Real-space orthogonal projector-augmented-wave method

Wenfei Li and Daniel Neuhauser
Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA

(Received 28 July 2020; revised 28 September 2020; accepted 20 October 2020; published 11 November 2020)

The projector-augmented-wave (PAW) method of Blöchl makes smooth but nonorthogonal orbitals. Here we show how to make a PAW orthogonal using a cheap transformation of the wave functions. We show that the resulting orthogonal PAW (OPAW), applied for density functional theory, reproduces (for a large variety of solids) band gaps from the ABINIT package. OPAW combines the underlying orthogonality of norm-conserving pseudopotentials with the large grid spacings and small energy cutoffs in PAW. The OPAW framework can also be combined with other electronic structure theory methods.

DOI: 10.1103/PhysRevB.102.195118

I. INTRODUCTION

A plane-wave basis set is natural when studying periodic systems with density functional theory (DFT) and post-DFT methods. Convergence with the basis set is simply verified by increasing a single parameter, the kinetic energy cutoff. However, due to the fast oscillation of atomic core states, a direct all-electron treatment is prohibitive. One way to circumvent this problem is to replace the effect of the chemically inert core states by an effective pseudopotential, and the resulting pseudo valence states are nonoscillatory [1,2]. DFT using pseudopotentials and a plane-wave basis set has therefore become one of the most popular choices in computational chemistry and materials science. However, despite the formal simplicity of norm-conserving pseudopotentials (NCPPs), treatment of first-row elements and transition metals is still computationally demanding due to the localized nature of $2p$ and $3d$ orbitals [3–5].

The projector-augmented wave (PAW) method proposed by Blöchl [6–9] seeks to make softer pseudo wave functions by relaxing the norm-conserving condition. There are several different implementations of the PAW method (e.g., [10–13]) with many successful applications.

In addition to the reduced kinetic energy cutoff, an advantage of the PAW method is that it provides a means for recovering the all-electron orbitals, and these orbitals possess the right nodal structures in the core region. Therefore, PAW enables the calculation of quantities such as hyperfine parameters, core-level spectra, electric-field gradients, and the NMR chemical shifts, which rely on a correct description of all-electron wave functions in the core region [14].

The PAW method is based on a map between the smoothed pseudo wave functions $\{\tilde{\psi}_m\}$ and the all-electron wave functions $\{\psi_m\}$. Unlike NCPPs, where the wave functions retain their orthogonality, the pseudo wave functions in PAW satisfy a generalized orthogonality condition,

$$\langle\tilde{\psi}_m|\hat{S}|\tilde{\psi}_n\rangle = \delta_{mn}, \tag{1}$$

which leads to a generalized eigenproblem, $\tilde{H}\tilde{\psi}_m = \epsilon_m\hat{S}\tilde{\psi}_m$, where we introduced the one-body Hamiltonian $\tilde{H}$ and overlap operator $\hat{S}$ (both detailed later).

However, the fact that the pseudo-orbitals are not orthogonal complicates the use of PAW for applications that rely on the orthogonality of molecular orbitals. These include some post-DFT methods, as well as several lower-scaling DFT methods, including the modified deterministic Chebyshev approach (see, e.g., [15]) or stochastic DFT methods [16,17], which are able to handle a large number of electrons (potentially hundreds of thousands for the stochastic approach) by filtering a function of an orthogonal Hamiltonian.

Here we solve the nonorthogonality problem by an efficient numerical transformation of the PAW problem to an orthogonal one,

$$(\hat{S}^{-\frac{1}{2}}\tilde{H}\hat{S}^{-\frac{1}{2}})\bar{\psi}_m = \epsilon_m\bar{\psi}_m, \tag{2}$$

with $\bar{\psi}_m = \hat{S}^{1/2}\tilde{\psi}_m$ forming an orthogonal set, with the same norm as the all-electron orbitals (to be proved later). The key is that we show how to numerically apply the $\hat{S}^{-1/2}$ (or $\hat{S}^{-1}$) operator efficiently, without significantly raising the cost of applying the Hamiltonian.

The resulting approach retains one of the desirable features of NCPP, orthogonality of molecular orbitals, and we therefore label it orthogonal PAW (OPAW). In addition to orthogonality, OPAW is also efficient because it is implemented in real space, exploiting the localization of atomic projector functions and partial waves [12,13].

OPAW provides a general framework and can be combined with different electronic structure methods. Here we apply the method with the Chebyshev-filtered subspace iteration (CheFS) DFT approach, concentrating on the fundamental band gap of solids. We show below excellent agreement with PAW calculations from the ABINIT package [11,18]. We also demonstrate that for many systems, PAW and OPAW band gaps converge with energy cutoff faster than NCPPs.

Section II presents the OPAW theory. Results are presented in Sec. III, and conclusions follow in Sec. IV. Technical details are deferred to the Appendixes.

II. THEORY

A. Orthogonal projector augmented wave

The basic relation in PAW is a map $\hat{T}$ yielding the true molecular eigenstates ψ_m from the smoother pseudo-orbitals,

$$|\psi_m\rangle = \hat{T}|\tilde{\psi}_m\rangle \equiv |\tilde{\psi}_m\rangle + \sum_{a,i} \left(|\phi_i^{(a)}\rangle - |\tilde{\phi}_i^{(a)}\rangle\right)\langle p_i^{(a)}|\tilde{\psi}_m\rangle, \quad (3)$$

where a is the atom index and i runs over all the partial wave channels (a combination of principal, angular momentum, and magnetic quantum numbers) associated with each atom; $\phi_i^{(a)}$ and $\tilde{\phi}_i^{(a)}$ are a true atomic orbital and a smoothed version which matches $\phi_i^{(a)}$ outside a small sphere around the atom (labeled the augmentation region). The atomic projectors $\{p_i^{(a)}\}$ are localized in the augmentation region and are built to span the space within each augmentation sphere, i.e., $\sum_i |\tilde{\phi}_i^{(a)}\rangle\langle p_i^{(a)}| \simeq 1$, in the sphere.

With some derivations, one arrives at the working equation of PAW, the generalized eigenproblem $\tilde{H}\tilde{\psi}_m = \epsilon_m \hat{S}\tilde{\psi}_m$, where

$$\hat{S} \equiv \hat{T}^\dagger T = \mathbb{I} + \sum_{ij,a} |p_i^{(a)}\rangle s_{ij}^{(a)} \langle p_j^{(a)}|, \quad (4)$$

with $s_{ij}^{(a)} \equiv \langle \phi_i^{(a)}|\phi_j^{(a)}\rangle - \langle \tilde{\phi}_i^{(a)}|\tilde{\phi}_j^{(a)}\rangle$, and

$$\tilde{H} = -\frac{\nabla^2}{2} + v_{KS}(\boldsymbol{r}) + \sum_{ij,a} |p_i^{(a)}\rangle D_{ij}^{(a)} \langle p_j^{(a)}|. \quad (5)$$

The expressions for the Kohn-Sham effective potential $v_{KS}(\boldsymbol{r})$ and for $D_{ij}^{(a)}$ are found in various references [6,11]. While $s_{ij}^{(a)}$ are only atom dependent, $v_{KS}(\boldsymbol{r})$ and $D_{ij}^{(a)}$ both depend on the on-site PAW atomic density matrices: $\rho_{ij}^{(a)} = \sum_m \langle p_j^{(a)}|\tilde{\psi}_m\rangle\langle\tilde{\psi}_m|p_i^{(a)}\rangle$, as well as the smooth density $\tilde{n}(\boldsymbol{r}) = \sum_m |\tilde{\psi}_m(\boldsymbol{r})|^2$ and the sum extends over the occupied states. The on-site atomic density matrices and the smooth density are the key components in PAW and together with the atomic information govern the updated quantities in each self-consistent field (SCF) cycle.

In many applications, however, it is desirable to work with an orthonormal collection of wave functions. As mentioned in the Introduction, this can be achieved by the transformation

$$\bar{\psi}_m = \hat{S}^{1/2}\tilde{\psi}_m, \quad (6)$$

resulting in

$$\bar{H}\bar{\psi}_m = \epsilon_m \bar{\psi}_m, \quad (7)$$

where $\bar{H} = \hat{S}^{-\frac{1}{2}}\tilde{H}\hat{S}^{-\frac{1}{2}}$.

As an example, in Fig. 1 we show three-dimensional (3D) isosurfaces of ψ, $\bar{\psi}$, and $\tilde{\psi}$ for the $2p_z$ orbital from a calculation of a single oxygen atom, as well as the associated one-dimensional (1D) radial part obtained by projecting the 3D orbital to 1D. The three orbitals differ only in the core region; ψ clearly has more structure in the core, while the oscillatory features are attenuated or absent in $\bar{\psi}$ and $\tilde{\psi}$. Furthermore, the magnitudes of $\bar{\psi}$ and $\tilde{\psi}$ are smaller than those of ψ.

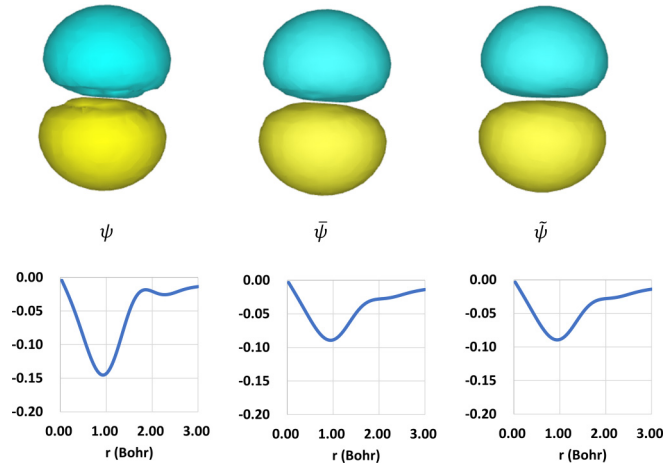

FIG. 1. Isosurfaces (top) and radial parts (bottom) of ψ, $\bar{\psi}$, and $\tilde{\psi}$ for the $2p_z$ orbital of a single oxygen atom. In the isosurface plot, blue color indicates positive value and yellow indicates negative value.

1. Obtaining $\hat{S}^{-1/2}$

An efficient implementation of OPAW thus requires fast application of $\hat{S}^{-1/2}$. For simplicity, we first consider the case where the augmentation spheres from different atoms do not overlap, so $\langle p_i^{(a)}|p_j^{(a')}\rangle = 0$ if $a \neq a'$. Therefore we can separately rotate the $\{p_i^{(a)}\}$ projectors around each atom so that $\hat{S}$ is transformed into

$$\hat{S} = \mathbb{I} + \sum_{i,a} |\eta_i^{(a)}\rangle o_i^{(a)} \langle \eta_i^{(a)}|, \quad (8)$$

where the rotated projectors $\{\eta_i^{(a)}\}$ are orthogonal and satisfy $\langle \eta_i^{(a)}|\eta_j^{(a')}\rangle = \delta_{ij}\delta_{a,a'}$ (see Appendix A). With this transformation, any power of $\hat{S}$ is easily expressed, e.g.,

$$\hat{S}^{-\frac{1}{2}} = \mathbb{I} + \sum_{j,a} |\eta_j^{(a)}\rangle\left[\left(1 + o_j^{(a)}\right)^{-\frac{1}{2}} - 1\right]\langle \eta_j^{(a)}|. \quad (9)$$

Since each $|\eta_j^{(a)}\rangle\langle\eta_j^{(a)}|$ is a projection operator (and all such operators are orthogonal), the proof of Eq. (9) becomes a trivial quantum mechanics exercise emanating from the simple equation $[\mathbb{I} + (a-1)P]^m = \mathbb{I} + (a^m - 1)P$, when P is a projection operator.

Next, note that the transformation operator between the orthogonal smooth molecular orbitals and the true ones is unitary,

$$|\psi_i^{(a)}\rangle = \hat{U}|\bar{\psi}_i^{(a)}\rangle, \quad \hat{U} = \hat{T}\hat{S}^{-\frac{1}{2}}, \quad (10)$$

so $\hat{U}^\dagger\hat{U} = \mathbb{I}$. Due to the unitarity, the norm of the true molecular orbitals and the orthogonal smooth ones is identical, as mentioned.

Overall, we note that except for the automatic orthogonality, the algorithm is identical to the usual PAW. In an SCF cycle, with a given one-body Hamiltonian the orthogonal molecular orbitals [the solutions of Eq. (7)] are first found; then we transform to the nonorthogonal orbitals, $\tilde{\psi}_i = \hat{S}^{-1/2}\bar{\psi}_i$, using Eq. (9), and use the usual prescription

TABLE I. Calculated band gaps of SiO_2 at different grid spacings. The Si atom PAW wave-function input data set based on GGA calculations has originally $o_1 = -1.005$, which was modified to $o_1 = -1 + \delta$; different choices of δ give essentially the same results (or slightly different for the largest δ), as does an analogous input file built based on LDA calculations where $o_1 > -1$. Note, of course, that with both data sets we did the same overall GGA (i.e., PBE) calculation; the difference was only in the PAW input functions.

Grid spacing (Bohr)		0.34	0.37	0.40	0.46
Gap (eV), LDA PAW		5.97	5.97	5.94	5.85
Gap (eV), GGA PAW	$\delta = 0.003$	5.97	5.97	5.94	5.85
	$\delta = 0.01$	5.97	5.97	5.94	5.85
	$\delta = 0.05$	5.95	5.95	5.92	5.83

of the PAW algorithm to update $v_{KS}(\boldsymbol{r}), D_{ij}^{(a)}$ in the PAW Hamiltonian.

Finally, note that the assumption of nonoverlapping augmentation spheres is quite accurate, as shown in a latter section by the agreement between our results and ABINIT. Nevertheless, it is not exact; we could go beyond it by viewing our expression for $\hat{S}^{-\frac{1}{2}}$ as a preconditioner, as shown in Appendix B, and this will be pursued in further publications.

2. Avoiding singularities

The one caveat in Eq. (9) is the formal singularity when any of the $o_i^{(a)}$ is close to or below -1. Fundamentally, a value of $o_i^{(a)} = -1$ indicates that the $\hat{S}$ operator projects out the subspace spanned by $|\eta_i^{(a)}\rangle o_i^{(a)} \langle \eta_i^{(a)}|$.

For a start, note that negative values of $o_i^{(a)}$ between -1 and 0 do not pose mathematical difficulties in our formulation but could indicate problems in the construction of the PAW parameters and in the eventual implementation, depending on the PAW code used (although they work fine in the ABINIT code used by us); see Ref. [19] for details.

In practice, for most atoms we tested, $o_i^{(a)}$ were well above -1. We did encounter one case where o_i is very close to -1—the GGA PAW parametrization of silicon taken from the website of the ABINIT PAW code [20,21], where $o_1^{(\mathrm{Si})} = -1.005$. Fortunately, the problem is trivially circumvented by replacing $o_1^{(a)}$ by $\max(o_1^{(a)}, -1 + \delta)$, where δ is a small positive number. The results are insensitive to δ. For example, for SiO_2 we tested (see Table I) three different choices, $\delta = 0.003, 0.01$, and 0.05. The two lower values of δ gave results that agree completely with those using the local density approximation (LDA) PAW file taken from the ABINIT website [20,22], where o_1 was higher than -1. Even the large shift parameter, $\delta = 0.05$, led to only a slight deviation.

We also note that numerical problems could also arise from the compensation charge being negative. A solution to this problem is discussed in the literature [19,23].

B. Application of OPAW in DFT and technical details

The OPAW algorithm is general and can be applied with any technique requiring an orthogonal Hamiltonian. Before talking about implementation of OPAW in DFT, note that a real-space implementation of OPAW will require the inner product between atomic projectors and wave functions: $\langle p_i^{(a)}|\bar{\psi}\rangle$. Such inner products are involved in determining the density matrices $\rho_{ij}^{(a)}$, as well as applying the operators $\hat{H}$ and $\hat{S}$. In a real-space formalism, the smooth wave functions $\bar{\psi}$ are defined on a 3D grid. For computational efficiency, as long as the accuracy of the results is not affected, the grid spacing for $\bar{\psi}$ should be made as large as possible. On the other hand, the projector functions are short ranged and in general show larger variation than the wave functions, so that evaluating the inner product directly on a coarse 3D grid would lead to large numerical errors.

To solve this problem, we adopted the method of Ono and Hirose [24], which connects the grid of the system with a set of finer grid points around each atom. Technical details regarding the Ono-Hirose method are given in Appendix C.

With a real-space implementation of OPAW in hand, we applied it along with the Chebyshev-filtered subspace iteration (CheFS) technique [15], resulting in an efficient DFT program (OPAW-DFT). The idea of CheFS is described in Appendix D, along with a summary of the algorithm in Appendix E.

Furthermore, since we are working with periodic systems, we did k-point sampling. A brief account of using k-point sampling with OPAW is supplied in Appendix F.

III. RESULTS AND DISCUSSION

A. Computational details

We did a set of calculations for periodic solids and report the calculated fundamental band gap. The geometries are taken from the ICSD database [25]. A 4×4×4 k-point mesh was used for each system.

We used the Perdew-Burke-Ernzerhof (PBE) generalized gradient approximation (GGA) functional in all calculations. For all calculations, the cutoff energy for the plane-wave basis set E_{cutoff} is related to the density cutoff energy by $E_{\mathrm{cutoff}}^{\mathrm{density}} = 4E_{\mathrm{cutoff}}$, as is typical in plane-wave calculations. Note that the latter is related to the grid spacing for the density by $E_{\mathrm{cutoff}}^{\mathrm{density}} = \frac{1}{2}(\frac{\pi}{dx})^2$. Thus, as usual, the grid used for the density is twice as dense (in each direction) than the spatial grid for the plane waves.

As mentioned, to assist the SCF convergence we applied a direct inversion of the iterative subspace (DIIS) procedure [26,27] when updating $v_{KS}(\boldsymbol{r})$. At times we have also applied a DIIS procedure for the Hamiltonian D_{ij} terms to assist SCF convergence.

For PAW calculations, we used the recommended atomic datasets from the ABINIT website [20]. There are two exceptions: the Sc atom, where the D_{ij} terms were large, more than 40 Hartree, and the Sr atom, where the D_{ij} terms exceed 1000 Hartree. In both cases this is due to a mismatch of the shape of the smooth and true atomic orbitals in the second, outer, d shell. To simplify, we therefore generated new PAW potentials for Sc and Sr from the ATOMPAW package [23] using only one d shell. For NCPP calculations, we used the recommended pseudopotentials from the ABINIT website [28]. More information on the PAW and NCPP datasets can be found in Supplemental Materials [29].

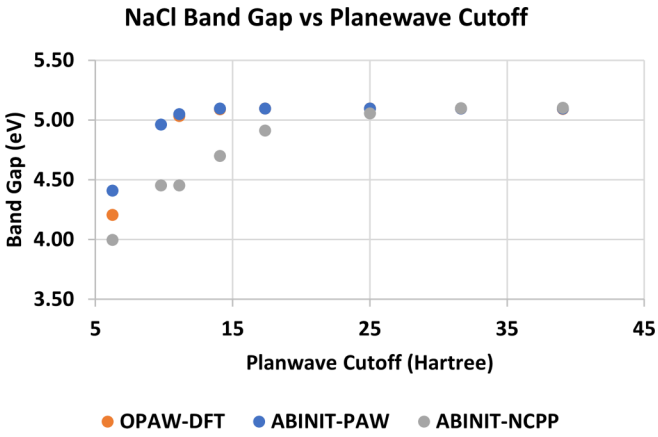

FIG. 2. Band gap vs energy cutoff for NaCl, with three methods: OPAW-DFT, ABINIT-PAW, and ABINIT-NCPP. For all the shown cutoff energies, except the lowest one, the OPAW-DFT and ABINIT-PAW results completely overlap on the scale of this graph.

B. Results

Overall, DFT calculations produce two types of information. The first is forces and total energy, important for binding and molecular dynamics. Here we concentrate on the second type of output from DFT: orbital energies and states, and here specifically the DFT highest and lowest occupied molecular orbital (HOMO and LUMO, respectively) gap. The DFT gap often serves as a preliminary approximation to the actual fundamental band gap [30], and the Kohn-Sham orbitals and their energies are the basic ingredients for most beyond-DFT methods. Future papers will also examine the total energy and forces with OPAW, as well as the shape of the band structure.

We first examine the band-gap convergence with energy cutoff for an NaCl solid. We compared OPAW-DFT with ABINIT simulations using PAW or NCPP. The results are shown in Fig. 2. For NaCl, our OPAW-DFT successfully reproduced the ABINIT results. Furthermore, the two PAW-based methods show better convergence with grid spacing than the NCPP-based method.

Second, we report the calculated fundamental band gap of a series of solids. A comparison of the converged results from ABINIT-PAW and OPAW-DFT is shown in Table II. We also present the reference value from the work of Borlido *et al.* [31]. The results indicate that OPAW-DFT reproduces ABINIT-PAW for a wide variety of systems, using generally the energy cutoff in ABINIT (with the advantage that in real space we use the localization of the projector functions, so the cost of applying the Hamiltonian on a single function scales linearly with the size of the system).

The table shows that for most solids both OPAW and ABINIT-PAW outperform NCPP, sometimes dramatically; e.g., for SiO_2, the energy cutoff required for converging the band gap is 15 Hartree for the two PAW-based methods and 29 Hartree for ABINIT-NCPP calculation; for InP the difference is even more dramatic.

To visualize the improvement in cutoff energy required for converging the fundamental band gap of solids to less than 0.05 eV, we use histograms in Fig. 3. The figure shows that

TABLE II. Calculated fundamental band gaps (in electronvolts) of selected solids. The values are reported along with the plane-wave cutoff (in Hartree) required for for a 0.05-eV gap convergence. The reference calculations use PAW in VASP [31].

	OPAW-DFT		ABINIT-PAW		ABINIT-NCPP		Ref. [31]
System	Gap	E_{cut}	Gap	E_{cut}	Gap	E_{cut}	Gap
NaCl	5.09	11	5.10	11	5.07	25	5.10
CaO	3.65	13	3.64	13	3.66	19	3.63
PbS	0.31	9	0.29	9	0.34	16	0.30
InP	0.68	10	0.65	10	0.69	23	0.71
Si	0.63	7	0.63	7	0.61	7	0.62
SiO_2	5.99	15	5.97	15	6.00	29	6.02
ScNiSb	0.28	17	0.25	15	0.29	34	0.30
NiScY	0.31	14	0.28	14	0.31	20	0.30
LiH	2.97	10	2.97	12	2.99	19	3.00
KBr	4.33	8	4.33	7	4.34	18	4.36
K_3Sb	0.75	8	0.74	5	0.75	6	0.77
$CaCl_2$	5.41	10	5.42	13	5.40	20	5.43
BN	4.46	18	4.45	24	4.53	34	4.45
$BaCl_2$	5.04	8	5.04	8	5.05	10	5.03
Ar	8.70	9	8.69	11	8.70	10	8.71
AlP	1.58	9	1.57	9	1.58	12	1.58
SrO	3.30	13	3.30	13	3.32	13	3.26

PAW gives excellent results with cutoff energies that can be as low as 7 Hartree and are generally (in the examples we studied) below 20 Hartree.

Finally, we note that in some approaches, for example, stochastic methods for DFT, time-dependent DFT (TDDFT), GW, and Bethe-Salpeter [16,17,32–36], the numerical cost is related directly to the number of spatial grid points rather than the number of plane waves; in those cases a choice of $E_{\text{cutoff}}^{\text{density}} = E_{\text{cutoff}}$ (rather than $4E_{\text{cutoff}}$) is better. The analog of Table II and Fig. 3 for this choice can be found in the Supplemental Material [29]. On average the $E_{\text{cutoff}}^{\text{density}}$ required when $E_{\text{cutoff}}^{\text{density}} = E_{\text{cutoff}}$ is much smaller than that required when using $E_{\text{cutoff}}^{\text{density}} = 4E_{\text{cutoff}}$ (as done above), i.e., setting $E_{\text{cutoff}}^{\text{density}} = E_{\text{cutoff}}$ allows a much sparser real-space grid.

IV. CONCLUSIONS

The results in the previous section show that our efficient OPAW reproduces traditional PAW. The OPAW algorithm is easy to implement and combines the best of both worlds: the lower cutoff energy typically enabled by PAW and the orthogonality of norm-conserving pseudopotential approaches.

With the efficient methodology for acting with the Hamiltonian and overlap/inverse overlap, i.e., the simple application (on any function f) of $\hat{S}f$, $\hat{H}f$ $\hat{S}^{-1}f$, $\hat{S}^{-\frac{1}{2}}f$, and $\hat{S}^{-\frac{1}{2}}\tilde{H}\hat{S}^{-\frac{1}{2}}f$, we can combine PAW with other electronic structure theory methods, including our linear scaling stochastic TDDFT and GW methods [32–34], opening the door to significant (in some cases an order of magnitude) improvements in overall grid size and the reduction of the spectral range, potentially including even larger improvements in the cost of beyond-DFT approaches. Finally, we note that an example where some of

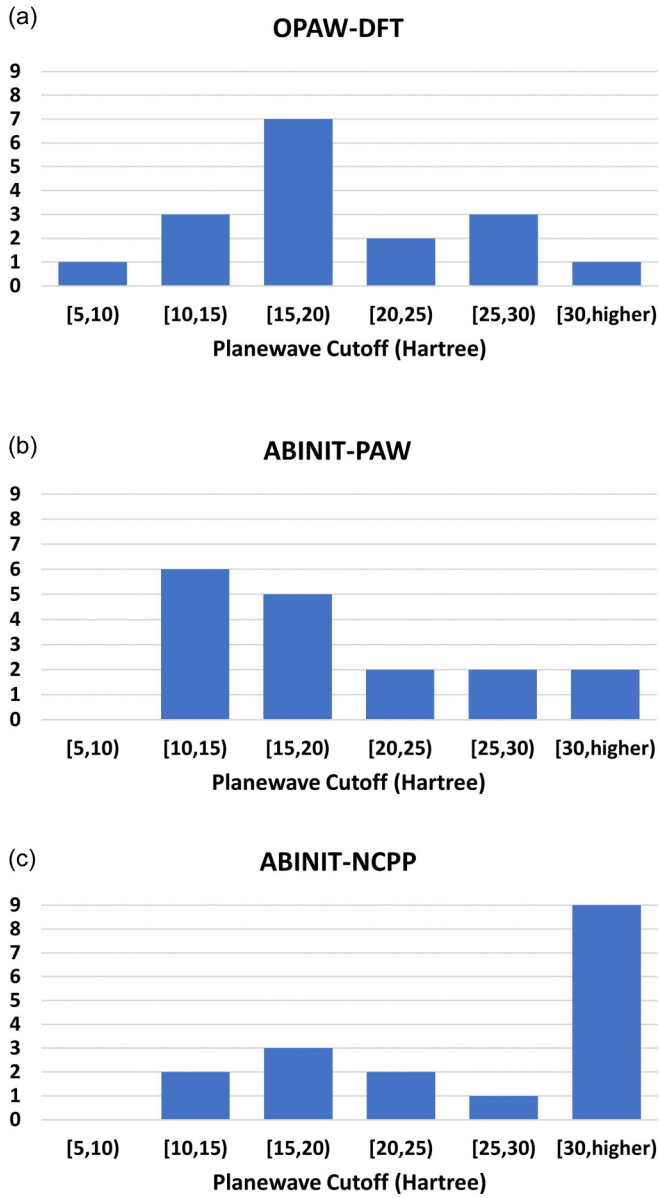

FIG. 3. Histogram of converged plane-wave cutoff for the solids in Table II, from (a) ABINIT-PAW, (b) OPAW-DFT, and (c) ABINIT-NCPP calculations.

the developments here were applied is our recent large-scale stochastic long-range exchange method for TDDFT using PAW [35].

ACKNOWLEDGMENTS

We are grateful to Roi Baer, Eran Rabani, Vojtech Vlcek, and Xu Zhang for helpful conversations. This work was supported by the NSF through Grant No. CHE-1763176. Computational resources were supplied by the XSEDE through Contract No. TG-CHE170058.

APPENDIX A: TRANSFORMATION THROUGH $\hat{S}$

We start by a proof of Eq. (1). Since the molecular orbitals are orthogonal, $\langle\psi_i|\psi_j\rangle = \delta_{ij}$, and since $|\psi_i\rangle = \hat{T}|\tilde{\psi}_i\rangle$, it follows that $\langle\tilde{\psi}_i|\hat{T}^\dagger\hat{T}|\tilde{\psi}_i\rangle = \delta_{ij}$, which, given the definition $\hat{S} \equiv \hat{T}^\dagger T$, yields Eq. (1). In the remainder we discuss the technical details of the transformation.

Given the initial operator

$$\hat{S} = \mathbb{I} + \sum_{ij,a} |p_i^{(a)}\rangle s_{ij} \langle p_j^{(a)}|, \tag{A1}$$

the first step is to orthonormalize the projectors. For each atom define a projector overlap matrix $L_{ij}^{(a)} = \langle p_i^{(a)}|p_j^{(a)}\rangle$ and diagonalize it: $L^{(a)} = U^{(a)}\lambda^{(a)}U^{(a)\dagger}$, with $U^{(a)}$ unitary. Then define a new set of projectors $\{\xi_i^{(a)}\}$,

$$|\xi_i^{(a)}\rangle = \frac{1}{\sqrt{\lambda_i^{(a)}}} \sum_j U_{ji}^{(a)} |p_j^{(a)}\rangle, \tag{A2}$$

that will be orthogonal, $\langle\xi_i^{(a)}|\xi_j^{(a)}\rangle = \delta_{ij}$. Inverting Eq. (A2) and substituting into Eq. (A1) then gives

$$\hat{S} = \mathbb{I} + \sum_{kl,a} |\xi_k^{(a)}\rangle O_{kl}^{(a)} \langle\xi_l^{(a)}|, \tag{A3}$$

where $O^{(a)} = \sqrt{\lambda^{(a)}}U^{(a)}s^{(a)}U^{(a)\dagger}\sqrt{\lambda^{(a)}}$.

The next step involves diagonalization of the matrix $O^{(a)}$, as $O^{(a)} = Q^{(a)}o^{(a)}Q^{(a)\dagger}$, with $Q^{(a)}$ unitary. It then readily follows that

$$\hat{S} = \mathbb{I} + \sum_{i,a} |\eta_i^{(a)}\rangle o_i^{(a)} \langle\eta_i^{(a)}|, \tag{A4}$$

where $|\eta_i^{(a)}\rangle = \sum_l Q_{li}^{(a)}|\xi_l^{(a)}\rangle$ are also orthogonal due to the unitarity of $Q^{(a)}$. (Note that a diagonal representation of projectors is also done in NCPP, where diagonal projectors are used in representing the nonlocal potential [5].)

Finally, when we apply the Ono-Hirose procedure, the bare $\eta_i^{(a)}$ are replaced by the processed ones, $\bar{\eta}_i^{(a)}$, as in Eq. (C3), i.e.,

$$\hat{S} = \mathbb{I} + \sum_{i,a} |\bar{\eta}_i^{(a)}\rangle o_i^{(a)} \langle\bar{\eta}_i^{(a)}|. \tag{A5}$$

These are not orthogonal on the rough grid surrounding each molecule. We therefore repeat the orthogonalization procedure, Eqs. (A1)–(A4), with the overlap matrix $L^{(a)}$ now being replaced by $\bar{L}_{ij}^{(a)} = dv \sum_{\boldsymbol{r}} \bar{\eta}_i^{(a)}(\boldsymbol{r})\bar{\eta}_j^{(a)}(\boldsymbol{r})$, leading eventually to

$$\hat{S} = \mathbb{I} + \sum_{i,a} |\bar{\xi}_i^{(a)}\rangle \bar{o}_i^{(a)} \langle\bar{\xi}_i^{(a)}|, \tag{A6}$$

where $\bar{\xi}_i^{(a)}$ are orthogonal on the rough grid, $\langle\bar{\xi}_i^{(a)}|\bar{\xi}_j^{(a)}\rangle = \delta_{ij}$.

APPENDIX B: GOING BEYOND THE NONOVERLAPPING AUGMENTATION SPHERES ASSUMPTION

In this Appendix we show how one could go beyond the nonoverlapping augmentation sphere assumption. Let us consider for simplicity expressions using $\hat{S}^{-1}$ rather than $\hat{S}^{-\frac{1}{2}}$. Then the generic relation $\hat{S}\psi = H\xi$ (the inversion of which is the crucial step in a Chebyshev propagation that iterates $\hat{S}^{-1}H$) can be rewritten as

$$(\mathbb{I} + \hat{B})\psi = \xi', \tag{B1}$$

where $\xi' \equiv \hat{S}_{NO}^{-1} H\xi$, and

$$\hat{B} \equiv \hat{S}_{NO}^{-1}\hat{S} - \mathbb{I}, \tag{B2}$$

while $\hat{S}_{NO}^{-1}$ is a nonoverlapping (NO) expression for $\hat{S}^{-1}$, as in Sec. II A,

$$\hat{S}_{NO}^{-1} = \mathbb{I} + \sum_J (o_J + 1)^{-1} - 1]P_J, \tag{B3}$$

and we use the abbreviated notation from there (but without assuming that different P_J are orthogonal). Note that this Appendix is the only place in the paper where we give an explicit subscript (NO) to expressions obtained under the nonoverlapping assumption.

Equation (B1) could be solved by a Taylor expression in B, which measures the deviation from the nonoverlapping spheres assumption. Recall that our results, obtained essentially by assuming that $B = 0$, are all quite accurate. Therefore even a single Taylor term should be extremely accurate, i.e.,

$$\psi = (\mathbb{I} - \hat{B})\xi' = \left(2\mathbb{I} - \hat{S}_{NO}^{-1}\hat{S}\right)\hat{S}_{NO}^{-1} H\xi, \tag{B4}$$

and as a reminder the definitions of the terms here come from Eqs. (4), (5), and (B3). This expression would not be much more expensive than the $B = 0$ expression we used throughout the rest of the paper ($\psi = \hat{S}_{NO}^{-1} H\xi$), since it only differs in the use of further overlaps.

APPENDIX C: THE ONO-HIROSE TRANSFORMATION WITH A SPLINE METHOD AND ITS IMPLICATIONS IN OPAW

The method of Ono and Hirose [24] is used to connect, for each atom, two sets of local grids. (The grids are specific to each atom, but for brevity we omit the atomic label in the following derivations.) One is a "rough grid" X^r, consisting of a small cubic region of the 3D wave-function grid, which encloses the augmentation sphere for the specific atom. The second is a "fine grid" X^f, spanning the same volume but with more grid points and smaller grid spacing.

The overlap of the wave functions and projectors should formally be performed on the fine grid. This requires, formally, interpolating the wave function from the rough grid (i.e., $\psi(\boldsymbol{r}), \boldsymbol{r} \in X^r$) to the fine grid, as

$$\psi(\boldsymbol{r}_f) = \sum_{\boldsymbol{r}\in X^r} B(\boldsymbol{r}_f, \boldsymbol{r})\psi(\boldsymbol{r}), \tag{C1}$$

where $B(\boldsymbol{r}_f, \boldsymbol{r})$ is a linear projection matrix. Earlier applications of the Ono-Hirose approach usually used cubic fitting [12,13,24], but here we used a spline fit.

The key observation of the Ono-Hirose approach is then that the fine-grid overlap of the atomic projectors and the wave functions,

$$\left\langle p_i^{(a)} \middle| \psi \right\rangle \equiv \sum_{\boldsymbol{r}_f \in X^f} p_i^{(a)}(\boldsymbol{r}_f)\psi(\boldsymbol{r}_f) dv_f,$$

can be written as a rough-grid overlap

$$\left\langle p_i^{(a)} \middle| \bar{\psi} \right\rangle = \sum_{\boldsymbol{r}\in X^r} \bar{p}_i^{(a)}(\boldsymbol{r})\bar{\psi}(\boldsymbol{r}) dv, \tag{C2}$$

where dv_f and dv are the fine-grid and rough-grid volume elements, and

$$\bar{p}_i^{(a)}(\boldsymbol{r}) = \frac{dv_f}{dv} \sum_{\boldsymbol{r}_f \in X^f} p_i^{(a)}(\boldsymbol{r}_f) B(\boldsymbol{r}_f, \boldsymbol{r}). \tag{C3}$$

The key practical aspect in the Ono-Hirose transformation is the smoothing matrix, $B(\boldsymbol{r}_f, \boldsymbol{r})$, connecting the fine and rough grids [Eq. (C1)]. Typically a cubic-fit approach is used; here we opted instead to use a spline fit matrix, which is separable,

$$B(\boldsymbol{r}_f, \boldsymbol{r}) = \beta(x_f, x)\beta(y_f, y)\beta(z_f, z), \tag{C4}$$

where the β matrices are obtained as explained below and depend on the element only, not the specific atoms (the derivation is done for the case of equal grid spacings, $dx = dy = dz$, and is trivially extended in the general case).

For each different element a small padding region is added around the augmented region (typically of size $r_{\text{pad}} = 0.5$ or 1 Bohr, the results do not change if either value is used). Then the set of all x points within a distance $\pm\bar{r}$ from the nucleus, where $\bar{r} = r_{\text{aug}} + r_{\text{pad}}$, is labeled as $\{x_i\}_{i=1,\ldots,n_{1d}}$. Here $n_{1d} \simeq 2\frac{\bar{r}}{dx}$ and will be typically 6–14 for our grid parameters. The set $\{x_i\}_{i=1,\ldots,n_{1d}}$ will be denoted as the rough-1d grid in the x direction.

We define then a fine 1D grid of size $n_f = 1 + (n_{1d} - 1)m_f$, where m_f is adjusted so that the fine-grid spacing, $dx_f = \frac{dx}{m_f}$, is quite small, about 0.1−0.15 Bohr (thus typically $n_f \sim$ 20–50). Further, we relabel $\beta(x_f, x)$ as a matrix, $\beta(i_f, i)$,with $1 \leqslant i \leqslant n_{1d}$, $1 \leqslant i_f \leqslant n_f$.

The $\beta(i_f, i)$ matrix is formally defined as the spline fit coefficient matrix, i.e., given a 1D function $g(x_i)$ on a rough grid, then the fine-grid spline interpolation is

$$g(x_{i_f}) = \sum_i \beta(i_f, i) g(x_i). \tag{C5}$$

While it is possible to derive $\beta(i_f, i)$ formally, the simplest approach is to use a set of δ functions. For example, to obtain $\beta(i_f, i = 1)$ use a spline-fit subroutine with a $g(x_i) = \delta_{1,i}$ input vector, feed it to a spline-fit interpolation program, and the resulting $g(x_{i_f})$ fine-grid vector will be exactly $\beta(i_f, i)$ for $i = 1$.

Given the $\beta(i_f, i)$ matrix [now again relabeled as $\beta(x_f, x)$], the next stage is to rotate each fine-grid function to the rough grid, Eq. (C3). This is easily done in stages due to the separability of Eq. (C4), so that the total cost to transform each function is only about $n_f^3 n_{1d}$, which works out to be about a one-time cost of 3000–100 000 operations for each atom and for each projector, i.e., an overall negligibly small cost.

A side note: As it stands, Eq. (C4) and therefore the remainder of our derivation only applies to orthogonal cells; however, it is trivially generalized to other crystallographic cells by replacing x, y, z by nonorthogonal coordinates that are parallel to the unit-cell directions. Finally, we note that there are alternatives to the Ono-Hirose technique, primarily the mask function technique, where the radial functions are smoothed [37].

APPENDIX D: CHEBYSHEV-FILTERED SUBSPACE ITERATION

The OPAW algorithm is general and can be applied with any technique requiring an orthogonal Hamiltonian. Here we combined our OPAW approach with the Chebyshev-filtered subspace iteration (CheFS) technique [15], resulting in an efficient DFT program (OPAW-DFT).

In CheFS, with each iteration a more refined subspace is obtained, spanned by the lower energy orbitals. The Chebyshev filter

$$F_J(\bar{H}) = C_J\left[\frac{\bar{H} - \frac{c+b}{2}\mathbb{I}}{\frac{c-b}{2}}\right]$$

selectively enhances the occupied orbitals. Here C_J is a Chebyshev polynomial of degree J (typically taken as $J \approx 20$) and its argument is a shifted Hamiltonian, where b is set to be a little bit higher than LUMO energy and c is set to be higher than the maximum eigenvalue of $\bar{H}$. The filter magnifies the weight of the lower end of the spectrum (energies below b). The number of states that the filter is operated on, labeled M, needs to be somewhat larger than the number of occupied molecular orbitals.

Obtaining the action of $F_J(\bar{H})$ on a function involves repeated applications of $\bar{H}$. In practice, we could either apply $F_J(\bar{H})$ directly or note that this is equivalent to $S^{1/2}F_J(\hat{S}^{-1}\tilde{H})S^{-1/2}$. The latter is numerically slightly more efficient, since it involves only one application of an S-type projector; practically, to obtain $\hat{S}^{-1}$ one simply needs to replace the $-\frac{1}{2}$ powers in Eq. (9) by -1. We verified that the two techniques give numerically the same results. A summary of the structure of the OPAW-DFT algorithm is given next.

APPENDIX E: SUMMARY OF ALGORITHM

For a given system, first

(1) At this stage a refers to each element in the system. From a given data set of atomic $\phi_i^{(a)}$, $\phi_i^{(a)}$, $p_i^{(a)}$ (typically contained in an "XML" file) construct the $s_{ij}^{(a)}$ matrix, as well as several small-atom matrices needed for the PAW algorithm. Construct a new set of orthogonal orbitals, $\eta_i^{(a)}$, that are a linear combination of $p_i^{(a)}$, and extract the $o_i^{(a)}$ coefficients (Appendix A). Shift $o_i^{(a)}$ to be above -1 if necessary.

(2) Starting at this next stage, a refers to each atom separately. Use the Ono-Hirose transformation (Appendix C) to form $\bar{p}_i^{(a)}(\boldsymbol{r})$, each on a small rough grid around each atom. Similarly form $\bar{\eta}_i^{(a)}$, and orthogonalize them (Appendix C) to form $\bar{\zeta}_i^{(a)}(\boldsymbol{r})$ that are orthogonal on the grid. A new set of $\bar{o}_i^{(a)}$ is then produced; again shift each $\bar{o}_i^{(a)}$ to be above -1 if necessary.

Then start the SCF algorithm, presented first in terms of the orthogonal Hamiltonian, $\bar{H}$. All expressions now refer to the sparse 3D grid.

Pick a set of M random plane-wave orbitals, $\bar{\psi}_{m\boldsymbol{k}}(\boldsymbol{q})$**.** (See Appendix F for details of the k-point sampling.) Orthogonalize them and then do the following loop until convergence:

(1) Fourier transform the orbitals to the equivalent density-based spatial grid, $\psi_{m\boldsymbol{k}}(\boldsymbol{r})$. Form $\tilde{\psi}_{m\boldsymbol{k}}(\boldsymbol{r}) = \langle \boldsymbol{r}|S^{\frac{1}{2}}|\psi_{m\boldsymbol{k}}\rangle$.

(2) From $\tilde{\psi}_{m\boldsymbol{k}}(\boldsymbol{r})$, calculate the atomic density-type matrices $\rho_{ij}^{(a)}$ and construct the smooth density, DFT potential, and the $D_{ij}^{(a)}$ terms. We adopted the routines of ABINIT for this stage.

(3) Starting at the second iteration, we apply at this stage a DIIS iteration on the DFT potential $v_{KS}(\boldsymbol{r})$ and potentially also on the D_{ij} terms.

(4) Apply the Jth-degree Chebyshev operator; symbolically assign $\bar{\psi}_{m\boldsymbol{k}} \leftarrow F_J(\bar{H}^{\boldsymbol{k}})\bar{\psi}_{m\boldsymbol{k}}$. This could be done either totally at the spatial grid level, $\bar{\psi}_{m\boldsymbol{k}}(\boldsymbol{r})$, or alternately, one could at each stage (i.e., after each application of $\bar{H}^{\boldsymbol{k}}$) transfer back to the plane-wave grid, $\bar{\psi}_{m\boldsymbol{k}}(\boldsymbol{q})$, keeping only values of $\boldsymbol{q}$ with energies below E_{cutoff} and then convert back to $\bar{\psi}_{m\boldsymbol{k}}(\boldsymbol{r})$. There is no difference in the accuracy using either approach.

(5) At the end of the Chebyshev iteration, transfer to the plane-wave grid, orthogonalize the resulting functions $\bar{\psi}_{m\boldsymbol{k}}(\boldsymbol{q})$, rotate back to $\boldsymbol{r}$ space, diagonalize the $M \times M$ matrix $h_{mm'}^{\boldsymbol{k}} = \langle \bar{\psi}_{m\boldsymbol{k}}|\bar{H}^{\boldsymbol{k}}|\bar{\psi}_{m'\boldsymbol{k}}\rangle$ in the resulting basis of M vectors, and rotate $\bar{\psi}_{m\boldsymbol{k}}(\boldsymbol{q})$ accordingly [with the resulting vectors again labeled $\bar{\psi}_{m\boldsymbol{k}}(\boldsymbol{q})$].

(6) Based on the resulting orbital energies, assign occupation numbers. Repeat the cycle until SCF convergence (typically 10–20 times).

The algorithm is only slightly modified if we choose to replace the orthogonal $\bar{H}^{\boldsymbol{k}}$ by $(\hat{S}^{\boldsymbol{k}})^{-1}\tilde{H}^{\boldsymbol{k}}$. In that case the only modifications are that we directly iterate $\tilde{\psi}_{m\boldsymbol{k}} \leftarrow F_J[(\hat{S}^{\boldsymbol{k}})^{-1}\tilde{H}^{\boldsymbol{k}}]\tilde{\psi}_{m\boldsymbol{k}}$, and at the end of each Chebyshev series we need to use general orthogonalization, so $\langle\tilde{\psi}_{m\boldsymbol{k}}|\hat{S}^{\boldsymbol{k}}|\tilde{\psi}_{m'\boldsymbol{k}}\rangle = \delta_{mm'}$.

APPENDIX F: K-POINT SAMPLING

For periodic systems, the plane-wave wave functions are given by Bloch waves, $e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\bar{\psi}_{m\boldsymbol{k}}(\boldsymbol{r})$, where $\boldsymbol{k}$ samples the first Brillouin zone, and $\bar{\psi}_{m\boldsymbol{k}}(\boldsymbol{r})$ are periodic. The modifications are therefore straightforward, exactly analogous to PAW and NCPP: Given a periodic Bloch state $\bar{\psi}_{m\boldsymbol{k}}(\boldsymbol{r})$ on a 3D unit-cell grid, define a $\boldsymbol{k}$-dependent Hamiltonian as $\bar{H}^{\boldsymbol{k}} = (\hat{S}^{\boldsymbol{k}})^{-\frac{1}{2}}\tilde{H}^{\boldsymbol{k}}(\hat{S}^{\boldsymbol{k}})^{-\frac{1}{2}}$, with (in the spatial basis)

$$(\hat{S}^{\boldsymbol{k}})^{-\frac{1}{2}}|\bar{\psi}_{m\boldsymbol{k}}\rangle = |\bar{\psi}_{m\boldsymbol{k}}\rangle + e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\sum_{i,a}|\bar{\zeta}_i^{(a)}\rangle\bar{o}_i^{(a)}\langle\bar{\zeta}_i^{(a)}|e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\bar{\psi}_{m\boldsymbol{k}}\rangle, \tag{F1}$$

i.e., in each application the $\bar{\psi}_{m\boldsymbol{k}}$ molecular orbital is multiplied once by $e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$, the projection performed for all atoms, and the resulting orbital is multiplied again by $e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}$.

Within the $\tilde{H}^{\boldsymbol{k}}$ operator, the D_{ij} terms are similarly calculated, and the kinetic energy with the kinetic energy operator obtained as usual by passing to Fourier space [i.e., producing $\bar{\psi}_{j\boldsymbol{k}}(\boldsymbol{G})$], multiplying by $\frac{1}{2}(\boldsymbol{k}+\boldsymbol{G})^2$, and transforming back.

[1] C. L. Reis, J. M. Pacheco, and J. L. Martins, Phys. Rev. B **68**, 155111 (2003).

[2] A. Willand, Y. O. Kvashnin, L. Genovese, Á. Vázquez-Mayagoitia, A. K. Deb, A. Sadeghi, T. Deutsch, and S. Goedecker, J. Chem. Phys. **138**, 104109 (2013).

[3] G. Kresse and J. Hafner, J. Phys.: Condens. Matter **6**, 8245 (1994).

[4] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).

[5] D. R. Hamann, Phys. Rev. B **88**, 085117 (2013).

[6] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[7] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[8] P. E. Blöchl, C. J. Först, and J. Schimpl, Bull. Mater. Sci. **26**, 33 (2003).

[9] N. A. W. Holzwarth, G. E. Matthews, R. B. Dunning, A. R. Tackett, and Y. Zeng, Phys. Rev. B **55**, 2005 (1997).

[10] A. Tackett, N. Holzwarth, and G. Matthews, Comput. Phys. Commun. **135**, 348 (2001).

[11] M. Torrent, F. Jollet, F. Bottin, G. Zerah, and X. Gonze, Comput. Mater. Sci. **42**, 337 (2008).

[12] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. Hansen *et al.*, J. Phys.: Condens. Matter **22**, 253202 (2010).

[13] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Phys. Rev. B **71**, 035109 (2005).

[14] C. J. Pickard and F. Mauri, Phys. Rev. B **63**, 245101 (2001).

[15] Y. Zhou, J. R. Chelikowsky, and Y. Saad, J. Comput. Phys. **274**, 770 (2014).

[16] R. Baer, D. Neuhauser, and E. Rabani, Phys. Rev. Lett. **111**, 106402 (2013).

[17] D. Neuhauser, R. Baer, and E. Rabani, J. Chem. Phys. **141**, 041102 (2014).

[18] X. Gonze, B. Amadon, G. Antonius, F. Arnardi, L. Baguet, J.-M. Beuken, J. Bieder, F. Bottin, J. Bouchet, E. Bousquet, N. Brouwer, F. Bruneval, G. Brunin, T. Cavignac, J.-B. Charraud, W. Chen, M. Côté, S. Cottenier, J. Denier, G. Geneste *et al.*, Comput. Phys. Commun. **248**, 107042 (2020).

[19] N. Holzwarth, Comput. Phys. Commun. **243**, 25 (2019).

[20] F. Jollet, M. Torrent, and N. A. W. Holzwarth, Comput. Phys. Commun. **185**, 1246 (2014).

[21] https://www.abinit.org/ATOMICDATA/014-si/Si.LDA_PW-JTH.xml.

[22] https://www.abinit.org/ATOMICDATA/014-si/Si.GGA_PBE-JTH.xml.

[23] N. Holzwarth, A. Tackett, and G. Matthews, Comput. Phys. Commun. **135**, 329 (2001).

[24] T. Ono and K. Hirose, Phys. Rev. Lett. **82**, 5016 (1999).

[25] https://icsd.fiz-karlsruhe.de/.

[26] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).

[27] P. Pulay, J. Comput. Chem. **3**, 556 (1982).

[28] https://www.abinit.org/psps_abinit.

[29] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.102.195118 for more information on the PAW and NCPP datasets and analog of Table II and Fig. 3 for the case where density cutoff equals planewave cutoff.

[30] C.-G. Zhan, J. A. Nichols, and D. A. Dixon, J. Phys. Chem. A **107**, 4184 (2003).

[31] P. Borlido, T. Aull, A. W. Huran, F. Tran, M. A. Marques, and S. Botti, J. Chem. Theory Comput. **15**, 5069 (2019).

[32] Y. Gao, D. Neuhauser, R. Baer, and E. Rabani, J. Chem. Phys. **142**, 034106 (2015).

[33] D. Neuhauser, Y. Gao, C. Arntsen, C. Karshenas, E. Rabani, and R. Baer, Phys. Rev. Lett. **113**, 076402 (2014).

[34] V. Vlcek, E. Rabani, D. Neuhauser, and R. Baer, J. Chem. Theory Comput. **13**, 4997 (2017).

[35] X. Zhang, G. Lu, R. Baer, E. Rabani, and D. Neuhauser, J. Chem. Theory Comput. **16**, 1064 (2020).

[36] E. Rabani, R. Baer, and D. Neuhauser, Phys. Rev. B **91**, 235302 (2015).

[37] M. Tafipolsky and R. Schmid, J. Chem. Phys. **124**, 174102 (2006).