JCTC Journal of Chemical Theory and Computation

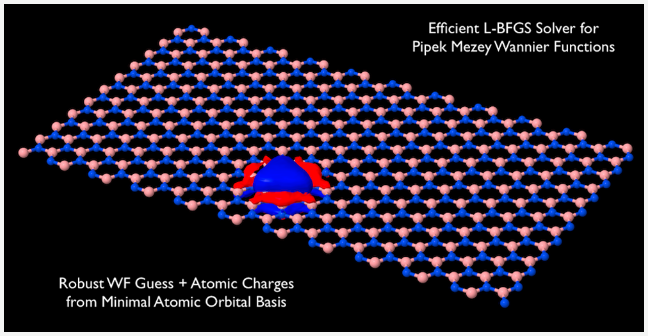
pubs.acs.org/JCTC Article

Robust Pipek–Mezey Orbital Localization in Periodic Solids

Marjory C. Clement, Xiao Wang, and Edward F. Valeev*

Cite This: *J. Chem. Theory Comput.* 2021, 17, 7406–7415 Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information



ABSTRACT: We describe a robust method for determining Pipek–Mezey (PM) Wannier functions (WF), recently introduced by Jónsson et al. (*J. Chem. Theor. Chem.* **2017**, *13*, 460), which provide some formal advantages over the more common Boys (also known as maximally-localized) Wannier functions. The Broyden–Fletcher–Goldfarb–Shanno-based PMWF solver is demonstrated to yield dramatically faster convergence compared to the alternatives (steepest ascent and conjugate gradient) in a variety of one-, two-, and three-dimensional solids (including some with vanishing gaps) and can be used to obtain Wannier functions robustly in supercells with thousands of atoms. Evaluation of the PM functional and its gradient in periodic linear combination of atomic orbital representation used a particularly simple definition of atomic charges obtained by Moore–Penrose pseudoinverse projection onto the minimal atomic orbital basis. An automated "canonicalize phase then randomize" method for generating the initial guess for WFs contributes significantly to the robustness of the solver.

1. INTRODUCTION

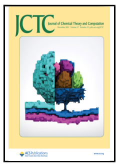
Localized orbitals, by eliminating the artifacts of symmetry and accidental degeneracy, are valuable for qualitative interpretation of electronic states in terms of traditional concepts of chemical bonding and as a computational basis underpinning many reduced complexity algorithms in electronic structure. Localized orbitals are particularly relevant for periodic solids, where due to the lattice symmetry, the eigenstates of observables are delocalized over the entire lattice; such delocalization becomes counterproductive when the unit cell size significantly exceeds the lengthscale of the decay of the 1-particle-reduced density matrix or when the focus is on localized features of the electronic structure (e.g., impurities and surface adsorbates).

For molecules, the most commonly used *black-box* localization methods are those due to Foster and Boys (FB)[1,2] and Pipek and Mezey (PM);[3] several historically important methods like Edmiston–Ruedenberg (ER)[4] and von Niessen (vN)[5] are rarely used nowadays. All of these orbitals are defined as the stationary points of the corresponding functional. For example, the FB orbitals minimize the sum of squared position uncertainties of the orbitals. PM orbitals maximize the sum of squares of atomic charges of each orbital; the original PM definition utilized Mulliken charges, which are meaningless for non-minimal basis sets, and other definitions of atomic charges are far more robust.[6−10] An often cited advantage of the PM orbitals is that, unlike the FB, ER, and vN orbitals, they usually "separate" the σ and π orbitals,[a] which is an important advantage when interpreting the electronic structure[b]. Further enhancements of these functionals include the use of higher-than-second power analogues of the FB[11] and PM[8,9] functionals.

In periodic solids, the analogues of molecular localized orbitals are referred to as *generalized* Wannier functions (WF),[12] which, in contrast to conventional Wannier functions[13] obtained from a single band, mix Bloch orbitals from several bands.[14] Marzari and Vanderbilt championed the use of the FB functional to determine the generalized Wannier functions, which they dubbed as *maximally localized* (generalized) Wannier functions (MLWF).[12] Since the implementation of an MLWF solver in the Wannier90 package,[15] the use of MLWF has become popular[14] for interpreting the electronic states and as a basis in a number of reduced-scaling many-body electronic structure methods.[16,17] Note that the MLWF formalism is not unique; for example, the FB localization of periodic orbitals by Zicovich-Wilson et al.,[18] as implemented in the Crystal package,[19] in contrast to the approach of Marzari and Vanderbilt, uses a purely real-space formulation. Furthermore, the choice of localization functional itself is not unique; MLWFs are just one of many plausible types of MLWF.

Recently, Jónsson et al. introduced the use of Pipek–Mezey Wannier functions (PMWF),[20] which, in contrast to the FB-based WFs of refs 12 and 18 and analogously to the molecular PM counterparts, do not mix the σ and π orbitals. Jónsson and Lehtola evaluated PMWFs using charges defined via real-space

Received: March 8, 2021
Published: November 5, 2021



partitioning; the atomic charges and PMWFs were found to be insensitive to the specifics of the atomic partitioning, just as molecular PM orbitals[9] were found to depend weakly on the choice of atomic charges used in the PM functional. The maximum of the PM functional was located by the conjugate gradient (CG) method, which for PM optimization in molecules was found to be superior to the steepest ascent (SA) method.[21] Unfortunately, the convergence of the CG solver for PMWFs, as documented by Jónsson et al. as well as revealed in our own experiments, can be fairly poor, requiring hundreds or even thousands of iterations. Also note that the CG is used to construct MLWFs in Wannier90.

As we discovered in our work, the performance of the SA and (nonlinear) CG solvers for PMWFs can be greatly improved by the Broyden−Fletcher−Goldfarb−Shanno (BFGS) solver.[22−25] While BFGS[26] as well as other quasi-Newton methods[27,28] have been used to solve for localized orbitals in molecules, to the best of our knowledge, its use for Wannier function optimization has not been considered. Thus, the main purpose of this manuscript is to document the implementation of the BFGS PMWF solver and compare its performance to that of SA and CG. Here, we also document the particularly simple definition of atomic charges that we devised via pseudoinverse projection on the minimal basis, which is very convenient when employing the linear combination of atomic orbitals (LCAO) representation of the periodic solid's orbitals. The result of this work is a robust periodic localizer that has been successfully applied to one-, two-, and three-dimensional systems with large Born−von-Kármán (BvK) unit cells, including some with vanishingly small band gaps.

2. FORMALISM

2.1. Basic Definitions. The objective of PM localization is to convert an input set of periodic (Bloch) orbitals (usually, Hartree−Fock or Kohn−Sham orbitals) to a set of localized orbitals. In our work, the orbitals are expressed in the LCAO representation, expanded in (contracted) Gaussian AOs, and obtained by the reduced-scaling Hartree−Fock method recently reported by some of us.[29] The jth Bloch orbital $|\psi_{j,\mathbf{k}}\rangle$ with crystal momentum wave vector $\mathbf{k}$ is a linear combination of Bloch AOs $\{\phi_{\nu,\mathbf{k}}\}$

$$|\psi_{j,\mathbf{k}}\rangle = \sum_{\nu} C_{\nu,\mathbf{k}}{}^{j}|\phi_{\nu,\mathbf{k}}\rangle \tag{1}$$

Bloch AOs in turn are translation-symmetry-adapted linear combinations of AOs $\{\phi_{\nu,\mathbf{R}}\}$

$$|\phi_{\nu,\mathbf{k}}\rangle = \frac{1}{\sqrt{N}}\sum_{\mathbf{R}} e^{ik\cdot R}|\phi_{\nu,\mathbf{R}}\rangle \tag{2}$$

with $\mathbf{R}$ denoting the origin of each primitive unit cell in the BvK unit cell ("supercell") composed of N primitive cells; by convention, $\mathbf{R} = 0$ corresponds to the *reference* unit cell. Please note the different normalizations of Bloch AOs in eq 2 than is traditional in the periodic LCAO literature.[29−31]

A generalized Wannier function centered in the unit cell at $\mathbf{R}$ will be expressed as a linear combination of Bloch orbitals

$$|\tilde{\psi}_{i,\mathbf{R}}\rangle = \frac{1}{\sqrt{N_k}}\sum_{j,\mathbf{k}} e^{-ik\cdot\mathbf{R}}|\psi_{j,\mathbf{k}}\rangle U_{j,\mathbf{k}}{}^{i} \tag{3}$$

where i indexes the Wannier functions (localizing o Bloch orbitals will produce o WFs) and N_k is the number of $\mathbf{k}$ points in the uniform (Monkhorst−Pack)[32] quadrature used to integrate the first Brillouin zone corresponding to the supercell in eq 2 (in this work, we impose $N_k = N$). Matrices $\{\mathbf{U}_\mathbf{k}\}$, defined as $(\mathbf{U}_\mathbf{k})_j{}^{i} \equiv U_{j,\mathbf{k}}{}^{i}$,[c] are unitary to ensure mutual orthonormality of WFs, both associated with the same unit cell and between WFs associated with different unit cells; this also ensures that WFs span the same space as the Bloch orbitals. No additional constraints, such as realness of the WFs, are imposed.

2.2. PM Functional with Pseudoinverse Minimal Basis Atomic Charges. The PM WFs are stationary points of the PM functional

$$P \equiv \sum_{\mathbf{R}}\sum_{A}\sum_{i} |Q_i{}^{A_\mathbf{R}}|^p \tag{4}$$

where A indexes the atoms in the primitive unit cell, and i indexes the Wannier functions. The atomic charge contribution $Q_i^{A_\mathbf{R}}$ is the charge associated with orbital i on atom A in unit cell $\mathbf{R}$, and the charge exponent, p, is equal to 2 (conventional PM functional) or 4 (fourth-order PM functional[8,9]).

The original work by PM used standard Mulliken charges[3]

$$Q_i{}^{A_\mathbf{R}} = \langle\tilde{\psi}_{i,0}|\hat{P}_{A_\mathbf{R}}|\tilde{\psi}_{i,0}\rangle \tag{5}$$

where

$$\hat{P}_{A_\mathbf{R}} \equiv \sum_{\mu\in A} |\phi_{\mu,\mathbf{R}}\rangle\langle\hat{\phi}_{\mu,\mathbf{R}}| \tag{6}$$

is the projector onto the AOs centered on atom $A_\mathbf{R}$ defined via the biorthogonal AO basis as

$$\langle\hat{\phi}_{\mu,\mathbf{R}}| \equiv \sum_{\nu,\mathbf{R}'} (S^{-1})_{\mu,\mathbf{R}}{}^{\nu,\mathbf{R}'}\langle\phi_{\nu,\mathbf{R}'}| \tag{7}$$

or its Bloch-AO equivalent

$$\langle\hat{\phi}_{\mu,\mathbf{R}}| \equiv \sum_{\nu,\mathbf{k}} (S^{-1})_{\mu,\mathbf{R}}{}^{\nu,\mathbf{k}}\langle\phi_{\nu,\mathbf{k}}| \tag{8}$$

and $\mathbf{S}^{-1}$ denotes the inverse overlap matrix.

The key issue with the Mulliken charges as defined by eq 5 is their ill-defined nature for non-minimal basis sets. This impacts the robustness of the PM localization method by making the PM orbitals sensitive to variations in the orbital basis set (OBS) and geometry. Luckily, as noted by Knizia,[8] Lehtola,[9] and others,[6] alternative charge definitions make the PM functional more robust (the interpretation of the approach in ref 6 as PM with non-Mulliken charge definitions was pointed out in ref 9). To this end, Jónsson et al. utilized real-space partitioning of orbital charge densities to obtain robust atomic charges for use in the PM functional for periodic solids.[20]

In this work, we defined the PM functional using atomic charges evaluated with the help of a pre-defined minimal basis set (MBS) of atomic orbitals. Such projection onto an MBS has been long employed to define OBS-independent atomic partitioning of densities and thus, due to its basis independent (intrinsic) nature, eliminates the undue sensitivity of charges to the orbital basis. The use of such MBS-projected charges, in the context of PM localization, has been frequently employed recently.[8−10,33d]

There is no unique method to use an MBS to define atomic charges, due to the arbitrariness of how to partition orbitals and/or their charge densities into atomic components. It is also far too easy to break desired invariants, such as the equality of the sum of atomic populations and the number of electrons. For example, replacing OBS AO $\phi_{\mu,\mathbf{R}}$ by MBS AO $\bar{\phi}_{\mu,\mathbf{R}}$ in eq 6 (such replacement is only meaningful if the overlap inverse is defined, i.e., if the MBS is a subset of the OBS) violates such an invariant. Thus, existing approaches[8,10,33,35,36] utilize the MBS indirectly, as a way to define an MBS-like subspace of the OBS. Here, we use a simpler approach. Please note that, throughout the following discussion, quantities with an overbar are expressed in the MBS AO representation.

Consider the OBS AO representation of the Bloch orbitals, obtained by plugging eqs 2 into 1

$$|\psi_{i,\mathbf{k}}\rangle = \sum_{\mu,\mathbf{R}} C_{\mu,\mathbf{R}}{}^{i,\mathbf{k}} |\phi_{\mu,\mathbf{R}}\rangle \tag{9}$$

with $C_{\mu,\mathbf{R}}{}^{i,\mathbf{k}}$ defined as

$$C_{\mu,\mathbf{R}}{}^{i,\mathbf{k}} \equiv \frac{1}{\sqrt{N}} C_{\mu,\mathbf{k}}{}^{i} e^{ik\cdot\mathbf{R}} \tag{10}$$

clearly, $C_{\nu,\mathbf{R}}{}^{j,\mathbf{k}} = C_{\nu,0}{}^{j,\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}}$. *Biorthogonal* mapping of Bloch functions on the MBS

$$|\overline{\psi_{i,\mathbf{k}}}\rangle = \sum_{\mu,\mathbf{R}} \bar{C}_{\mu,\mathbf{R}}{}^{i,\mathbf{k}} |\bar{\phi}_{\mu,\mathbf{R}}\rangle \tag{11}$$

is trivially obtained by solving

$$\delta_{ij} = \langle \psi_{i,\mathbf{k}} | \overline{\psi_{j,\mathbf{k}}} \rangle = \sum_{\mu,\mathbf{R}} \langle \psi_{j,\mathbf{k}} | \bar{\phi}_{\mu,\mathbf{R}} \rangle \bar{C}_{\mu,\mathbf{R}}{}^{j,\mathbf{k}} \tag{12}$$

for coefficients $\bar{C}_{\mu,\mathbf{R}}{}^{j,\mathbf{k}}$. A least-squares solution is produced by the Moore−Penrose pseudoinverse of the overlap matrix between the target set of Bloch orbitals for the given $\mathbf{k}$ with the MBS AOs in the reference cell, $\bar{\phi}_{\mu,\mathbf{0}}$

$$(\mathbf{S})_{i,\mathbf{k}}{}^{\mu} \equiv \langle \psi_{i,\mathbf{k}} | \bar{\phi}_{\mu,0} \rangle = \sum_{\nu,\mathbf{R}} (C_{\nu,\mathbf{R}}{}^{i,\mathbf{k}})^{*} \langle \phi_{\nu,\mathbf{R}} | \bar{\phi}_{\mu,0} \rangle \tag{13}$$

Only simple Gaussian AO overlaps are needed to evaluate eq 13, with the lattice sum in eq 13 geometrically convergent. Evaluation in other numerical representations, such as plane waves (PWs), should be also straightforward. Extension of such *pseudoinverse* MBS mapping to the molecular case is obvious, where such procedure is related to how the *corresponding* orbitals are constructed.[37] Note also that the pseudoinverse-mapped orbitals are not orthonormal; hence, the pseudoinverse charges differ from the existing definitions for minimal-basis-derived charges;[8,10,33,35,36] however, the relationship between the pseudoinverse MBS charges and other MBS-based charges is outside of the scope of this article and will be discussed elsewhere.

The OBS AO coefficients of the WF

$$W_{\mu,\mathbf{R}}{}^{i} \equiv \frac{1}{\sqrt{N_k}} \sum_{j,\mathbf{k}} C_{\mu,\mathbf{R}}{}^{j,\mathbf{k}} U_{j,\mathbf{k}}{}^{i} \tag{14}$$

are mapped straightforwardly to the MBS AO coefficients

$$\bar{W}_{\mu,\mathbf{R}}{}^{i} \equiv \frac{1}{\sqrt{N_k}} \sum_{j,\mathbf{k}} \bar{C}_{\mu,\mathbf{R}}{}^{j,\mathbf{k}} U_{j,\mathbf{k}}{}^{i} \tag{15}$$

The minimal-basis pseudoinverse charges are thus obtained by replacing the OBS with the MBS in eq 5 and replacing the corresponding OBS AO coefficient of the WF, $\langle \bar{\phi}_{\mu,\mathbf{R}} \| \tilde{\psi}_{i,\mathbf{0}} \rangle \equiv W_{\mu,\mathbf{R}}{}^{i}$, with the corresponding MBS WF AO coefficient

$$\bar{Q}_{i}{}^{A_{\mathbf{R}}} = \frac{1}{2}\left[\sum_{\mu\in\mathbf{A}} \langle \tilde{\psi}_{i,0} \| \bar{\phi}_{\mu,\mathbf{R}} \rangle \bar{W}_{\mu,\mathbf{R}}{}^{i} + \text{h. c.}\right] \tag{16}$$

where "h.c." denotes the Hermitian conjugate. Evaluation of these charges in the LCAO representation leverages the Bloch-MBS overlaps (eq 13) and is therefore completely straightforward. For a fixed MBS, these charges are expected to depend weakly on the OBS and have a well-defined basis set limit.

To avoid introducing a new symbol, P will henceforth denote the PM functional (eq 4) defined with the MBS pseudoinverse charges.

2.3. PM Functional Maximization. *2.3.1. Initial WF Guess.* It is easy to see that the WFs, defined by the unitary matrices $\{\mathbf{U}_{\mathbf{k}}\}$ in eq 3, are in general not uniquely defined by the corresponding functional. The nonuniqueness stems from several factors. First, PM and other functionals defining WFs are invariant with respect to arbitrary permutations of the sequence of WFs. This may appear trivial, but in general, it means that comparing sets of WFs produced in two separate computations is not straightforward (see e.g. ref 20). Second, the WF functional is invariant with respect to all or some of the geometric transformations of the space group of the crystal (such as shifting a WF by a lattice vector). Third, the functionals defining WFs routinely have multiple maxima for a given system; hence, finding the global maximum is NP-hard. Thus, WF computation relies on heuristics to generate initial guesses for $\{\mathbf{U}_{\mathbf{k}}\}$; this initial guess and other solver details determine which functional maximum will be located.

Initial guesses for generalized WFs are typically obtained by projecting Bloch orbitals onto some trial functions. For example, Marzari and Vanderbilt utilized Gaussians located at expected centers of charge of Wannier functions, such as midbond centers;[12] such user-controlled guess construction is also utilized by Wannier90.[15] A more automated approach was used by Zicovich-Wilson et al.,[18] who approximately projected Bloch orbitals onto the reference cell's AO basis (since they expanded the Bloch orbitals in an AO basis already, such projection was trivial); unfortunately, such a choice is not appropriate when covalent bonds cross boundaries of the unit cell. A similar approach was used by Mustafa et al.,[38] who projected Bloch orbitals (expanded in PW basis) onto an appropriate set of AOs that spanned a space containing the target Wannier set. To account for the covalent bonds crossing the unit cell boundaries, the projection AO set must include AOs not only in the reference cell but also in its nearest periodic images.

Note that these projection-based approaches are still not robust enough to deal with multiple minima, since it is necessary to generate multiple initial guesses to probe the global optimality of the resulting WFs. Thus, Jónsson et al.[20] simply generated guess WFs using randomly-generated $\{\mathbf{U}_{\mathbf{k}}\}$ and ran multiple computations.

Here, we have devised a novel automated "canonicalize phase then randomize" (CPR) method for generating guess WFs that can be applied to arbitrary Bloch orbitals expressed in LCAO and non-LCAO representations. The first step in this method is motivated by the realization that, to produce localized Wannier functions even for bands composed of a single atomic orbital (e.g., core bands), it is helpful to canonicalize the phases of the AO coefficients at different $\mathbf{k}$ points. Such phase canonicalization can be viewed as removing the gauge freedom of the Bloch orbitals; once the arbitrariness of the gauge is removed, then the original Wannier prescription,[13] eq 3 with $\{\mathbf{U}_{\mathbf{k}}\}$ set to identity, will recover the maximally localized state, namely, the atomic orbital in the reference cell. Of course, the generalized Wannier functions can compensate for the gauge freedom of the Bloch orbitals via the **k**-dependence of $\{\mathbf{U}_\mathbf{k}\}$; thus, the generalized Wannier orbital for a single-AO band with arbitrary gauge transformation will still be a single AO (although it will not necessarily be the reference cell AO). However, by including the phase canonicalization of the input Bloch orbitals, the WF functional maximization becomes more robust by starting from a good initial guess.

Phase canonicalization in the CPR method proceeds as follows:

- The set of Bloch orbitals at the Γ point ($\mathbf{k} = \mathbf{0}$) is split into degenerate subsets (bands). A set of orbitals is considered degenerate if its eigenenergies are within a prescribed energy tolerance ϵ.
- For each band α, the *phase-defining* subset of AOs, $\{\mu_\alpha\}$, includes AOs μ that have the largest occupancies in the band:

$$\rho_{\mu,0}{}^{\alpha,0} \equiv \sum_{i_\alpha} |C_{\mu,0}{}^{i_\alpha,0}|^2 \tag{17}$$

where i_α are the Bloch orbitals in band α. The phase-defining AO set, $\{\mu_\alpha\}$, thus includes the AOs with the greatest contribution to the band; it can include a single AO (most common) or multiple AOs (due to geometric symmetry and band degeneracy).

- The phase of every Bloch orbital, i, at the Γ point is aligned so that the coefficient of the *first* phase-defining AO, $\mu_{\alpha i}^{(0)}$, for its band, α_i, is positive

$$C_{\mu,0}{}^{i} \to C_{\mu,0}{}^{i} \times \left(\frac{|C_{\mu_{\alpha_i}^{(0)},0}{}^{i}|}{C_{\mu_{\alpha_i}^{(0)},0}{}^{i}} \right) \tag{18}$$

This step eliminates possible arbitrary phase factors that were introduced by the SCF solver and makes all AO coefficients at the Γ point real.

- Bloch orbitals at every **k** point are next mapped to the matching orbital at the Γ point. To perform such matching, consider orbitals i and j at two neighboring points **k** and **k**′, respectively. Their reference cell MBS "overlap" is defined as the overlap of the Bloch orbital $\psi_{i,\mathbf{k}}$ with the Bloch orbital $\psi_{j,\mathbf{k}'}$ pseudoinverse-projected (see Section 2.2) onto the reference cell's MBS AO basis:

$$\bar{S}_{j,\mathbf{k}'}{}^{i,\mathbf{k}} \equiv \sum_{\mu} \langle \psi_{i,\mathbf{k}} || \bar{\phi}_{\mu,0} \rangle \bar{C}_{\mu,0}{}^{j,\mathbf{k}'} \tag{19}$$

All matrix elements in this equation were already evaluated when computing the MBS AO projections of the Bloch orbitals for the purpose of computing atomic charges. For every $i \in [0, o)$, where o is the number of Bloch orbitals being localized, $\psi_{i,\mathbf{k}}$ *matches* orbital $\psi_{j,\mathbf{k}'}$ point if $|\bar{S}_{j,\mathbf{k}'}{}^{i,\mathbf{k}}|$ is the largest among all j; if the match candidate $\psi_{j,\mathbf{k}'}$ has been declared a match for a another orbital $i' < i$, the next best candidate is chosen. If orbital $\psi_{i,\mathbf{k}}$ was matched to orbital $\psi_{j,\mathbf{k}'}$, its phase is canonicalized such that its reference-cell MBS overlap is real

$$C_{\mu,\mathbf{k}}{}^{i} \to C_{\mu,\mathbf{k}}{}^{i} \times \left(\frac{|\bar{S}_{j,\mathbf{k}'}{}^{i,\mathbf{k}}|}{\bar{S}_{j,\mathbf{k}'}{}^{i,\mathbf{k}}} \right) \tag{20}$$

Since the goal is to align the phases of bands at all **k** points to the bands at the Γ point, the band matching is performed using sequences of **k** points originating from the Γ point that span the entire first Brillouin zone mesh. Assuming that points in a uniform mesh of **k** points are indexed by triplets (i, j, k), with $i, j, k = 0, \pm1, \pm2, ..., \pm\lfloor N_k/2 \rfloor$, for one-dimensional structures, bands at the mesh points (1, 0, 0) and (−1, 0, 0) are matched to the canonicalized bands at the (0,0,0) (Γ point), and then bands at mesh points (2, 0, 0) and (−2, 0, 0) are matched to the canonicalized bands at (1, 0, 0) and (−1, 0, 0), respectively, and so on. Similarly, for a two-dimensional structure, the bands at (i, 1, 0) and (i, −1, 0) are matched to the canonicalized bands at (i, 0, 0) and so forth.

- To account for band crossings, bands at point **k** are sorted to appear in the same order as their matching bands at point **k**′.

Despite its relative simplicity, the phase canonicalization is fairly robust and significantly improves the quality of the trial Wannier functions (see Supporting Information). However, the described algorithm is not perfect: (1) it requires a dense Brillouin zone mesh to track high-dispersion bands across the Brillouin zone reliably, (2) it relies on an ad hoc way of matching bands, and (3) it does not account for arbitrary rotations among the degenerate bands. Work to address these shortcomings is underway and will be presented elsewhere.

Performing the phase canonicalization ensures that using identity for $\{\mathbf{U}_\mathbf{k}\}$ produces well-localized WFs for many bands. Note, however, that the "intra-cell" localization is not assisted by the phase canonicalization; thus, it alone will not be sufficient for systems with large unit cells. To be able to locate the global maxima of the PMWF functional by sampling the initial guesses, by default, we initialize $\{\mathbf{U}_\mathbf{k}\}$ with a (quasi)random unitary matrix generated from a user-supplied seed. Note that the same unitary matrix is used for every **k** in order to preserve the benefit of phase canonicalization.

All computations reported in the manuscript used the CPR guess generated with the same (default) seed value. In the Supporting Information, we report additional computations that, after the phase canonicalization, initialized $\{\mathbf{U}_\mathbf{k}\}$ with identity matrices as well as with random $\{\mathbf{U}_\mathbf{k}\}$ generated nonuniformly across the first Brillouin zone (i.e., a different quasirandom matrix was used for every **k** quadrature point, thereby canceling the benefit of phase canonicalization). The performance of the default CPR and "identity" CPR was found to be similar, whereas using the random nonuniform guess required significantly more iterations to reach convergence; however, the final PM functional value was found to be the same for all initial guesses.

Clearly, since the CPR method is defined intrinsically without any reference to the LCAO representation, it can be utilized in the context of non-LCAO (PW and other) representations as long as the Bloch orbitals can be projected onto the MBS AOs. As discussed above, projection on localized states is common in preparing trial Wannier functions;[12,18,38] the use of a full *minimal* AO basis for projection makes the CPR guess (1) well defined even in the limit of a complete OBS (in contrast to the approach of ref 18) and (2) more black-box by eliminating the need to guess positions or composition of target Wannier functions (in contrast to the approaches of refs 12 and 38).

2.3.2. PM Functional Gradient. To find a maximum of the PM functional, we will use global gradient-based optimization. The gradient of P with respect to $\{\mathbf{U}_\mathbf{k}\}$ is expressed straightforwardly

$$\begin{aligned}\frac{\partial P}{\partial (U_{j,\mathbf{k}}{}^{i})^{*}} &= \frac{p}{\sqrt{N_k}} \sum_{A,\mathbf{R}} |\bar{Q}_{i}{}^{A_{\mathbf{R}}}|^{p-1} \sum_{\mu \in A} [\langle \psi_{j,\mathbf{k}} \| \bar{\phi}_{\mu,\mathbf{R}} \rangle \bar{W}_{\mu,\mathbf{R}}{}^{i} \\ &\quad + (\bar{C}_{\mu,\mathbf{R}}{}^{j,\mathbf{k}})^{*} \langle \bar{\phi}_{\mu,\mathbf{R}} \| \tilde{\psi}_{i,0} \rangle] \\ &= \frac{p}{\sqrt{N_k}} \sum_{A,\mathbf{R}} |\bar{Q}_{i}{}^{A_{\mathbf{R}}}|^{p-1} e^{-i\mathbf{k}\cdot\mathbf{R}} \\ &\quad \sum_{\mu \in A} [\langle \psi_{j,\mathbf{k}} \| \bar{\phi}_{\mu,0} \rangle \bar{W}_{\mu,\mathbf{R}}{}^{i} + (\bar{C}_{\mu,\mathbf{R}}{}^{j,\mathbf{k}})^{*} \langle \bar{\phi}_{\mu,\mathbf{R}} | \\ &\quad |\tilde{\psi}_{i,0} \rangle]\end{aligned} \tag{21}$$

eq 21 uses the standard complex-valued form[39] of the derivative of a real-valued function of complex-valued parameters, $(\partial f/\partial z^{*}) \equiv (\partial f/\partial \mathrm{Re} z) + \mathrm{i}(\partial f/\partial \mathrm{Im} z)$, which makes the notation more compact. The evaluation of the gradient again leverages the Bloch-MBS overlaps (eq 13) and is straightforward.

It is of course more convenient to express the PMWF functional in terms of nonredundant variables by introducing the standard exponential parametrization of a unitary matrix, $\mathrm{U} \equiv \exp(\boldsymbol{\kappa} - \boldsymbol{\kappa}^{\dagger})$, where $\boldsymbol{\kappa}$ is a complex triangular matrix. It is straightforward to convert the "Euclidean" gradient arranged as a matrix for each $\mathbf{k}$

$$(\Gamma_{\mathbf{k}})_{j}^{\;i} \equiv \frac{\partial P}{\partial (C_{j,\mathbf{k}}{}^{i})^{*}} \tag{22}$$

to its "curvilinear" counterpart

$$(\mathbf{G}_{\mathbf{k}})_{j}^{\;i} \equiv \frac{\partial P}{\partial (\kappa_{j,\mathbf{k}}{}^{i})^{*}} \tag{23}$$

as follows[40,41]

$$\mathbf{G}_{\mathbf{k}} = (\mathbf{U}_{\mathbf{k}})^{\dagger}\Gamma_{\mathbf{k}} - (\Gamma_{\mathbf{k}})^{\dagger}\mathbf{U}_{\mathbf{k}} \tag{24}$$

Note that $\mathbf{G}_{\mathbf{k}}$ is antihermitian, just like $\boldsymbol{\kappa}_{\mathbf{k}} - \boldsymbol{\kappa}_{\mathbf{k}}^{\dagger}$. Maximization of the PMWF functional expressed in terms of $\boldsymbol{\kappa}_{\mathbf{k}}$ is a standard (unconstrained) nonlinear optimization problem; its solution is described in the following section.

2.3.3. Direction Choice: SA, CG, and BFGS.

The main focus on our work is how to solve for the PMWFs robustly. As the reference methods for locating the PM functional maxima, we will use the SA and (nonlinear) CG methods. In particular, we have chosen three specific varieties of CG for comparison: the Polak−Ribière formulation (CGPR),[42] the Fletcher−Reeves formulation (CGFR),[43] and the Hestenes−Stiefel formulation (CGHS).[44] Due to the well-known nature of SA and nonlinear CG (see, e.g., any textbook on numerical optimization), we will not discuss their implementation details here, except to note that, for each of the three different CG variants considered, we also varied the number of SA steps taken before beginning CG. The numbers of initial SA steps considered in this work were 1, 2, 5, 10, and 15, meaning that, for each system, we ran a total of 15 different CG calculations. Also note that the optimization problem of the real-valued PMWF functional P was recast (as usual) in terms of real and imaginary components of the complex-valued parameters $U_{j,\mathbf{k}}{}^{i}$, that is, henceforth the gradient and other vectors will consist of $o(o-1)N_k$ real numbers only, where o is the number of Bloch orbitals being localized;[e] complex-valued formulations of the optimization problem[45] were not considered here.

The BFGS method, though also well-known, warrants a bit of discussion. In particular, we have employed the "two-loop recursion" form of the limited-memory BFGS[46] (L-BFGS; henceforth, we will omit the "L-" prefix unless this algorithmic detail is relevant) algorithm for updating the estimated inverse Hessian; the initial estimate of the inverse Hessian was chosen to be an identity matrix. Because each BFGS iteration depends on some number of prior iterations (the "history") to generate an updated estimate of the inverse Hessian, it is necessary to select the size of this history (i.e., the number of iterations kept). In addition, regardless of the history size, the first update must be necessarily be SA since the history does not yet exist. Of course, it is also possible to perform any number of SA steps before beginning the BFGS procedure, and it is these two parameters (the history size and the number of initial SA steps) that define the BFGS algorithm as implemented here. For all systems, we used five different initial SA values and five different history sizes, for a total of 25 BFGS solver setups. The initial SA values considered were 1, 2, 5, 10, and 15; the history sizes considered were also 1, 2, 5, 10, and 15. In future discussion, BFGS parameters are indicated as ⟨no. of SA steps, history size⟩.

2.3.4. Line Search.

Regardless of how the direction was chosen (SA, CG, and BFGS), the line search was performed in the same manner, using a low-order polynomial approximation of the objective function along the trial direction. First, the proposed direction is checked to point uphill (if not, the trial direction is reversed). Then, given a fitting range upper bound T_{μ} (see below) and the polynomial order $n = 4$, the PM functional is evaluated at $n + 1$ evenly-spaced points, $\{\mu_0 \equiv 0, \mu_1 \equiv T_{\mu}/n, \mu_2 \equiv 2T_{\mu}/n... \}$, in $[0, T_{\mu}]$. The $\{P(\mu_i)\}$ set is then used to construct a polynomial fit, $f(\mu_i)$, and the bisection method is used to find a root of $f'(\mu_i)$. The fitting range and polynomial orders are determined as follows.

- Iteration 0: T_{μ} is estimated from the shortest orbital rotation frequency $\omega_{\max}$ along the given direction via eq 15 of ref 41 (the largest $\omega_{\max}$ is chosen among all $\mathbf{k}$ points).
- Iteration i: the upper bound from the previous iteration is used as a trial upper bound. If $P(\mu_1)$ is less than $P(0)$, then T_{μ} is reset to μ_1, else $P(\mu_k)$ are evaluated for increasing k until $P(\mu_k) < P(\mu_{k-1})$ is found. If such k is found, values $\{P(\mu_i)\}$, $i \in [0, k]$ are fitted to polynomial of order $k - 1$; else, T_{μ} is reset to $5T_{\mu}$ and the process is repeated.

Finally, if we fail to find an acceptable upper bound along a chosen direction, we will reset to the SA direction. If the upper bound is not located in the SA direction or the root finder fails, then the upper bound is recomputed as done at the start. These resets are rarely necessary when performing BFGS.

Note that, in addition to resetting the CG direction to the SA direction whenever an acceptable upper bound cannot be found, it is also necessary to reset the CG direction every n iterations, where n is the number of orbitals being localized. This is because a system with n variables can only have n conjugate directions.

3. COMPUTATIONAL DETAILS

All calculations were carried out in the developmental version of the Massively Parallel Quantum Chemistry (MPQC) package (version 4.0.0).[47] All computations were performed on the Cascades commodity cluster at Virginia Tech.[48]

Hartree−Fock computations were carried out using the reduced-scaling LCAO formalism described in ref 29. The Coulomb potential was evaluated using multipole-accelerated real-space summation and density-fitting approximation, whereas the exchange potential was evaluated using concentric atomic density fitting.[29] Table 1 lists the test systems as well as the

https://doi.org/10.1021/acs.jctc.1c00238
J. Chem. Theory Comput. 2021, 17, 7406−7415

Table 1. Test Systems Used to Assess the Performance of PMWF Solvers, along with the Relevant Computational Details[a]

system	OBS	Monkhorst−Pack mesh size[a]	PM gradient norm
trans-$(C_2H_2)_\infty$	6-31G*	101	10^{-8}
$(C_2H_4)_\infty$	6-31G*	101	10^{-8}
(4,0) nanotube	6-31G*	51	10^{-8}
graphene	6-31G	21 × 21	10^{-8}
h-BN	6-31G	21 × 21	10^{-8}
LiH	CR-cc-pVDZ[53]	11 × 11 × 11	10^{-8}
diamond	6-31G*	11 × 11 × 11	10^{-8}
silicon	6-31G*	11 × 11 × 11	10^{-7}

[a] All systems except h-BN utilized primitive unit cells; an orthorhombic non-primitive unit cell was used for h-BN.

corresponding OBS, the Monkhorst−Pack mesh size, and the PM convergence threshold employed for each. In all cases, the def2-universal-JFIT basis set[49] was used as the density fitting basis (in ref 29, we showed that the errors due to the use of this fitting basis and the local fitting approximations were sufficiently small for all systems considered in this study). The pseudoinverse atomic charges were evaluated using the Huzinaga MINI basis set[50−52] as the minimal AO basis; a limited comparison to the results obtained with other MBS is presented in the Supporting Information. All occupied orbitals (including core) were localized. The PM functional with $p = 4$ (see eq 4) was utilized throughout. Complete input files (which specify unit cell parameters) and geometries can be found in the Supporting Information.

Note that the convergence thresholds that we use in this work are significantly tighter than the typical threshold of 10^{-5} utilized in the comparable studies.[21,28] A slightly looser convergence in bulk Si was due to the greater range of the lengthscales of the Wannier functions in this system; localizing core and valence orbitals separately would alleviate these issues, but was not pursued in order to keep the solver assessment protocol as uniform and as stringent as possible.

4. RESULTS

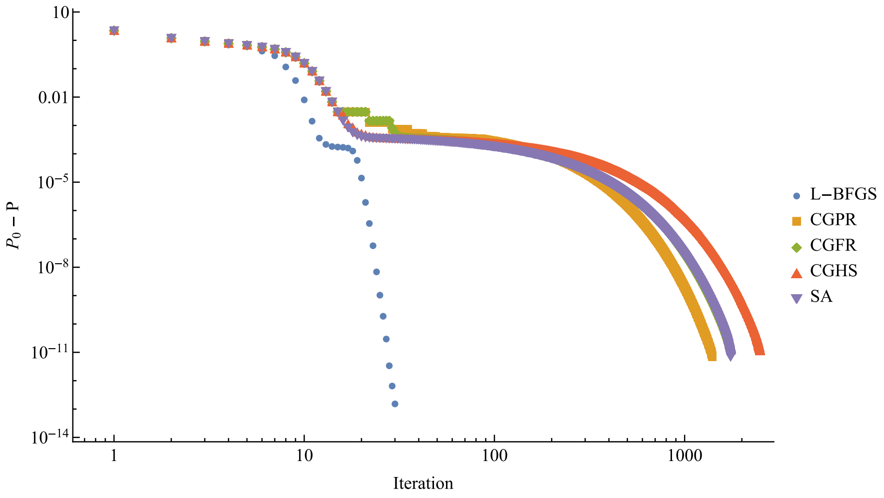
As discussed in Section 2.3.1, the PM functional does not specify WFs uniquely; thus, to compare computed WFs, we compare the corresponding values of the PM functional; two WF sets will be referred to as *equivalent* if they correspond to the same value of the PM functional up to target precision. For every test system, all PMWF solvers considered in this work (BFGS with 25 parameter settings, CG with 15 parameter settings, and SA; see Section 2.3), when converged successfully, produced WFs that were practically equivalent; only for silicon and the carbon nanotube did the variance of the final P value exceeded 10^{-8} (see the Supporting Information for more details). Limited testing also indicated that the use of the standard (CPR) and nonstandard guesses produced equivalent sets of Wannier functions.

Although all solvers produced equivalent sets of PMWFs, the number of iterations needed to locate the maxima of the PM functional differed dramatically between the different classes of solvers. Column "min" in Table 2 lists the minimum number of iterations needed to arrive at the solution, broken down by the system and solver class. In all cases except the zero-gap system, the BFGS solver converged in fewer than 60 iterations in the best-case scenario, and only two systems required more than 100 iterations in the worst-case scenario. Each of the three CG variants managed to converge in under 100 iterations for at least one system, but even CGFR, which saw the most success in converging in less than 100 iterations, only did so for three

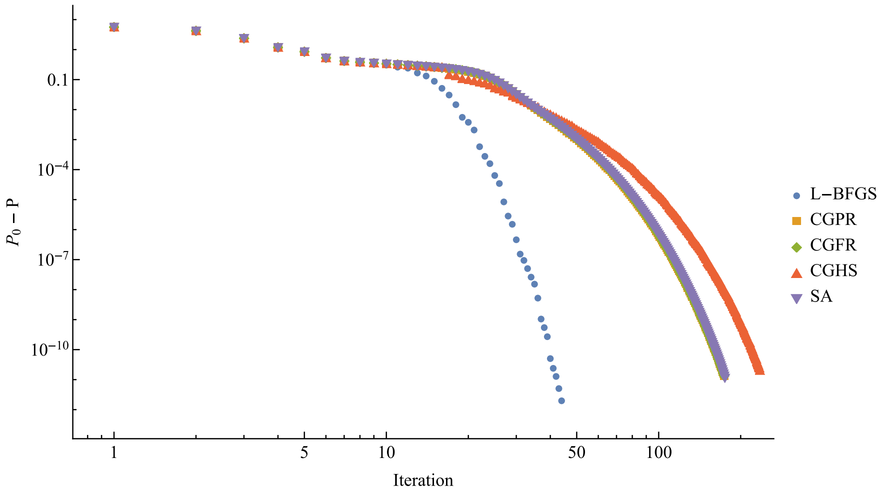
Table 2. Length Summary

system	solver	min	max	mean	st. dev.	system	solver	Min	max	mean	st. dev.
C_2H_2	BFGS	31	47	40.5	3.5	h-BN	BFGS	45	57	50.2	3.3
	CG	1398	2534	2007.5	389.1		CG	174	808	260.8	160.9
	CGPR	1398	1934	1708.6	209.0		CGPR[c]	174	808	339.3	312.7
	CGFR	1752	1832	1801.0	35.5		CGFR	174	250	210.6	35.5
	CGHS	2483	2534	2513.0	22.5		CGHS	235	257	248.2	8.3
	SA	1745					SA	175			
C_2H_4	BFGS	24	37	29.0	3.2	LiH	BFGS	10	22	14.5	4.2
	CG	66	729	230.7	209.8		CG	14	1441	783.2	607.2
	CGPR	69	425	215.0	133.2		CGPR	786	1062	900.6	122.7
	CGFR	109	187	158.6	33.8		CGFR	14	23	17.6	3.8
	CGHS	66	729	318.4	344.6		CGHS	1395	1441	1431.4	20.4
	SA	52					SA	1457			
diamond	BFGS	23	28	24.7	1.4	nanotube	BFGS	75	101	86.1	8.3
	CG	39	212	89.6	65.3		CG	713	2275	1174.5	471.0
	CGPR	43	212	133.0	69.6		CGPR	935	2275	1309.0	581.3
	CGFR	39	51	46.2	5.1		CGFR	1015	1903	1447.8	334.5
	CGHS[a]						CGHS	713	853	766.8	55.4
	SA[b]						SA	485			
graphene	BFGS	53	186	86.0	42.0	silicon	BFGS	21	29	24.1	2.0
	CG	2657	3666	3296.8	275.8		CG	32	2975	954.9	1240.7
	CGPR[c]	2657	3666	3234.8	439.4		CGPR	113	221	176.4	44.3
	CGFR	3321	3376	3346.4	21.6		CGFR	32	83	55.4	19.9
	CGHS[a]						CGHS	2243	2975	2633.0	310.8
	SA[b]						SA	3171			

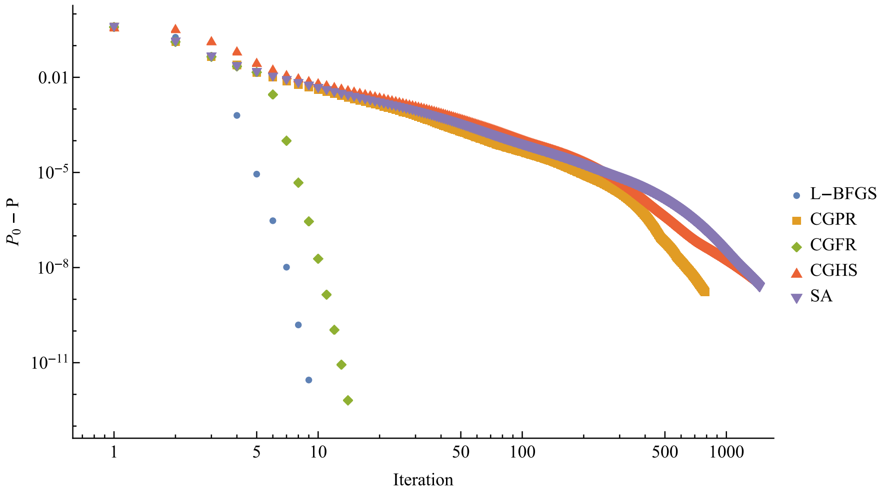
[a] All five calculations failed to converge. [b] Calculation failed to converge. [c] One calculation failed to converge.



(a) trans-$(C_2H_2)_\infty$



(b) h-BN



(c) LiH

Figure 1. Plots of the difference between the PM functional value P and the converged value P_0 vs the iteration count. For each system represented, the shortest calculation in each category was chosen; P_0 was taken to be the greatest final P value across the selected calculations for a given system.

systems. In other cases, the various CG variants could take thousands of iterations to converge, and CGHS failed to converge in 4,000 iterations for both diamond and graphene, regardless of the starting number of SA steps. Depending on the system, SA could converge in under 50 iterations or take hundreds or thousands to converge; in two cases, it failed to converge at all within 4000 iterations. Even in the best-case scenario (i.e., minimum iterations to convergence), BFGS was always superior to CG and SA, though the latter two could sometimes come close. But even when CG and SA were nearly comparable to BFGS in terms of number of iterations to solution, their convergence behavior could not be counted on, as is evidenced by the much larger calculation length standard deviations for these solvers compared with the same metric for BFGS. Overall, the performance of the SA and CG solvers can, at best, be characterized as unreliable; in contrast, BFGS is reliably rapid.

Due to the significant variation in the performance of the CG and SA solvers for different systems, it is difficult to pinpoint the origin of their struggles. Figure 1 illustrates the convergence patterns observed for representative 1-d, 2-d, and 3-d test systems. The superior performance of BFGS compared to CG and SA is plainly visible. Also note the extended plateau exhibited by the CG and SA solvers in the 1-d system, which is a typical CG convergence pattern[54] and which suggests a high condition number of the PM Hessian in this system. The BFGS solver also exhibits a plateau in this system, but it is much shorter. Lastly, the quality of the initial guess can vary greatly from system to system: in the 1-d system, the initial guess is clearly significantly worse than in the 2-d and 3-d systems, as indicated by significantly larger initial deviations.

The performance of BFGS is also relatively insensitive to the choice of its parameters, namely, the number of SA steps at the start and the history size, as illustrated in Table 2. The small standard deviation of the BFGS solvers' performance (<10 for all systems other than graphene) indicates that approximately the same number of iterations is needed to locate the maximum regardless of the BFGS parameter values. In contrast, the CG solvers' performance can depend strongly on the number of starting SA steps. This correlates with the unreliable performance of the CG solvers that we noted previously.

In addition, the number of iterations needed for the CG solver also correlates as expected with the findings of Jónsson et al.,[20] whose implementation of CG required hundreds of iterations in periodic systems. Cases where our CG implementation converged in fewer iterations may be a result of differences in initial guess (CPR in our work, random guess in their work), while situations in which our CG implementation took longer may be due to the tighter convergence thresholds employed herein. Furthermore, the only 3-d system that Jónsson et al. reported was a benzene crystal,[20] which, like other molecular crystals, would have low-dispersion band structure and for which it should be easier to generate a localized initial guess (see our discussion in Section 2.3.1). We focused on the more challenging ionic and covalent 3-d systems.

Although the BFGS-based PMWF solver is already highly robust, some parameter choices are systematically better than others. Thus, we analyzed the distribution of the number of iterations needed to locate the solution for a given system with the given BFGS solver parameter values relative to the *smallest* number of iterations needed for that system; the results of this analysis are listed in Table 3 (see the Supporting Information for the raw number of iterations for each system). We highlighted the smallest and largest values in each column to make it easier to locate the fastest and slowest solvers. The ⟨2, 15⟩ BFGS solver is overall the fastest, both on average and in the worst-case scenario, and thus is the recommended default choice.

Table 3. Statistical Analyses of the Number of Iterations to Solution for Each BFGS Solver Parameter Set[a,b,c,d]

⟨ # of SA, History ⟩	Min	Max	Mean	St. Dev
⟨1, 1⟩	1.00000	2.83019	1.33181	0.61693
⟨1, 2⟩	1.00000	1.81132	1.19016	0.27206
⟨1, 5⟩	1.04348	1.35849	1.20877	0.11738
⟨1, 10⟩	1.03774	1.41935	1.19625	0.14074
⟨1, 15⟩	1.01887	1.38095	1.18609	0.12330
⟨2, 1⟩	1.00000	3.50943	1.41816	0.85422
⟨2, 2⟩	1.00000	1.43396	1.14882	0.13846
⟨2, 5⟩	1.04348	1.22581	1.13897	0.06151
⟨2, 10⟩	1.03774	1.41935	1.14482	0.12524
⟨2, 15⟩	1.01887	1.19355	1.11134	0.05189
⟨5, 1⟩	1.00000	3.39623	1.43594	0.80097
⟨5, 2⟩	1.00000	2.05660	1.22799	0.34514
⟨5, 5⟩	1.04000	1.41509	1.17032	0.13938
⟨5, 10⟩	1.05660	1.41935	1.17322	0.12475
⟨5, 15⟩	1.01887	1.30000	1.15381	0.09378
⟨10, 1⟩	1.00000	1.70000	1.26942	0.28516
⟨10, 2⟩	1.00000	1.88679	1.28996	0.33131
⟨10, 5⟩	1.00000	1.70000	1.21055	0.25964
⟨10, 10⟩	1.01887	1.70000	1.23150	0.23658
⟨10, 15⟩	1.03774	1.80000	1.24892	0.24966
⟨15, 1⟩	1.04444	3.01887	1.60319	0.66128
⟨15, 2⟩	1.10667	2.30189	1.45403	0.47729
⟨15, 5⟩	1.06667	2.10000	1.32343	0.33587
⟨15, 10⟩	1.05660	2.10000	1.29656	0.34030
⟨15, 15⟩	1.05660	2.20000	1.33176	0.36673

[a]Colored boxes identify the smallest and largest values in each column as follows: green (smallest), blue (second smallest), red (largest), orange (second largest). [b]A "min" value of 1 indicates that this solver was the fastest (required the fewest iterations) for at least 1 test system. [c]A "max" value of 1.5 means that, for every test system, this solver required at most 50% more iterations than the fastest solver for that system. [d]A "mean" value of 1.2 means that, for a give test system, this solver required on average 20% more iterations than the fastest solver for that system.

5. SUMMARY AND PERSPECTIVE

We described a robust BFGS-based solver to obtain (generalized) PMWF in periodic solids whose use was pioneered recently by Jónsson et al.[20] The PM functional used in this work utilized atomic charges using a simple pseudoinverse projection onto a pre-defined minimal AO basis, thus making its evaluation convenient in a periodic LCAO representation. An essential contributor to the robustness of the solver is the novel automated CPR method for generating the initial guess. The limited-memory BFGS solver converged very tightly in fewer than 60 iterations in one-, two-, and three-dimensional systems featuring a variety of bonding patterns (covalent and ionic) and gaps, even in systems with very large BvK unit cells (thousands of atoms). The sole exception was one system with a vanishing gap, where ~80 iterations were needed. This is a significant improvement on the more traditional SA-based solver that can require hundreds or thousands of iterations, or the nonlinear CG solvers that often converge faster than SA, but can unpredictably converge very slowly or even fail to converge at all. Although the performance of the solver was relatively insensitive to the BFGS history size, a history size of 15 iterations was found to be near optimal.

Clearly, the BFGS-based solver should be robustly useable for computing other generalized Wannier functions, such as the Boys (maximally-localized) WFs. Although here we only explored localization of occupied states, the solver should be also applicable to the unoccupied states with valence character. The automated CPR method for generating initial WF guesses could be used in conjunction with PW-based representations of Bloch orbitals, potentially improving on the existing approaches.[12,38] Lastly, it is also worthwhile to assess the efficacy of BFGS-based solvers for other challenging orbital optimization problems in molecules and solids, such as localization of unoccupied (virtual) orbitals[21,28] and for the orbital optimization in the context of Perdew−Zunger self-interaction-corrected DFT (notably, some limited use of BFGS in this context was recently reported by Lehtola et al.[55]).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00238.

Additional computational data (PDF)

Complete program inputs and molecular geometries (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Edward F. Valeev − *Department of Chemistry, Virginia Tech, Blacksburg, Virginia 24061, United States;* orcid.org/0000-0001-9923-6256; Email: efv@vt.edu

Authors

Marjory C. Clement − *Department of Chemistry, Virginia Tech, Blacksburg, Virginia 24061, United States*

Xiao Wang − *Department of Chemistry, Virginia Tech, Blacksburg, Virginia 24061, United States; Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, United States;* orcid.org/0000-0003-1402-7522

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c00238

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the U.S. National Science Foundation (awards 1550456 and 1800348) and by the Virginia Tech Institute for Critical Technology and Applied Science (M.C.). We also acknowledge Advanced Research Computing at Virginia Tech (www.arc.vt.edu) for providing computational resources and technical support that have contributed to the results reported within this paper. The Flatiron Institute is a division of the Simons Foundation.

■ ADDITIONAL NOTES

[a] Gradient-driven optimization of the four aforementioned localization functionals starting from unmixed σ and π orbitals, such as the canonical orbitals, preserves the σ/π separation. However, the stationary points of the FB, ER, and vN functionals that mix σ and π orbitals are usually more optimal than the unmixed solutions; thus to arrive at the optimal solution, the σ/π symmetry is broken by, for example, randomizing the initial guess. In contrast, the PM localization will usually separate σ and π orbitals even if the starting guess involves their linear combinations. However, the σ/π separation does not always occur in PM localization (simple counterexamples include diatomic molecules).

[b] However, unlike the ER and vN orbitals, the PM orbitals lack the intra-atomic localizaton.

[c] The superscript and subscript indices on matrix elements refer to the columns and rows, respectively.

[d] Curiously, even Mulliken recognized that the free-atom orbitals are ideal for defining charges by remarking: "the ideal LCAO-MO population analysis would perhaps be in terms of free atom SCF AO's".[34]

[e] In practice, the implementation uses pairs of real square matrices to represent $\boldsymbol{\kappa}_k$, $\mathbf{G}_k$, thus using $2o^2N_k$ (instead of $o(o-1)N_k$) parameters. This is due to the lack of support for the (anti)symmetric matrix format in the Eigen library used for numerical manipulations in the PM solver.

■ REFERENCES

(1) Boys, S. F. Construction of Some Molecular Orbitals to Be Approximately Invariant for Changes from One Molecule to Another. *Rev. Mod. Phys.* **1960**, *32*, 296−299.

(2) Foster, J. M.; Boys, S. F. Canonical Configurational Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300−302.

(3) Pipek, J.; Mezey, P. G. A Fast Intrinsic Localization Procedure Applicable for Ab Initio and Semiempirical Linear Combination of Atomic Orbital Wave Functions. *J. Chem. Phys.* **1989**, *90*, 4916−4926.

(4) Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457−464.

(5) von Niessen, W. Density Localization of Atomic and Molecular Orbitals. I. *J. Chem. Phys.* **1972**, *56*, 4290−4297.

(6) Cioslowski, J. Partitioning of the Orbital Overlap Matrix and the Localization Criteria. *J. Math. Chem.* **1991**, *8*, 169−178.

(7) Alcoba, D. R.; Lain, L.; Torre, A.; Bochicchio, R. C. An Orbital Localization Criterion Based on the Theory of "Fuzzy" Atoms. *J. Comput. Chem.* **2006**, *27*, 596−608.

(8) Knizia, G. Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and Chemical Concepts. *J. Chem. Theory Comput.* **2013**, *9*, 4834−4843.

(9) Lehtola, S.; Jónsson, H. Pipek−Mezey Orbital Localization Using Various Partial Charge Estimates. *J. Chem. Theory Comput.* **2014**, *10*, 642−649.

(10) Janowski, T. Near Equivalence of Intrinsic Atomic Orbitals and Quasiatomic Orbitals. *J. Chem. Theory Comput.* **2014**, *10*, 3085−3091.

(11) Høyvik, I.-M.; Jansik, B.; Jørgensen, P. Orbital Localization Using Fourth Central Moment Minimization. *J. Chem. Phys.* **2012**, *137*, 224114.

(12) Marzari, N.; Vanderbilt, D. Maximally Localized Generalized Wannier Functions for Composite Energy Bands. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1997**, *56*, 12847−12865.

(13) Wannier, G. H. The Structure of Electronic Excitation Levels in Insulating Crystals. *Phys. Rev.* **1937**, *52*, 191−197.

(14) Marzari, N.; Mostofi, A. A.; Yates, J. R.; Souza, I.; Vanderbilt, D. Maximally Localized Wannier Functions: Theory and Applications. *Rev. Mod. Phys.* **2012**, *84*, 1419−1475.

(15) Pizzi, G.; Vitale, V.; Arita, R.; Blügel, S.; Freimuth, F.; Géranton, G.; Gibertini, M.; Gresch, D.; Johnson, C.; Koretsune, T.; Ibañez-Azpiroz, J.; Lee, H.; Lihm, J.-M.; Marchand, D.; Marrazzo, A.; Mokrousov, Y.; Mustafa, J. I.; Nohara, Y.; Nomura, Y.; Paulatto, L.; Poncé, S.; Ponweiser, T.; Qiao, J.; Thöle, F.; Tsirkin, S. S.; Wierzbowska, M.; Marzari, N.; Vanderbilt, D.; Souza, I.; Mostofi, A. A.; Yates, J. R. Wannier90 as a Community Code: New Features and Applications. *J. Phys.: Condens. Matter* **2020**, *32*, 165902.

(16) Maschio, L.; Usvyat, D.; Manby, F. R.; Casassa, S.; Pisani, C.; Schütz, M. Fast Local-MP2 Method with Density-Fitting for Crystals. I. Theory and Algorithms. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2007**, *76*, 075101.

(17) Usvyat, D.; Civalleri, B.; Maschio, L.; Dovesi, R.; Pisani, C.; Schütz, M. Approaching the Theoretical Limit in Periodic Local MP2 Calculations with Atomic-Orbital Basis Sets: The Case of LiH. *J. Chem. Phys.* **2011**, *134*, 214105.

(18) Zicovich-Wilson, C. M.; Dovesi, R.; Saunders, V. R. A General Method to Obtain Well Localized Wannier Functions for Composite Energy Bands in Linear Combination of Atomic Orbital Periodic Calculations. *J. Chem. Phys.* **2001**, *115*, 9708−9719.

(19) Dovesi, R.; Erba, A.; Orlando, R.; Zicovich-Wilson, C. M.; Civalleri, B.; Maschio, L.; Rérat, M.; Casassa, S.; Baima, J.; Salustro, S.; Kirtman, B. Quantum-Mechanical Condensed Matter Simulations with CRYSTAL. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1360.

(20) Jónsson, E. Ö.; Lehtola, S.; Puska, M.; Jónsson, H. Theory and Applications of Generalized Pipek−Mezey Wannier Functions. *J. Chem. Theory Comput.* **2017**, *13*, 460−474.

(21) Lehtola, S.; Jónsson, H. Unitary Optimization of Localized Molecular Orbitals. *J. Chem. Theory Comput.* **2013**, *9*, 5365−5372.

(22) Broyden, C. G. The Convergence of a Class of Double-Rank Minimization Algorithms 1. General Considerations. *IMA J. Appl. Math.* **1970**, *6*, 76−90.

(23) Fletcher, R. New Approach to Variable Metric Algorithms. *Comput. J.* **1970**, *13*, 317−322.

(24) Goldfarb, D. A Family of Variable-Metric Methods Derived by Variational Means. *Math. Comput.* **1970**, *24*, 23.

(25) Shanno, D. F. Conditioning of Quasi-Newton Methods for Function Minimization. *Math. Comput.* **1970**, *24*, 647.

(26) Kari, R. Parametrization and Comparative Analysis of theBFGS Optimization Algorithm for the Determination of Optimum Linear Coefficients. *Int. J. Quantum Chem.* **1984**, *25*, 321−329.

(27) Leonard, J. M.; Luken, W. L. Quadratically Convergent Calculation of Localized Molecular Orbitals. *Theor. Chim. Acta* **1982**, *62*, 107−132.

(28) Høyvik, I.-M.; Jansik, B.; Jørgensen, P. Trust Region Minimization of Orbital Localization Functions. *J. Chem. Theory Comput.* **2012**, *8*, 3137−3146.

(29) Wang, X.; Lewis, C. A.; Valeev, E. F. Efficient Evaluation of Exact Exchange for Periodic Systems via Concentric Atomic Density Fitting. *J. Chem. Phys.* **2020**, *153*, 124116.

(30) Harris, F. E. In *Theoretical Chemistry*; Eyring, H., Henderson, D., Eds.; Elsevier, 1975; pp 147−218.

(31) Pisani, C.; Dovesi, R. Exact-Exchange Hartree-Fock Calculations for Periodic Systems. I. Illustration of the Method. *Int. J. Quantum Chem.* **1980**, *17*, 501−516.

(32) Monkhorst, H. J.; Pack, J. D. Special Points for Brillouin-Zone Integrations. *Phys. Rev. B: Solid State* **1976**, *13*, 5188−5192.

(33) West, A. C.; Schmidt, M. W.; Gordon, M. S.; Ruedenberg, K. A Comprehensive Analysis of Molecule-Intrinsic Quasi-Atomic, Bonding, and Correlating Orbitals. I. Hartree-Fock Wave Functions. *J. Chem. Phys.* **2013**, *139*, 234107.

(34) Mulliken, R. S. Criteria for the Construction of Good Self-Consistent-Field Molecular Orbital Wave Functions, and the Significance of LCAO-MO Population Analysis. *J. Chem. Phys.* **1962**, *36*, 3428−3439.

(35) Lu, W. C.; Wang, C. Z.; Schmidt, M. W.; Bytautas, L.; Ho, K. M.; Ruedenberg, K. Molecule Intrinsic Minimal Basis Sets. I. Exact Resolution of Ab Initio Optimized Molecular Orbitals in Terms of Deformed Atomic Minimal-Basis Orbitals. *J. Chem. Phys.* **2004**, *120*, 2629−2637.

(36) Laikov, D. N. Intrinsic Minimal Atomic Basis Representation of Molecular Electronic Wavefunctions. *Int. J. Quantum Chem.* **2011**, *111*, 2851−2867.

(37) King, H. F.; Stanton, R. E.; Kim, H.; Wyatt, R. E.; Parr, R. G. Corresponding Orbitals and the Nonorthogonality Problem in Molecular Quantum Mechanics. *J. Chem. Phys.* **1967**, *47*, 1936−1941.

(38) Mustafa, J. I.; Coh, S.; Cohen, M. L.; Louie, S. G. Automated Construction of Maximally Localized Wannier Functions: Optimized Projection Functions Method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *92*, 165134.

(39) Brandwood, D. H. A complex gradient operator and its application in adaptive array theory. *IEE Proc.-Part F Commun. Radar Signal Process.* **1983**, *130*, 11−16.

(40) Abrudan, T. E.; Eriksson, J.; Koivunen, V. Steepest Descent Algorithms for Optimization under Unitary Matrix Constraint. *IEEE Trans. Signal Process.* **2008**, *56*, 1134−1147.

(41) Abrudan, T.; Eriksson, J.; Koivunen, V. Conjugate gradient algorithm for optimization under unitary matrix constraint. *Signal Process.* **2009**, *89*, 1704−1714.

(42) Polak, E.; Ribière, G. Note sur la convergence de méthodes de directions conjuguées. *ESAIM Math. Model. Numer. Anal.* **1969**, *3*, 35−43.

(43) Fletcher, R.; Reeves, C. M. Function minimization by conjugate gradients. *Comput. J.* **1964**, *7*, 149−154.

(44) Hestenes, M. R.; Stiefel, E. Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.* **1952**, *49*, 409−435.

(45) Sorber, L.; Barel, M. V.; Lathauwer, L. D. Unconstrained Optimization of Real Functions in Complex Variables. *SIAM J. Optim.* **2012**, *22*, 879−898.

(46) Nocedal, J. Updating Quasi-Newton Matrices with Limited Storage. *Math. Comput.* **1980**, *35*, 773.

(47) Peng, C.; Lewis, C. A.; Wang, X.; Clement, M. C.; Pierce, K.; Rishi, V.; Pavošević, F.; Slattery, S.; Zhang, J.; Teke, N.; Kumar, A.; Masteran, C.; Asadchev, A.; Calvin, J. A.; Valeev, E. F. Massively Parallel Quantum Chemistry: A high-performance research platform for electronic structure. *J. Chem. Phys.* **2020**, *153*, 044120.

(48) Cascades compute cluster at Virginia Tech. https://arc-rtd.readthedocs.io/en/latest/resources/compute/02cascades.html (accessed Oct 24, 2021).

(49) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057−1065.

(50) van Duijneveldt, F. B. *Gaussian Basis Sets for the Atoms H-Ne for Use in Molecular Calculations*, IBM Research Report RJ945, 1971.

(51) Andzelm, J.; Huzinaga, S.; Klobukowski, M.; Radzio-Andzelm, E.; Sakai, Y.; Tatewaki, H. In *Gaussian Basis Sets for Molecular Calculations*; Huzinaga, S., Ed.; Physical Sciences Data; Elsevier, 1984; Vol. *16*, Chapter Gaussian Basis Sets, pp 27−426.

(52) Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L. Basis Set Exchange: A Community Database for Computational Sciences. *J. Chem. Inf. Model.* **2007**, *47*, 1045−1052.

(53) Lorenz, M.; Maschio, L.; Schütz, M.; Usvyat, D. Local Ab Initio Methods for Calculating Optical Bandgaps in Periodic Systems. II. Periodic Density Fitted Local Configuration Interaction Singles Method for Solids. *J. Chem. Phys.* **2012**, *137*, 204119.

(54) Axelsson, O. Iteration Number for the Conjugate Gradient Method. *Math. Comput. Simulat.* **2003**, *61*, 421−435.

(55) Lehtola, S.; Head-Gordon, M.; Jónsson, H. Complex Orbitals, Multiple Local Minima, and Symmetry Breaking in Perdew−Zunger Self-Interaction Corrected Density Functional Theory Calculations. *J. Chem. Theory Comput.* **2016**, *12*, 3195.