Real Space Norm-Conserving Projector-Augmented-Wave Method

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

(Dated: 29 July 2020)

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

I. INTRODUCTION

A plane wave basis set is natural when studying periodic systems with DFT and post-DFT methods. Convergence with 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 pseudo-potential, and the resulting pseudo valence states are non-oscillatory. One way to circumvent this problem is to replace the effect of the chemically inert core states by an effective pseudo-potential, and the resulting pseudo valence states are non-oscillatory. DFT using pseudo-potentials 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 pseudo-potentials (NCPP), treatment of first-row elements and transition metals is still computationally demanding, due to the localized nature of 2p and 3d orbitals.

The projector-augmented wave (PAW) method proposed by Blöchl seeks to make softer pseudo wavefunctions by relaxing the norm-conserving condition. There are several different implementations of the PAW method with many successful applications.

In addition to the reduced kinetic energy cutoff, an advantage of the PAW method is that it provides 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 wavefunctions in the core region.

The PAW method is based on a map between the smoothed pseudo wavefunctions \( \{ \tilde{\psi}_i \} \) and the all electron wavefunctions \( \{ \psi_i \} \). Unlike NCPP where the wavefunctions retain their orthogonality, the pseudo wavefunctions in PAW satisfy a generalized orthogonality condition:

\[
\langle \tilde{\psi}_i | \hat{S} | \tilde{\psi}_j \rangle = \delta_{ij},
\]

which leads to a generalized eigenproblem: \( \hat{H} \tilde{\psi}_i = \epsilon_i \tilde{\psi}_i \), where we introduced the 1-body Hamiltonian \( \hat{H} \) and overlap operator \( \hat{S} \) (both detailed later).

The fact that the pseudo-orbitals are not orthogonal complicates, however, 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., stochastically DFT methods 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 non-orthogonality problem by an efficient numerical transformation of the PAW problem to an orthogonal one,

\[
\left( \hat{S}^{-\frac{1}{2}} \hat{H} \hat{S}^{-\frac{1}{2}} \right) \tilde{\psi}_i = \epsilon_i \tilde{\psi}_i
\]

with \( \tilde{\psi}_i = \hat{S}^{1/2} \bar{\psi}_i \) 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 Norm-Conserving PAW (NCPAW). In addition to orthogonality, NCPAW is also efficient because it is implemented in real space, exploiting the localization of atomic projector functions and partial waves.

NCPAW provides a general framework, and can be combined with different electronic structure methods. Here we apply the method with a deterministic DFT approach, concentrating on the fundamental band gap of solids. We show below excellent agreement with PAW calculations from the ABINIT package. We also demonstrate that for many systems, PAW and NCPAW converge with grid spacing faster than NCPP.

Section II presents the NCPAW theory. A brief discussion of the application to a deterministic modified Chebyshev algorithm follows in Section III. Results are presented in Section IV, and conclusions follow in Section V. Technical details are deferred to appendices.
II. THEORY

A. Norm-conserving projector augmented wave

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

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

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., $\sum_i |\tilde{\phi}_i^{(a)}\rangle \langle p_i^{(a)} | \approx 1$ in the sphere.

With some derivations, one arrives at the working equation of PAW, the generalized eigenproblem $\hat{H} |\tilde{\psi}_i\rangle = \varepsilon_i |\tilde{\psi}_i\rangle$ where

$$\hat{H} = \hat{T}\hat{T}^\dagger + \sum_{ij,a} |\bar{p}_i^{(a)}\rangle \langle \bar{p}_j^{(a)} |,$$

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

$$\hat{H} = -\frac{1}{2} \nabla^2 + \nu_K S(r) + \sum_{ij,a} |\bar{p}_i^{(a)}\rangle \langle \bar{D}_{ij}^{(a)} | \bar{p}_j^{(a)} |.$$

The expressions for the Kohn-Sham effective potential $\nu_K S(r)$ and for $\bar{D}^{(a)}_{ij}$ are found in various references. While $\bar{s}^{(a)}_{ij}$ are only atom-depended, $\nu_K S(r)$ and $\bar{D}^{(a)}_{ij}$ both depend on the on-site PAW atomic density matrices: $\bar{p}^{(a)}_{ij} = \sum_m |\bar{p}_i^{(a)}\rangle \langle \bar{p}_m^{(a)} | \psi_m(r) | \bar{p}_j^{(a)} \rangle$, as well as the smooth density $\bar{n}(r) = \sum_m |\bar{\psi}_m(r)|^2$ 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 SCF cycle.

In many applications, however, it is desirable to work with an orthonormal collection of wavefunctions. As mentioned in the introduction, this can be achieved by the transformation: $\tilde{\psi}_i = \hat{S}^{1/2} |\tilde{\psi}_i\rangle$, resulting in

$$\hat{H} |\tilde{\psi}_i\rangle = \varepsilon_i |\tilde{\psi}_i\rangle,$$

where $\hat{H} = \hat{S}^{-1/2} \hat{T} \hat{H} \hat{S}^{-1/2}$.

An efficient implementation of NCPAW thus requires fast application of $\hat{S}^{-1/2}$. This is easily done by realizing first that the augmentation spheres from different atoms generally do not overlap, so: $\langle \bar{p}_i^{(a)} | \bar{p}_j^{(a')} \rangle = 0$ if $a \neq a'$.

| 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    | 5.97 | 5.97 | 5.94 | 5.85 |
| $\delta = 0.005$     | 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 |

**TABLE I:** Calculated band gaps of SiO$_2$ at different grid spacings. The Si atom PAW wavefunction input data set based on GGA calculations has originally $\alpha_1 = -1.005$, which was modified to $\alpha_1 = -1 + \delta$; different choices of $\delta$ give essentially the same results (or slightly different for the largest $\delta$) as does an analogous input file built based on LDA calculations where $\alpha_1 > -1$. 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.

Therefore, we can separately rotate the $\{\bar{p}_i^{(a)}\}$ projectors around each atom, so that the $\hat{S}$ is transformed into:

$$\hat{S} = 1 + \sum_{i,a} |\eta_i^{(a)}\rangle \langle \eta_i^{(a)} |,$$

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.,

$$\hat{S}^{-\frac{1}{2}} = 1 + \sum_{i,a} |\eta_i^{(a)}\rangle \left( \left( 1 + o_i^{(a)} \right)^{-\frac{1}{2}} - 1 \right) \langle \eta_i^{(a)} |.$$

The one caveat in this expression is the formal singularity when any of the $o_i^{(a)}$ is close to or below $-1$. Fundamentally, a value of $o^{(a)}_1 = -1$ indicates that the $\hat{S}$ operator projects out the subspace spanned by $|\eta_i^{(a)}\rangle \langle \eta_i^{(a)} |$.

In practice, for most atoms we tested, $o^{(a)}_1$ were well above $-1$. We did encounter one case where $o_1$ is very close to $-1$ – the GGA PAW parametrization of silicon taken from the website of the ABINIT PAW code where $o^{(Si)}_1 = -1.005$. Fortunately the problem is trivially circumvented by replacing $o^{(a)}_1$ by max($o^{(a)}_1, -1 + \delta$) where $\delta$ is a small positive number. The results are insensitive to $\delta$. 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 $\delta$ gave results that agree completely with those using the LDA PAW file taken from the ABINIT website where $o_1$ was higher than $-1$. Even the large shift parameter, $\delta = 0.05$, led to only a slight deviation.

Note also that the transformation operation between the orthogonal smooth molecular orbitals and the true ones is unitary

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

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

A key component in PAW is the inner product between an atomic projector and a wavefunction: $\langle \bar{p}_{1}^{(a)} | \bar{\psi} \rangle$. Such inner products are involved in determining the density matrices $\rho_{ij}^{(a)}$, as well as applying the operators $\hat{H}$ and $\hat{S}^k$. In a real space formalism, the smooth wavefunctions $\bar{\psi}$ 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 range and in general show larger variation than the wavefunctions, so that evaluating the inner product directly on a coarse 3D grid would lead to large numerical errors.

To solve this problem, we use the method of Ono and Hirose\cite{Hirose_1993} 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 wavefunction 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 wavefunctions and projectors should formally be performed on the fine grid. This requires, formally, interpolating the wavefunction from the rough grid (i.e., $\psi(r)$, $r \in X^r$) to the fine grid, as

$$\psi(r_f) = \sum_{r \in X^f} B(r_f, r) \psi(r),$$

where $B(r_f, r)$ is a linear projection matrix. Earlier applications of the Ono-Hirose approach usually used cubic fitting\cite{Rohringer_1994, Hirose_1995} but here we used a spline fit, as explained in Appendix B.

The key observation of the Ono-Hinose approach is then that the fine-grid overlap of the atomic projectors and the wavefunctions,

$$\langle p_{i}^{(a)} | \psi \rangle \equiv \sum_{r \in X^f} p_{1}^{(a)}(r_f) \psi(r_f) dv_f,$$

can be written as a rough-grid overlap

$$\langle p_{i}^{(a)} | \bar{\psi} \rangle = \sum_{r \in X^r} \bar{p}_{1}^{(a)}(r) \bar{\psi}(r) dv,$$

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

$$\bar{p}_{1}^{(a)}(r) = \frac{dv_f}{dv} \sum_{r \in X^f} p_{1}^{(a)}(r_f) B(r_f, r).$$

The rewriting of the overlaps to be on the rough grid implies that we can use these rough-grid projector functions in the Hamiltonian, which becomes:

$$\hat{H} = -\frac{\nabla^2}{2} + v_{KS} + \sum_{ij,a} |\bar{p}_{1}^{(a)}(i)|^2 \delta_{ij}^{(a)}.$$

Note that the ket/\bra here refer to the specific discrete (rough) 3D grid, and not to continuum wavefunctions. Thus, the conclusion of the Ono-Hirose approach is that the modified Hamiltonian has the same matrix elements with non-interpolated wavefunctions, as the original Hamiltonian would have had with a fine-grid calculation and interpolated wavefunctions.

Next, in the $S$ matrix (Eq. (1)) we similarly replace $\bar{q}_{1}^{(a)}$ by rough-grid functions $\bar{q}_{1}^{(a)}$, calculated analogously to Eq. (1). Note, however, that these $\bar{q}_{1}^{(a)}$ vectors are not exactly orthogonal on the rough grid (though they are quite close), i.e.,

$$\langle \bar{q}_{i}^{(a)} | \bar{q}_{j}^{(a)} \rangle \equiv dv \sum_{r \in X^r} \bar{q}_{i}^{(a)}(r) \bar{q}_{j}^{(a)}(r) \neq \delta_{ij}.$$  (13)

We therefore need to reorthogonalize them, yielding new functions, $\tilde{\zeta}_{i}^{(a)}(r)$, that are linear combinations of $\bar{q}_{i}^{(a)}(r)$ and therefore of $\bar{p}_{i}^{(a)}(r)$, that are orthogonal on the rough grid

$$\langle \tilde{\zeta}_{i}^{(a)} | \tilde{\zeta}_{j}^{(a)} \rangle = dv \sum_{r \in X^r} \tilde{\zeta}_{i}^{(a)}(r) \tilde{\zeta}_{j}^{(a)}(r) = \delta_{ij}.$$  (14)

Appendix A gives further details. The final expression for $S^{-1/2}$ is then

$$\hat{S}^{-\frac{1}{2}} = I + \sum_{i,a} \tilde{\zeta}_{i}^{(a)} \left( \left( 1 + \bar{\alpha}_{i}^{(a)} \right)^{-\frac{1}{2}} - 1 \right) \langle \tilde{\zeta}_{i}^{(a)} \rangle,$$

where the coefficients $\bar{\alpha}_{i}^{(a)}$ are the eigenvalues of the final overlap operator, see Appendix A.

The NCPAW method is therefore simple: solve the orthogonal eigenvalue problem (Eq. (1)) with $\hat{H} = \hat{S}^{-\frac{1}{2}} \hat{H} \hat{S}^{-\frac{1}{2}}$ and $\hat{S}$ and $\hat{S}^{-\frac{1}{2}}$ defined by Eqs. (12) and (15). There are various approaches to the solution with such an orthogonal Hamiltonian, and the specific approach we use here is discussed in Section III.

C. k-point sampling

For periodic systems, the wavefunctions are given by Bloch waves, $e^{ik \cdot r} \tilde{\psi}_{mk}(r)$ where $k$ samples the first Brillouin zone, and $\tilde{\psi}_{mk}(r)$ are periodic. The modifications are therefore straightforward, exactly analogous to PAW and NCPP: Given a periodic Bloch state $\psi_{mk}(r)$ on a 3D unit cell grid, define a $k$-dependent Hamiltonian as

$$\hat{H}^k = \left( \hat{S}^k \right)^{-\frac{1}{2}} \hat{H} \left( \hat{S}^k \right)^{-\frac{1}{2}},$$

with

$$\left( \hat{S}^k \right)^{-\frac{1}{2}} | \tilde{\psi}_{mk} \rangle = | \tilde{\psi}_{mk} \rangle + e^{-ik \cdot r} \sum_{i,a} \tilde{\zeta}_{i}^{(a)} \bar{\alpha}_{i}^{(a)} e^{ik \cdot r} \tilde{\psi}_{mk}.$$

I.e., in each application the $\tilde{\psi}_{mk}$ molecular orbital is multiplied once by $e^{ik \cdot r}$, the projection performed for all atoms, and the resulting orbital is multiplied again by $e^{-ik \cdot r}$. 
Within the $\tilde{H}^k$ operator, the projection terms in Eq. 12 are similarly calculated, and the kinetic energy with the kinetic energy operator obtained as usual by passing to Fourier space (i.e., producing $\tilde{\psi}_{jk}(G)$), multiplying by $\frac{1}{2}(k + G)^2$, and transforming back.

### III. APPLICATION OF NCPAW IN DFT

#### A. Chebyshev-filtered subspace iteration

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

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

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

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 $\tilde{H}$. 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 $F_J(\tilde{H})$ on a function involves repeated applications of $\tilde{H}$. In practice, we could either apply $F_J(\tilde{H})$ directly, or note that this is equivalent to $S^{1/2}F_J(\tilde{S}^{-1}\tilde{H})S^{-1/2}$. The latter is numerically slightly more efficient, since it involves only one application of an $S$-type projector; practically, to obtain $\tilde{S}^{-1} \bar{S}$ one simply need to replace the $-\frac{1}{2}$ powers in Eq. (15) by $-1$.

We verified that the two techniques give numerically the same results.

A summary of the structure of the NCPAW-DFT algorithm is given next.

#### B. Algorithm summary

For a given system, first,

- 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 $S^{(a)}$ 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.

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

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

Pick a set of $M$ random orbitals, $\tilde{\psi}_{mk}(r)$. Orthogonalize them, and then do the following loop till convergence:

- Form $\tilde{\psi}_{mk}(r) = S^\frac{1}{2} |\psi_{mk}\rangle$.

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

- Apply the $J$-th degree Chebyshev operator; symbolically assign $\tilde{\psi}_{mk} \leftarrow F_J(\tilde{H}^k)\tilde{\psi}_{mk}$

- Orthogonalize the resulting functions $\tilde{\psi}_{mk}$, diagonalize the $M \times M$ matrix $h_{mm'}^{k} = \langle \tilde{\psi}_{mk} | \tilde{H}^k | \tilde{\psi}_{mk'} \rangle$ in the resulting basis of $M$ vectors, and rotate them accordingly (with the resulting vectors again labeled $\tilde{\psi}_{mk}(r)$).

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

The algorithm is only slightly modified if we choose to replace the orthogonal $\tilde{H}^k$ by $\left(\tilde{S}^k\right)^{-1} \tilde{H}^k$. In that case the only modifications are that we directly iterate $\tilde{\psi}_{mk} \leftarrow F_J \left(\left(\tilde{S}^k\right)^{-1} \tilde{H}^k\right) \tilde{\psi}_{mk}$, and we need to use general orthogonalization, so $\langle \tilde{\psi}_{mk} | \tilde{S}^k | \tilde{\psi}_{mk'} \rangle = \delta_{mm'}$.

### IV. RESULTS AND DISCUSSION

#### A. Computational details

We did a set of calculations for periodic solids and report the calculated fundamental band gap. The geometries of are taken from the ICSD database. A $4 \times 4 \times 4$ k-point mesh was used for each system.
We used the PBE GGA functional in all calculations. For all calculations, the cutoff energy for the plane wave basis set is taken as the maximum allowed by ABINIT for a given grid spacing, i.e., \( E_{\text{cutoff}} = \frac{1}{2} \left( \frac{\pi}{dx} \right)^2 \); this is to ensure that the only convergence parameter is the grid spacing. We also note that the same grid was used for the density and for the wavefunctions. This turns out to be the most efficient choice (i.e., leading to optimized results for a given total numerical effort).

A side note is that, as usual, while we view and define the wavefunctions as grid-based, they can be viewed as a plane wave basis set; the number of plane wave coefficients \( \psi_{mk}(G) \) is exactly the same as the grid ones, \( \psi_{mk}(r) \), related to it by Fourier transform.

As mentioned, to assist the SCF convergence we applied a DIIS procedure\(^\text{22,23}\) when updating \( \nu_{KS}(r) \). At times, we have also applied a DIIS procedure for the Hamiltonian \( D_{ij} \) coefficients to assist SCF convergence.

For PAW calculations, we used the recommended atomic datasets from the ABINIT website\(^\text{25}\). There are two exceptions: the Sc atom, where the \( D_{ij} \) coefficients were large, more than 40 Hartree, and the Sr atom, where the \( D_{ij} \) coefficients exceed 1000 Hartree. In both cases this gap 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\(^\text{23}\) using only one d-shell. For NCPP calculations, we used the recommended pseudo-potentials from the ABINIT website\(^\text{25}\).

### 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 HOMO-LUMO gap. The DFT gap often serves as preliminary approximation to the actual fundamental band gap\(^\text{29}\) and the Kohn-Sham orbitals are the basic ingredients for most beyond-DFT methods. Future papers will also examine the total energy and forces with NCPAW, as well as the shape of the band structure.

We first examine the band-gap convergence with grid spacing for an NaCl solid. We compared NCPAW-DFT with ABINIT simulations using PAW or NCPP. The results are shown in Figure 1. For NaCl, our NCPAW-DFT successfully reproduced the ABINIT results. Furthermore, the two PAW-based methods show better convergence with grid spacing than the NCPP-based method.

Secondly, we report the calculated fundamental band gap of a series of solids. A comparison of the converged results from ABINIT-PAW and NCPAW-DFT is shown in Table II. We also present the reference value from the work of Borlido et al.\(^\text{27}\). The results indicate that our NCPAW-DFT method is able to reproduce ABINIT-PAW for a wide variety of systems, using generally the same grid spacing as in ABINIT (with the advantage, of course, of a formally linear-scaling application of \( H \)).

The table shows that for most solids PAW outperforms NCPP, sometimes dramatically; e.g., for the two Ni solid compounds, the \( dx \) required in NCPAW is at 0.40-0.42Bohr, vs. 0.24-0.27Bohr in NCPP.

To visualize the improvement in grid spacing required for converging the fundamental band gap of solids to less than 0.05 eV, we use histograms in Figure 2. The figure shows that PAW gives excellent results with grid spacings that can be as high as 0.65Bohr, and are generally (in the examples we studied) above 0.4Bohr.

![NaCl Band Gap vs Grid spacing](image)

**FIG. 1:** Band gap vs. grid spacing for NaCl, with three methods: ABINIT-PAW, NCPAW-DFT and ABINIT-NCPP. Recall that the grid spacing corresponds to the cutoff energy, by \( E_{\text{cutoff}} = \frac{1}{2} \left( \frac{\pi}{dx} \right)^2 \). For example, for \( dx = 0.6 \text{Bohr} \), \( E_{\text{cutoff}} = 13.7 \text{Hartree} \).

| System | ABINIT-PAW | NCPAW-DFT | ABINIT-NCPP | Reference\(^\text{27}\) |
|--------|-------------|------------|--------------|-------------------|
| NaCl   | 5.09        | 5.10       | 5.07         | 5.10              |
| CaO    | 3.65        | 3.64       | 3.66         | 3.63              |
| PbS    | 0.31        | 0.29       | 0.34         | 0.30              |
| InP    | 0.68        | 0.65       | 0.69         | 0.71              |
| Si     | 0.63        | 0.63       | 0.61         | 0.68              |
| SiO\(_2\) | 5.99      | 5.97       | 6.00         | 6.02              |
| ScNiSb | 0.28        | 0.25       | 0.29         | 0.34              |
| LiH    | 2.97        | 2.97       | 2.99         | 3.00              |
| KBr    | 4.33        | 4.33       | 4.34         | 4.36              |
| K\(_2\)Sb | 0.75   | 0.74       | 0.75         | 0.77              |
| CaCl\(_2\) | 5.41    | 5.42       | 5.40         | 5.43              |
| BN     | 4.46        | 4.45       | 4.53         | 4.45              |
| BaCl\(_2\) | 5.04     | 5.04       | 5.05         | 5.03              |
| Ar     | 8.70        | 8.69       | 8.70         | 8.71              |
| AlP    | 1.58        | 1.57       | 1.58         | 1.58              |
| SrO    | 3.30        | 3.30       | 3.32         | 3.26              |

**TABLE II:** Calculated fundamental band gaps (in eV) of selected solids. The values are reported along with the grid spacing (in Bohr) required for a 0.05eV gap convergence.
stochastic long-range exchange method for TDDFT using developments here were applied is our recent large scale beyond-DFT approaches. Overall grid size and the reduction of the spectral range, in some cases an order of magnitude, improvements in GW methods, including our linear scaling stochastic TDDFT.\\\(\hat{S}\\) Hamiltonian and overlap/inverse overlap, i.e., the simple conserving approaches.\\\(c\) NCPAW-DFT; and c) ABINIT-NCPP calculations.

FIG. 2: Histogram of converged grid spacings for the solids in Table II from a) ABINIT-PAW; b) NCPAW-DFT; and c) ABINIT-NCPP calculations.

V. CONCLUSIONS

The results in the previous section show that our efficient grid-based NCPAW reproduces traditional PAW. The NCPAW algorithm is easy implement and combines the best of both worlds: the large-grid spacing typically enabled by PAW and the orthogonality of norm-conserving approaches.

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

Finally, we note that an example where some of the developments here were applied is our recent large scale stochastic long-range exchange method for TDDFT using PAW[31].

ACKNOWLEDGEMENTS

We are grateful to Roi Baer, Eran Rabani, Vojtech Vlcek and Xu Zhang for helpful conversations. This work was supported by the NSF CHE-1763176 grant. Computational resources were supplied through the XSEDE allocation TG-CHE170058.

APPENDIX A: TRANSFORMATION OF \(\hat{S}\)

Given the initial operator:

\[
\hat{S} = 1 + \sum_{ij,\alpha} |p_i^{(\alpha)}\rangle s_{ij} \langle p_j^{(\alpha)}|,
\]

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

\[
|\xi_i^{(\alpha)}\rangle = \frac{1}{\sqrt{\lambda_i^{(\alpha)}}} \sum_j U_{ji}^{(\alpha)} |p_j^{(\alpha)}\rangle
\]

that will be orthogonal, \(\langle \xi_i^{(\alpha)}| \xi_j^{(\alpha)}\rangle = \delta_{ij}\). Inverting Eq. (18) and substituting into Eq. (17) then gives:

\[
\hat{S} = 1 + \sum_{kl,\alpha} |\xi_k^{(\alpha)}\rangle O^{(\alpha)}_{kl} |\xi_l^{(\alpha)}\rangle
\]

where \(O^{(\alpha)} = \sqrt{\lambda^{(\alpha)}} U^{(\alpha)} \hat{S}^{(\alpha)} U^{(\alpha)^\dagger}/\sqrt{\lambda^{(\alpha)}}\).

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

\[
\hat{S} = 1 + \sum_{i,\alpha} |\eta_i^{(\alpha)}\rangle a_i^{(\alpha)} |\eta_i^{(\alpha)}\rangle,
\]

where \(|\eta_i^{(\alpha)}\rangle = \sum_k Q_{ki}^{(\alpha)} |\xi_k^{(\alpha)}\rangle\) are also orthogonal due to the unitarity of \(Q^{(\alpha)}\).

Finally, when we apply the Ono-Hirose procedure, the bare \(\eta_i^{(\alpha)}\) are replaced by the processed ones, \(\tilde{\eta}_i^{(\alpha)}\) as in Eq. (11), i.e.,

\[
\hat{S} = 1 + \sum_{i,\alpha} |\tilde{\eta}_i^{(\alpha)}\rangle a_i^{(\alpha)} |\tilde{\eta}_i^{(\alpha)}\rangle.
\]

These are not orthogonal on the rough-grid surrounding each molecule, per Eq. (13). We therefore repeat the orthogonalization procedure, Eqs. (17)-(20), with the overlap matrix \(L^{(\alpha)}\) now being replaced by \(\bar{L}_{ij}^{(\alpha)} = dv \sum_r \bar{\eta}_i^{(\alpha)}(r) \bar{\eta}_j^{(\alpha)}(r)\), leading eventually to

\[
\hat{S} = 1 + \sum_{i,\alpha} |\tilde{\zeta}_i^{(\alpha)}\rangle a_i^{(\alpha)} |\tilde{\zeta}_i^{(\alpha)}\rangle,
\]

where \(\tilde{\zeta}_i^{(\alpha)}\) are orthogonal on the rough grid (Eq. (14)).
APPENDIX B: THE ONO-HIROSE TRANSFORMATION WITH A SPLINE METHOD AND ITS IMPLICATIONS IN NCPAW

The key practical aspect in the Ono-Hirose transformation is the smoothing matrix, \( B(\mathbf{r}_f, \mathbf{r}) \), connecting the fine and rough grids (Eq. (9)). Typically a cubic-fit approach is used; here we opted instead to use a spline fit matrix, which is separable.

\[
B(\mathbf{r}_f, \mathbf{r}) = \beta(x_f, x)\beta(y_f, y)\beta(z_f, z),
\]

where the \( \beta \) 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 \( r_{pad} = 0.5 \) or 1Bohr, the results do not change if either value is used). Then the set of all \( x \) points within a distance \( \pm r \) from the nucleus, where \( r = r_{aug} + r_{pad} \), is labeled as \( \{ x_i \}_{i=1,...,n_{id}} \). Here, \( n_{id} \approx 2 \beta_f \), and will be typically 6-14 for our grid parameters. The set \( \{ x_i \}_{i=1,...,n_{id}} \) will be denoted as the rough-1d grid in the \( x \) direction.

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

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

\[
g(x_{i_f}) = \sum_i \beta(i_f, i) g(x_i).
\]

While it is possible to derive \( \beta(i_f, i) \) formally, the simplest approach is to use a set of delta-functions. For example, to obtain \( \beta(i_f, i = 1) \) use a spline fit subroutine with a \( g(x_i) = \delta_{i,1} \) input vector, feed it to a spline-fit interpolation program, and 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. (17). This is easily done in stages due to the separability of Eq. (23), so that the total cost to transform each function is only about \( n_f^3 n_{id} \), which works out to be about a one-time cost of 3,000-100,000 operations for each atom and for each projector, i.e., an overall negligibly small cost.

1Carlos L Reis, JM Pacheco, and José Luís Martins. First-principles norm-conserving pseudopotential with explicit incorporation of semicore states. Physical Review B, 68(15):155111, 2003.

2Alex Willand, Yaroslav O Kvasnič, Luigi Genovese, Álvaro Vázquez-Mayagoitia, Arpan Krishna Deb, Ali Sadeghi, Thierry Deutsch, and Stefan Goedecker. Norm-conserving pseudopotentials with chemical accuracy compared to all-electron calculations. The Journal of chemical physics, 138(10):104109, 2013.

3G Kresse and J Hafner. Norm-conserving and ultrasoft pseudopotentials for first-row and transition elements. Journal of Physics: Condensed Matter, 6(40):8245–8257, 1994.

4Peter E Blöchl. Projector augmented-wave method. Physical review B, 50(24):17953, 1994.

5Georg Kresse and Daniel Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. Physical review b, 59(3):1758, 1999.

6Peter E Blöchl, Clemens J Först, and Johannes Schimpf. Projector augmented wave method: ab inito molecular dynamics with full wave functions. Bulletin of Materials Science, 26(1):33–41, 2003.

7NAW Holwarth, GE Matthews, RB Dunning, AR Tackett, and Y Zeng. Comparison of the projector augmented-wave, pseudopotential, and linearized augmented-plane-wave formalisms for density-functional calculations of solids. Physical Review B, 55(4):2005, 1997.

8AR Tackett, NAW Holwarth, and GE Matthews. A projector augmented wave (paw) code for electronic structure calculations, part ii: pwakeup for periodic solids in a plane wave basis. Computer Physics Communications, 135(3):348–376, 2001.

9Marc Torrent, F Jollet, Francois Bottin, G Zerah, and Xavier Gonze. Implementation of the projector augmented-wave method in abinit code: Application to the study of iron under pressure. Computational Materials Science, 42(2):337–351, 2008.

10JE Enkovaara, Carsten Rostgaard, J Jørgen Mortensen, Jingzhe Chen, M Dulsak, Lara Ferrighi, Jeppe Gavnholt, Christian Glinsvad, V Haikola, HA Hansen, et al. Electronic structure calculations with gpaw: a real-space implementation of the projector augmented-wave method. Journal of Physics: Condensed Matter, 22(25):253202, 2010.

11Jens Jørgen Mortensen, Lars Bruno Hansen, and Karsten Wedel Jacobsen. Real-space grid implementation of the projector augmented wave method. Physical Review B, 71(3):035109, 2005.

12Chris J Pickard and Francesco Mauri. All-electron magnetic response with pseudopotentials: Nnr chemical shifts. Physical Review B, 63(24):245101, 2001.

13Yunkai Zhou, James R Chelikowsky, and Yousef Saad. Chebyshev-filtered subspace iteration method free of sparse diagonalization for solving the kohn-sham equation. Journal of Computational Physics, 274:770–782, 2014.

14Roi Baer, Daniel Neuhauser, and Eran Rabani. Self-averaging stochastic kohn-sham density-functional theory. Physical review letters, 111(10):106402, 2013.

15Daniel Neuhauser, Roi Baer, and Eran Rabani. Communication: Embedded fragment stochastic density functional theory. The Journal of chemical physics, 141(4):044102, July 2014.

16Xavier Gonze, Bernard Amadon, Gabriel Antonius, Frédéric Arnardi, Lucas Baguet, Jean-Michel Beuken, Jordan Bieder, François Bottin, J o h a n n B o u c h e t, Eric Bouset, Nils Brouwer, Fabien Brunelle, Guillaume Brunin, Théo Cavigneac, Jean-Baptiste Charrraud, Wei Chen, Michel Côté, Stefaan Cottien, Jules Denier, Grégory Geneste, Philippe Ghosez, Matteo Giangiamm, Yannick Gillet, Olivier Gingras, Donald R. Hamann, Geoffroy Hautier, Xu He, Nicole Helbig, Natalie Holzwarth, Yongchao Jia, François Jollet, William Lafargue-Dit-Hauret, Kurt Lejaeghere, Miguel A. L. Marques, Alexandre Martin, Cyril Martins, Henrique P. C. Miranda, Francesco Naccarato, Kristin Persson, Guido Petretto, Valentin Planes, Yann Pouillon, Sergei Prokhorenko, Fabio Ricci, Gian-Mario Rignanese, Aldo H. Romero, Michael Marcus Schmitt, Marc Torrent, Michiel J. van Setten, Benoit Van Troeye, Matthieu J. Verstraete, Gilles Zérah, and Josef W. Zwanziger. The abinit project: Impact, environment and recent developments. Comput. Phys. Commun., 248:107042, 2020.

17F Jollet, Marc Torrent, and N A W Holzwarth. Generation of projector augmented-wave atomic data: A 71 element validated table in the xml format. Computer Physics Communications, 185(4):1246–1254, 2014.

18https://www.abinit.org/ATOMICDATA/014-si/Si.LDA_PW-
Tomoya Ono and Kikuji Hirose. Timesaving double-grid method for real-space electronic-structure calculations. Physical Review Letters, 82(25):5016–5019, 1999.

Peter Pulay. Convergence acceleration of iterative sequences. the case of scf iteration. Chemical Physics Letters, 73(2):393–398, 1980.

P Pulay. Improved scf convergence acceleration. Journal of Computational Chemistry, 3(4):556–560, 1982.

NAW Holzwarth, AR Tackett, and GE Matthews. A projector augmented wave (paw) code for electronic structure calculations, part i: atompaw for generating atom-centered functions. Computer Physics Communications, 135(3):329–347, 2001.

Chang-Guo Zhan, Jeffrey A Nichols, and David A Dixon. Ionization potential, electron affinity, electronegativity, hardness, and electron excitation energy: molecular properties from density functional theory orbital energies. The Journal of Physical Chemistry A, 107(20):4184–4195, 2003.

Pedro Borlido, Thorsten Aull, Ahmad W Huran, Fabien Tran, Miguel AL Marques, and Silvana Botti. Large-scale benchmark of exchange–correlation functionals for the determination of electronic band gaps of solids. Journal of chemical theory and computation, 15(9):5069–5079, 2019.

Daniel Neuhauser, Yi Gao, Christopher Arntsen, Cyrus Karshenas, Eran Rabani, and Roi Baer. Breaking the theoretical scaling limit for predicting quasiparticle energies the stochastic gw approach. Physical Review Letters, 113(7):076402, 2014.

Vojtech Vlcek, Eran Rabani, Daniel Neuhauser, and Roi Baer. Stochastic gw calculations for molecules. Journal of Chemical Theory and Computation, 13(10):4997–5003, 2017.

Yi Gao, Daniel Neuhauser, Roi Baer, and Eran Rabani. Sublinear scaling for time-dependent stochastic density functional theory. Journal of Chemical Physics, 142(3):034106, 2015.

Xu Zhang, Gang Lu, Roi Baer, Eran Rabani, and Daniel Neuhauser. Linear-response time-dependent density functional theory with stochastic range-separated hybrids. Journal of Chemical Theory and Computation, 16(2):1064–1072, 2020.