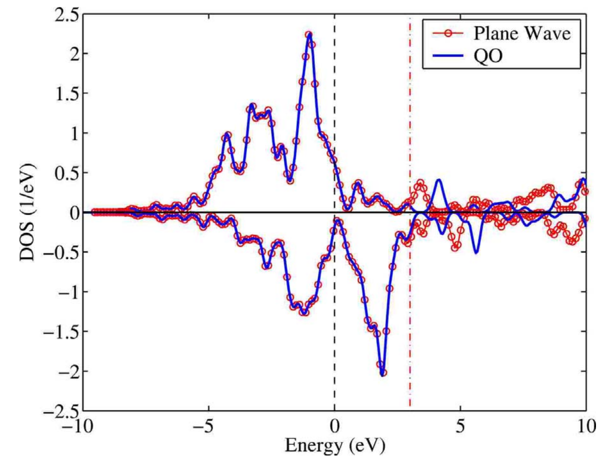

FIG. 13. (Color online) Electronic density of states in bcc Fe. Top panel: majority spin; bottom: minority spin. (Circle dot line: plane-wave DFT calculation; solid line: TB calculation; dashed line: Fermi energy; and dash-dot line: energy threshold with $\varepsilon_{\rm th}$ =3 eV.)

Figures 14(a) and 14(b) present two Fermi surfaces in the first Brillouin zone for the majority spin and minority spin, respectively. In the majority-spin case, the closed surface around Γ point holds electrons while the open surfaces on the zone faces and another two types of small surfaces around H enclose holes. These open surfaces are connected to other surfaces of the same type in the second Brillouin zone forming open orbits across Brillouin zones. In the case of minority spin, the large surfaces around H and those around N near the zone faces form hole pockets, while one octahedral closed surface around Γ and six small spheres inside the Brillouin zone form electron pockets. The computation of the high-resolution Fermi surface in reciprocal space requires thousands of Hamiltonian diagonalization on a very fine grid, which is expensive for plane-wave DFT calculations even if the symmetry property of the Brillouin zone is taken into account. However, QO-based TB method makes the calculation very efficient since we can easily diagonalize the small TB Hamiltonian and overlap matrices. So these high-resolution Fermi surfaces again demonstrate the utility of QO analysis for solids.

D. Minimal basis for bcc Mo crystal

In a previous paper[24] we applied the original QUAMBO method to one of the transition metals, bcc Mo, and obtained $\{s,d\}$ QUAMBOs as the minimal basis. Most of the QUAMBO-based tight-binding band structure (Fig. 3 of Ref. 24) agrees very well with the DFT results; however it shows some deviations around high-symmetry point *N*. In particular, the Γ-*N* and *P*-*N* bands crossing the Fermi energy have several strong wiggles even below ε_F. The original explanation of such deviations is related to the coarse **k**-point sampling which will affect the slope of the band structures near Fermi energy. However, the Monkhorst-Pack grid of $16\times16\times16$ used in Ref. 24 is already quite dense. Therefore, there is more important physical reason responsible for the large deviations around *N* point below ε_F.

To solve the above puzzle, we have constructed two sets of QO basis, $\{s,d\}$ and $\{s,p,d\}$ with $\varepsilon_{\rm th}$=0 and 8 eV, respectively. Pseudoatomic orbitals *s*, *p*, and *d* are rescaled by $e^{-\eta|\mathbf{x}|}$, with η=1.0, 1.5, and 0.5 Å^{-1}, respectively, and then

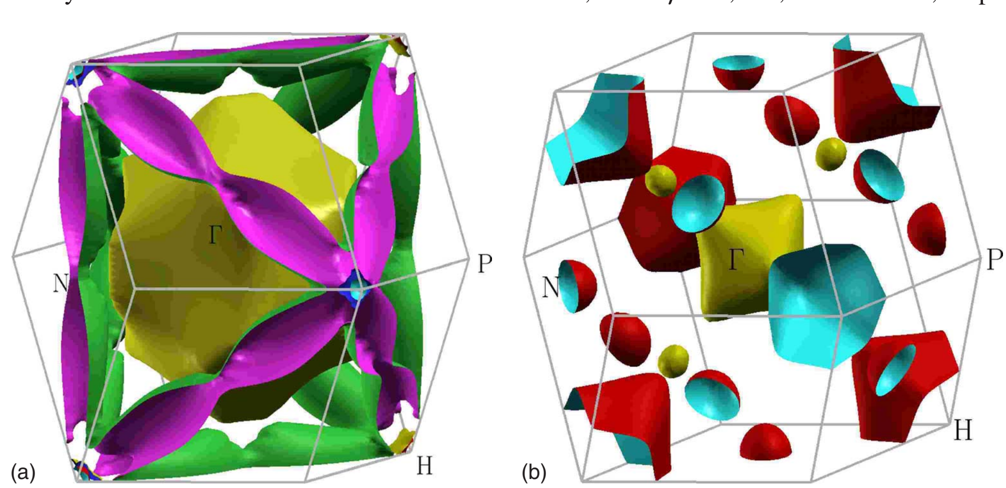

FIG. 14. (Color online) The Fermi surface of bcc Fe with (a) majority spin and (b) minority spin. [Plotted using XCRYSDEN (Refs. 78–80).]

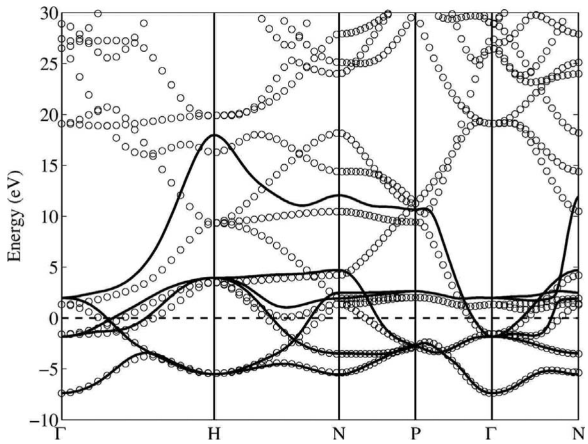

FIG. 15. Band structure of bcc Mo with $\{s,d\}$ QO basis. (Circle dot: plane-wave DFT calculation; solid line: TB calculation based on six QOs; dashed line: Fermi energy; energy threshold with $\varepsilon_{\rm th}=0$ eV.)

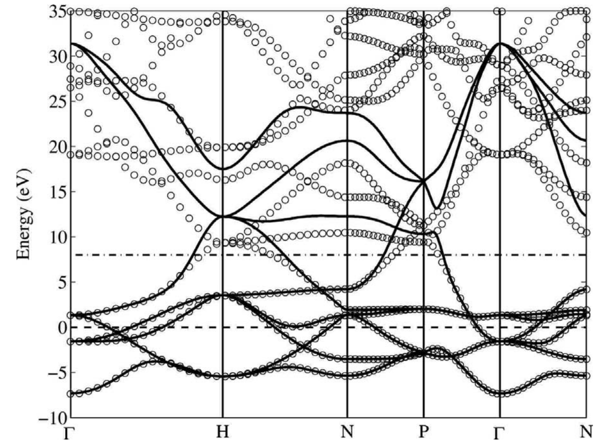

FIG. 16. Band structure of bcc Mo with $\{s,p,d\}$ QO basis. (Circle dot: plane-wave DFT calculation; solid line: TB calculation based on nine QOs; dashed line: Fermi energy; and dash-dot line: energy threshold with $\varepsilon_{\rm th}=8$ eV.)

renormalized. The corresponding tight-binding band structures are presented in Figs. 15 and 16. Although the band structure using the $\{s,d\}$ QO basis is very smooth as shown in Fig. 15, we still observe a strong deviation around N below ε_F. But in Fig. 16 the band structure with $\{s,p,d\}$ QO basis agrees with the DFT result very well, especially for those problematic bands around point N. This indicates that the p component may play an important role around N.

We then use VASP to perform AO-projected band-structure analysis as shown in the color-encoded plot, Fig. 17(a), where the specific color is from the linear weight of d, s, and p components corresponding to red, green, and blue, respectively, as shown in the color triangle of Fig. 17(c). We can immediately see that around point N those Kohn-Sham bands crossing the Fermi energy ε_F have strong blue and red components corresponding to the p and d characters. In contrast we do not find clear s component in these bands. This is very crucial since we were expecting the $\{s,d\}$ QOs as the minimal-basis set for bcc Mo; however due to this strong p component around N the $\{s,d\}$ QOs are not enough to pre-

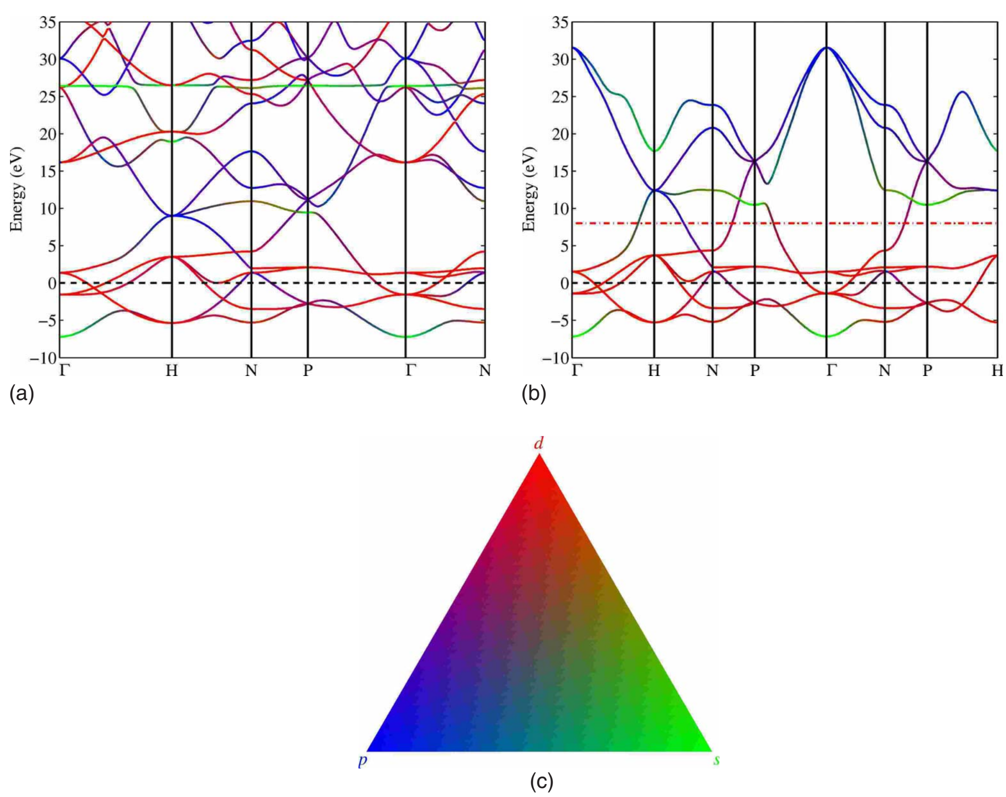

FIG. 17. (Color online) (a) AO-projected band structure of bcc Mo with $\{s,p,d\}$ QO basis; (b) QO-projected tight-binding band structure of bcc Mo with $\{s,p,d\}$ QO basis; (c) color triangle: red for d orbitals, green for s orbital, and blue for p orbitals.

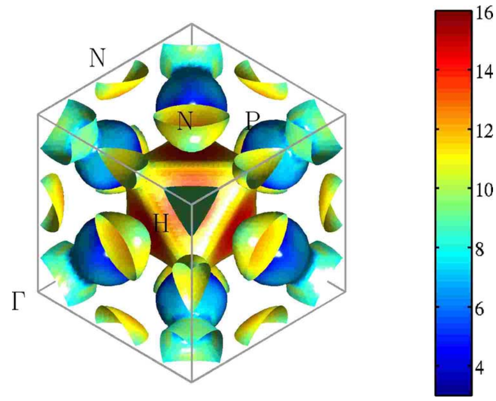

FIG. 18. (Color online) The Fermi-velocity-encoded Fermi surface of bcc Mo with $\{s,p,d\}$ QO basis in the reciprocal cell. The velocity is in the unit of Å/fs.

serve the full DFT band structure below the energy threshold accurately, thus give rise to the strong deviations in both Fig. 3 of Ref. 24 and Fig. 15. Figure 17(b) shows the color-encoded QO-projected tight-binding band structure with $\{s,p,d\}$ QO basis set and it preserves the general distribution of AO components in the band structure. Therefore, the minimal basis for bcc Mo should be the $\{s,p,d\}$ QOs.

With this $\{s,p,d\}$ QO basis set we have calculated the high-resolution Fermi surface of bcc Mo using a dense $32\times32\times32$ grid. Here in Fig. 18 we show the Fermi-velocity-encoded Fermi surface where the magnitude of velocity $|\mathbf{v}_F|$ is represented by different colors defined in the color bar. Fermi velocity is calculated from $\mathbf{v}_F = dE(\mathbf{k})/\hbar d\mathbf{k}$. It should be mentioned that Fig. 18 displays the Fermi surface in reciprocal cell, instead of the first Brillouin zone. Thus high-symmetry points Γ, H, N, and P are located at the corner, the center of the cell, the middle of surfaces and edges, and the center of equilateral triangles on the surfaces, respectively. From our calculation the minimal and maximal magnitudes of Fermi velocity of bcc Mo are 3.36 and 15.02 Å/fs. Obviously the magnitude of Fermi velocity is very different on different sheets of the Fermi surface. The central octahedral surface around point H encloses *holes* which have higher velocity than the electrons or holes on the other sheets. This is also clearly reflected by the large slope of the Kohn-Sham bands crossing ε_F at both P-H and H-N in Fig. 17(b). In contrast Fermi electrons in the other bands at Γ-H, Γ-N, and Γ-P have smaller velocity showing blue color in Fig. 18.

VI. COMPARISON BETWEEN QO AND OTHER LOCALIZED ORBITALS

A. Comparison between QO and MLWF

MLWF developed by Marzari and Vanderbilt[32] is the most localized orthogonal Wannier function, and it could achieve even better localization if the orthogonality condition is relaxed, which is an advantage compared to QO. In general both the center and shape of MLWF are unknown before the construction is fully finished. It could be atomic-orbital- or bonding-orbital-like, which is determined by the information included in the selected Bloch subspace. In contrast, the center and pseudoangular momentum of QO are known *before* the construction. Algorithmically, QO is a noniterative projection-based scheme, whereas MLWF is based on nonlinear optimization and needs to search for the global minimum iteratively. Due to the nonlinear nature of the MLWF scheme, the selection of Bloch subspace is of utmost importance, whereas the present QO scheme represents infinite band result cheaply, and therefore might be simpler to use. The maximal similarity and pseudoangular momentum of QO also allow for easier labeling and interpretation. From another point of view, QO method may (a) give an upper bound of the energy of the highest unoccupied Bloch states one need to include in the MLWF scheme in order to obtain a set of atomic-orbital-like MLWFs, (b) provide a simple way to disentangle the Bloch wave functions in solids, and (c) perform as a good initial guess for MLWFs as well.

B. Comparison between QO and QUAMBO

The original QUAMBO method[20–24] selects the optimized combination subspace $\mathcal{C}(\mathbf{k})$ from the large unoccupied Bloch subspace $\overline{\mathcal{R}}(\mathbf{k})$. This method is also implemented in our code.[27] In the Appendix we rigorously prove that QO is equivalent to QUAMBO in the infinite band limit. However, practically with QUAMBO method one needs to include enough Kohn-Sham bands to capture all bonding and antibonding Bloch states for construction of the corresponding quasiatomic orbitals. It is difficult to predict where the corresponding highest antibonding Bloch state is. Even if it is predictable, those antibonding states, unfortunately, are often pushed to very high energies. There could be hundreds of Bloch states between the bonding and antibonding Bloch states, which are irrelevant to the construction of QUAMBO. In conventional DFT calculations it is very inefficient to calculate and very memory consuming to store a large number of bands. In the QUAMBO method most of time could be wasted on calculating atomic projections on these irrelevant bands. The alternative QO construction is totally independent of unoccupied Bloch eigenstates since one *directly* constructs the optimized combination Bloch states and the only additional cost is non-self-consistent evaluation of Hamiltonian matrix elements between them.

The theoretical basis for QO and QUAMBO method is the idea of Slater and Koster[16] of linear combination of atomic orbitals (LCAO), thus the localization of QO and QUAMBO depends on whether the specific material can be well described by the LCAO idea for the low-energy chemistry. As long as the idea of LCAO works for the materials one is interested in, the low-energy bands should be dominated by quantum numbers of atomic orbitals (antibonding Bloch states are usually smeared out among the unoccupied Bloch subspace, but they are not far from Fermi level). Meanwhile, by definition QO is maximally similar to AO; therefore the quasiangular quantum numbers should be still preserved while the radial part and the detailed local shape of QO largely depend on the bonding nature of QO with other orbitals on its neighboring atoms. Practically speaking, the pseudoatomic orbitals from pseudopotential generators have already provided us the clue about the relevant angular quantum numbers. As long as density-functional theory with these pseudopotentials can describe the specific material well, we can always obtain localized QOs which can accurately describe the electronic structure below a few eV above the Fermi energy by forming the bonding-antibonding closure. For higher energy regions, we may have to include additional radial quantum numbers for *s*-state, *p*-states, etc. And certainly, it would be difficult for QO to describe unbound electron states.

C. Comparison between QO and PAO

The construction of optimized combination subspace from atomic-orbital Bloch subspace in QO scheme is similar to the PAO scheme of Sæbø and Pulay,[66–68] which has been widely used in quantum chemistry. However our QO scheme is applicable to molecules, surfaces, and solids within one program, enabling the construction of transferable local basis functions and comparison of bonding chemistry from molecules to surfaces to solids. It can be embedded in or interfaced to any DFT package using plane-wave, Gaussian, or mixed bases. As we have shown in the various applications above, QO can be constructed not only for insulators and semiconductors but also for metallic systems. Another difference is that we use the pseudized atomic orbitals as the similarity objects with less nodes in their wave functions. Moreover without considering the core wave functions we have much less number of basis orbitals to construct and diagonalize in *ab initio* TB calculations. QO is a true minimal basis scheme, and consequently we can efficiently perform TB analysis and parametrizations.

VII. SUMMARY

Quasiatomic orbital is derived and implemented for different types of materials. The accuracy, efficiency, and robustness of QO for *ab initio* tight-binding analysis are demonstrated through band structure, density of states, QO-projected density of states, the Fermi surface, the Mulliken charge, and bond order analysis. We have shown that QO is equivalent to the infinite band limit of QUAMBO without the need to explicitly compute and store a large number of unoccupied Bloch wave functions. Furthermore, the most important property of QO is that it retains all electronic structure information below a certain energy threshold while possessing both quasiangular momentum quantum number and reasonably good localization, which fulfills the true spirit of the LCAO of Slater and Koster.[16] Therefore, QO may be used as a transferable local basis for the calculations of total energy, electrical conductance, and the development of linear-scaling DFT. For ease of checking, all source codes and relevant data used in this paper are put at a permanent website.[27]

ACKNOWLEDGMENTS

We thank Clemens Först and Wencai Lu for valuable discussions and suggestions. Work of J.L. and L.Q. was supported by Honda Research Institute USA, Inc., NSF under Grant No. DMR-0502711 and USAFOSR. Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No. DE-AC02-07CH11358. This work was supported by the Director for Energy Research, Office of Basic Energy Sciences including a grant of computer time at the National Energy Research Supercomputing Center (NERSC) in Berkeley.

APPENDIX: MATRIX ALGEBRA PROOF OF QO EQUIVALENCE TO QUAMBO IN THE INFINITE BAND LIMIT

To prove the equivalence between QO and QUAMBO in the infinite band limit, we first expand matrix element $(\mathbf{W_k})_{Ii,Jj}$ in Eq. (33) as the following:

$$\begin{aligned}(\mathbf{W_k})_{Ii,Jj} &= \langle A_{Ii}|\left(\sum_{\bar{n}} \hat{P}^{\dagger}_{\psi_{\bar{n}\mathbf{k}}}\right)\hat{S}\left(\sum_{\bar{m}} \hat{P}_{\psi_{\bar{m}\mathbf{k}}}\right)|A_{Jj}\rangle \\ &= \langle A_{Ii}|\hat{S}\left(\sum_{\bar{n}} \hat{P}_{\psi_{\bar{n}\mathbf{k}}}\right)|A_{Jj}\rangle \\ &= \sum_{\bar{n}} \langle A_{Ii}|\hat{S}|\psi_{\bar{n}\mathbf{k}}\rangle\langle\psi_{\bar{n}\mathbf{k}}|\hat{S}|A_{Jj}\rangle. \end{aligned} \tag{A1}$$

We use $(\mathbf{M_k})_{\bar{n},Jj}$ to represent the matrix element $\langle\psi_{\bar{n}\mathbf{k}}|\hat{S}|A_{Jj}\rangle$. Then we will have the simple form of $\mathbf{W_k}$ for QO,

$$\mathbf{W_k} = \mathbf{M_k^{\dagger}M_k}, \tag{A2}$$

where the size of $\mathbf{W_k}$ and $\mathbf{M_k}$ is $qN\times qN$ and dim $\overline{\mathcal{R}}(\mathbf{k})\times qN$ (or $\infty\times qN$), respectively. However, in the original QUAMBO method of Lu *et al.*[22] in the limit of infinite bands the overlap matrix $\widetilde{\mathbf{W}}_\mathbf{k}$ is defined as

$$\widetilde{\mathbf{W}}_\mathbf{k} = \mathbf{M_k M_k^{\dagger}}, \tag{A3}$$

where the size of $\widetilde{\mathbf{W}}_\mathbf{k}$ is $\infty\times\infty$. $\mathbf{W_k}$ and $\widetilde{\mathbf{W}}_\mathbf{k}$ are the so-called Gramian matrix. We then perform singular value decomposition (SVD) of matrix $\mathbf{M_k}$,

$$\mathbf{M_k} = \mathbf{U_k \Sigma_k V_k^{\dagger}}, \tag{A4}$$

where $\mathbf{U_k}$ and $\mathbf{V_k}$ are the unitary transformation matrices with the sizes of $\infty\times\infty$ and $qN\times qN$, respectively, and they satisfy $\mathbf{U_k^{\dagger}U_k=I}$ and $\mathbf{V_k^{\dagger}V_k=I}$. Matrix $\mathbf{\Sigma_k}$ with the size of $\infty\times qN$ contains the singular values, and it has N_M nonzero values, where $N_M\le\min\{qN,\infty\}=qN$. Thus, $\mathbf{W_k}=\mathbf{V_k\Sigma_k^{\dagger}\Sigma_k V_k^{\dagger}}$ and $\widetilde{\mathbf{W}}_\mathbf{k}=\mathbf{U_k\Sigma_k\Sigma_k^{\dagger}U_k^{\dagger}}$. Let $\mathbf{Y_k}=\mathbf{\Sigma_k^{\dagger}\Sigma_k}$ and $\widetilde{\mathbf{Y}}_\mathbf{k}=\mathbf{\Sigma_k\Sigma_k^{\dagger}}$. Both $\mathbf{Y_k}$ and $\widetilde{\mathbf{Y}}_\mathbf{k}$ are the diagonal matrices with the sizes of $qN\times qN$ and $\infty\times\infty$, respectively; however they contain exactly the same N_M positive eigenvalues. It immediately leads to three conclusions: (a) $\mathbf{W_k}$ and $\widetilde{\mathbf{W}}_\mathbf{k}$ have the same rank as $\mathbf{M_k}$; (b) $\mathbf{W_k}$ and $\widetilde{\mathbf{W}}_\mathbf{k}$ share the same eigenvalues; and (c) $\mathbf{V_k}$ and $\mathbf{U_k}$ contain the corresponding eigenvectors of $\mathbf{W_k}$ and $\widetilde{\mathbf{W}}_\mathbf{k}$, respectively. We then have

$$\widetilde{\mathbf{W}}_\mathbf{k}\mathbf{M_k V_k} = \mathbf{M_k W_k V_k} = \mathbf{M_k V_k Y_k}, \tag{A5}$$

which means corresponding to the ith positive eigenvalue, the ith eigenvector $(\mathbf{U_k})_i$ of $\widetilde{\mathbf{W}}_\mathbf{k}$ is the ith vector $(\mathbf{M_k V_k})_i$ multiplied by a factor $(\mathbf{\Lambda_k})_{ii}$,

$$(\mathbf{U_k})_i = (\mathbf{\Lambda_k})_{ii}(\mathbf{M_k V_k})_i. \tag{A6}$$

$\mathbf{\Lambda_k}$ is a diagonal matrix with the size of $qN \times qN$. Since $\mathbf{U_k}$ is a unitary matrix,

$$\mathbf{I}_{qN\times qN} = (\mathbf{U_k^\dagger U_k})_{qN\times qN} = \mathbf{\Lambda_k^\dagger \Sigma_k^\dagger \Sigma_k \Lambda_k}, \tag{A7}$$

which leads to $(\mathbf{\Lambda_k})_{ii} = (\mathbf{\Sigma_k^\dagger \Sigma_k})_{ii}^{-1/2} = (\mathbf{Y_k})_{ii}^{-1/2}$ and thus $(\mathbf{U_k})_i = (\Lambda_\mathbf{k})_{ii}(\mathbf{M_k V_k})_i$ corresponding to the ith positive eigenvalue $(\mathbf{Y_k})_{ii}$. Finally, in the original QUAMBO method[22] the $C_\mathbf{k}$ eigenvectors associated with the largest $C_\mathbf{k}$ eigenvalues of $\tilde{\mathbf{W}}_\mathbf{k}$ are selected to form the optimized combination subspace $\mathcal{C}(\mathbf{k})$. Therefore, the optimized combination state $|\tilde{c}_{m\mathbf{k}}\rangle$ can be expanded as the following:

$$\begin{aligned}
|\tilde{c}_{m\mathbf{k}}\rangle &= \sum_{\bar{n}} (\mathbf{U_k})_{\bar{n},m}|\psi_{\bar{n}\mathbf{k}}\rangle \\
&= \sum_{\bar{n},Ii} (\mathbf{\Lambda_k})_{mm}(\mathbf{M_k})_{\bar{n},Ii}(\mathbf{V_k})_{Ii,m}|\psi_{\bar{n}\mathbf{k}}\rangle \\
&= \sum_{\bar{n},Ii} (\mathbf{\Lambda_k})_{mm}(\mathbf{V_k})_{Ii,m}|\psi_{\bar{n}\mathbf{k}}\rangle\langle\psi_{\bar{n}\mathbf{k}}|\hat{S}|A_{Ii}\rangle \\
&= \sum_{Ii} (\mathbf{\Lambda_k})_{mm}(\mathbf{V_k})_{Ii,m}(\hat{I}_\mathbf{k} - \sum_n \hat{P}_{\psi_{\bar{n}\mathbf{k}}})|A_{Ii}\rangle \\
&= \sum_{Ii} (\mathbf{\Lambda_k})_{mm}(\mathbf{V_k})_{Ii,m}|A^\perp_{Ii,\mathbf{k}}\rangle = |c_{m\mathbf{k}}\rangle.
\end{aligned} \tag{A8}$$

Therefore, in the end we have $|\tilde{c}_{m\mathbf{k}}\rangle = |c_{m\mathbf{k}}\rangle$. This means that the selected $C_\mathbf{k}$ eigenvectors associated with the largest $C_\mathbf{k}$ eigenvalues of $\tilde{\mathbf{W}}_\mathbf{k}$ in the QUAMBO method in the limit of infinite bands are exactly the same as those associated with the largest $C_\mathbf{k}$ eigenvalues of $\mathbf{W_k}$ in the QO method. The above proof shows that although $\mathbf{W_k}$ and $\tilde{\mathbf{W}}_\mathbf{k}$ defined for QO and QUAMBO are different, in the infinite band limit both matrices have exactly the same positive eigenvalues, leading to the same optimized combination subspace $\mathcal{C}(\mathbf{k})$. More importantly, by using the definition of identity operator we only need the finite occupied Bloch subspace $\mathcal{R}(\mathbf{k})$ for the construction of QO, while the construction of QUAMBO requires infinite unoccupied Bloch subspace $\overline{\mathcal{R}}(\mathbf{k})$ to reach the same result as QO. As shown in Eq. (41), the only additional but little cost is to evaluate Hamiltonian matrix elements between any two of the directly constructed finite $\{c_{m\mathbf{k}}\}$.

In practical implementations "infinite bands" refer to full occupied and unoccupied Bloch space defined on particular basis. For example, in plane-wave DFT calculations we use large but finite plane waves as the basis. Therefore, at each $\mathbf{k}$ point the dimension of full Bloch space or infinite bands is the total number of plane waves. In practice when using the original QUAMBO scheme we have to truncate unoccupied Bloch space due to the limited computational power and memory, which leads to different eigenvalues and different optimized combination subspace $\mathcal{C}(\mathbf{k})$ compared to the QO method. The above truncation could give rise to the finite UBTE problem discussed in the beginning of this work. The situation will be even worse when we apply the QUAMBO method in strongly confined systems where particular antibonding Bloch bands will be pushed up to very high energy and cannot be captured in finite unoccupied Bloch subspace. Then the rank of $\mathbf{U_k}$ will be smaller than $qN - R_\mathbf{k}$, leading to the incomplete optimized combination subspace $\{\tilde{c}_{m\mathbf{k}}\}$ and consequently the singularity of TB Hamiltonian under the QUAMBO basis set. The QO method does not suffer from this UBTE.

*Author to whom correspondence should be addressed; liju@seas.upenn.edu

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[3] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).
[4] K. Laasonen, R. Car, C. Lee, and D. Vanderbilt, Phys. Rev. B **43**, 6796 (1991).
[5] K. Laasonen, A. Pasquarello, R. Car, C. Lee, and D. Vanderbilt, Phys. Rev. B **47**, 10142 (1993).
[6] G. Kresse and J. Furthmuller, Phys. Rev. B **54**, 11169 (1996).
[7] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[8] M. S. Tang, C. Z. Wang, C. T. Chan, and K. M. Ho, Phys. Rev. B **53**, 979 (1996).
[9] H. Haas, C. Z. Wang, M. Fahnle, C. Elsasser, and K. M. Ho, Phys. Rev. B **57**, 1461 (1998).
[10] C. Z. Wang, B. C. Pan, and K. M. Ho, J. Phys.: Condens. Matter **11**, 2043 (1999).
[11] J. Li, D. Y. Liao, S. Yip, R. Najafabadi, and L. Ecker, J. Appl. Phys. **93**, 9072 (2003).
[12] *Handbook of Materials Modeling*, edited by S. Yip (Springer, Dordrecht, 2005).
[13] W. A. Harrison, *Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond* (Freeman, San Francisco, 1980).
[14] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).
[15] W. J. Hehre, R. F. Stewart, and J. A. Pople, J. Chem. Phys. **51**, 2657 (1969).
[16] J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).
[17] R. Resta, J. Phys.: Condens. Matter **14**, R625 (2002).
[18] O. K. Andersen and T. Saha-Dasgupta, Phys. Rev. B **62**, R16219 (2000).
[19] X. Qian, J. Li, X. Lin, and S. Yip, Phys. Rev. B **73**, 035408 (2006).
[20] W. C. Lu, C. Z. Wang, M. W. Schmidt, L. Bytautas, K. M. Ho, and K. Ruedenberg, J. Chem. Phys. **120**, 2629 (2004).
[21] W. C. Lu, C. Z. Wang, M. W. Schmidt, L. Bytautas, K. M. Ho, and K. Ruedenberg, J. Chem. Phys. **120**, 2638 (2004).
[22] W. C. Lu, C. Z. Wang, T. L. Chan, K. Ruedenberg, and K. M. Ho, Phys. Rev. B **70**, 041101(R) (2004).
[23] W. C. Lu, C. Z. Wang, K. Ruedenberg, and K. M. Ho, Phys. Rev. B **72**, 205123 (2005).
[24] T.-L. Chan, Y. X. Yao, C. Z. Wang, W. C. Lu, J. Li, X. F. Qian, S. Yip, and K. M. Ho, Phys. Rev. B **76**, 205119 (2007).
[25] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
[26] S. R. Bahn and K. W. Jacobsen, Comput. Sci. Eng. **4**, 56 (2002).

[27] http://alum.mit.edu/www/liju99/QO/ (QO).
[28] G. H. Wannier, Phys. Rev. **52**, 191 (1937).
[29] G. F. Koster and J. C. Slater, Phys. Rev. **95**, 1167 (1954).
[30] W. Kohn, Phys. Rev. **115**, 809 (1959).
[31] W. Kohn, Phys. Rev. B **7**, 4388 (1973).
[32] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).
[33] G. Berghold, C. J. Mundy, A. H. Romero, J. Hutter, and M. Parrinello, Phys. Rev. B **61**, 10040 (2000).
[34] I. Souza, N. Marzari, and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2001).
[35] J. J. Mortensen and M. Parrinello, J. Phys.: Condens. Matter **13**, 5731 (2001).
[36] C. M. Zicovich-Wilson, R. Dovesi, and V. R. Saunders, J. Chem. Phys. **115**, 9708 (2001).
[37] K. S. Thygesen, L. B. Hansen, and K. W. Jacobsen, Phys. Rev. Lett. **94**, 026405 (2005).
[38] R. A. Evarestov, D. E. Usvyat, and V. P. Smirnov, Theor. Chem. Acc. **114**, 19 (2005).
[39] U. Birkenheuer and D. Izotov, Phys. Rev. B **71**, 125116 (2005).
[40] J. Bhattacharjee and U. V. Waghmare, Phys. Rev. B **73**, 121102(R) (2006).
[41] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).
[42] A. J. Williamson, R. Q. Hood, and J. C. Grossman, Phys. Rev. Lett. **87**, 246406 (2001).
[43] J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, and D. Sanchez-Portal, J. Phys.: Condens. Matter **14**, 2745 (2002).
[44] R. Iftimie, J. W. Thomas, and M. E. Tuckerman, J. Chem. Phys. **120**, 2169 (2004).
[45] A. Calzolari, N. Marzari, I. Souza, and M. Buongiorno Nardelli, Phys. Rev. B **69**, 035108 (2004).
[46] K. S. Thygesen and K. W. Jacobsen, Chem. Phys. **319**, 111 (2005).
[47] Z. Y. Li and D. S. Kosov, J. Phys.: Condens. Matter **18**, 1347 (2006).
[48] T. Thonhauser, D. Ceresoli, D. Vanderbilt, and R. Resta, Phys. Rev. Lett. **95**, 137205 (2005).
[49] I. Schnell, G. Czycholl, and R. C. Albers, Phys. Rev. B **65**, 075103 (2002).
[50] S. Fabris, S. de Gironcoli, S. Baroni, G. Vicario, and G. Balducci, Phys. Rev. B **71**, 041102(R) (2005).
[51] A. Yamasaki, M. Feldbacher, Y. F. Yang, O. K. Andersen, and K. Held, Phys. Rev. Lett. **96**, 166401 (2006).
[52] E. J. Bylaska, K. Tsemekhman, and F. Gao, Phys. Scr., T **124**, 86 (2006).
[53] L. X. He and D. Vanderbilt, Phys. Rev. Lett. **86**, 5341 (2001).
[54] C. Brouder, G. Panati, M. Calandra, C. Mourougane, and N. Marzari, Phys. Rev. Lett. **98**, 046402 (2007).
[55] J. W. Thomas, R. Iftimie, and M. E. Tuckerman, Phys. Rev. B **69**, 125105 (2004).
[56] R. Iftimie, P. Minary, and M. E. Tuckerman, Proc. Natl. Acad. Sci. U.S.A. **102**, 6654 (2005).
[57] J. M. Foster and S. F. Boys, Rev. Mod. Phys. **32**, 300 (1960).
[58] S. Baroni, A. D. Corso, S. de Gironcoli, P. Giannozzi, C. Cavazzoni, G. Ballabio, S. Scandolo, G. Chiarotti, P. Focher, A. Pasquarello, K. Laasonen, A. Trave, R. Car, N. Marzari, and A. Kokalj, http://www.pwscf.org (PWSCF).
[59] http://www.cpmd.org (CPMD).
[60] P. E. Blöchl, C. J. Först, and J. Schimpl, Bull. Mater. Sci. **26**, 33 (2003).
[61] X. Gonze, J. M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G. M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, Ph. Ghosez, J. Y. Raty, and D. C. Allan, Comput. Mater. Sci. **25**, 478 (2002).
[62] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).
[63] K. Ruedenberg, M. W. Schmidt, M. M. Gilbert, and S. T. Elbert, Chem. Phys. **71**, 41 (1982).
[64] J. C. Slater, J. Chem. Phys. **41**, 3199 (1964).
[65] J. P. Lewis, K. R. Glaesemann, G. A. Voth, J. Fritsch, A. A. Demkov, J. Ortega, and O. F. Sankey, Phys. Rev. B **64**, 195103 (2001).
[66] P. Pulay, Chem. Phys. Lett. **100**, 151 (1983).
[67] S. Sæbø and P. Pulay, Chem. Phys. Lett. **113**, 13 (1985).
[68] S. Sæbø and P. Pulay, Annu. Rev. Phys. Chem. **44**, 213 (1993).
[69] W. Kohn, Phys. Rev. Lett. **2**, 393 (1959).
[70] B. N. Brockhouse, T. Arase, G. Caglioti, K. R. Rao, and A. D. B. Woods, Phys. Rev. **128**, 1099 (1962).
[71] M. Methfessel and A. T. Paxton, Phys. Rev. B **40**, 3616 (1989).
[72] R. S. Mulliken, J. Chem. Phys. **23**, 1833 (1955).
[73] J. Li and S. Yip, Phys. Rev. B **56**, 3524 (1997).
[74] X. Qian, Ph.D. thesis, Massachusetts Institute of Technology, 2008.
[75] http://dcwww.camp.dtu.dk/campos/Dacapo/ (DACAPO).
[76] B. Hammer, L. B. Hansen, and J. K. Nørskov, Phys. Rev. B **59**, 7413 (1999).
[77] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).
[78] http://www.xcrysden.org/ (XCRYSDEN).
[79] A. Kokalj, J. Mol. Graphics Modell. **17**, 176 (1999).
[80] A. Kokalj, Comput. Mater. Sci. **28**, 155 (2003).