Linear-Scaling Density Functional Theory using the Projector Augmented Wave Method

Nicholas D. M. Hine

Department of Physics, University of Warwick, Coventry, CV4 7AL

Abstract. Quantum mechanical simulation of realistic models of nanostructured systems, such as nanocrystals and crystalline interfaces, demands computational methods combining high-accuracy with low-order scaling with system size. Blöchl's Projector Augmented Wave (PAW) approach enables all-electron (AE) calculations with the efficiency and systematic accuracy of plane-wave pseudopotential calculations. Meanwhile, Linear-Scaling (LS) approaches to Density Functional Theory (DFT) allow for simulation of thousands of atoms in feasible computational effort. This article describes an adaptation of PAW for use in the LS-DFT framework provided by the ONETEP LS-DFT package. ONETEP uses optimisation of the density matrix through *in-situ*-optimised local orbitals rather than the direct calculation of eigenstates as in traditional PAW approaches. The method is shown to be comparably accurate to both PAW and AE approaches and to exhibit improved convergence properties compared to norm-conserving pseudopotential methods.

PACS numbers: 31.15.E-,71.15.Mb,34.20.Gj,31.15.-p,31.15.A-

1. Introduction

Within the field of Density Functional Theory (DFT) [1, 2] there exists a wide range of methodologies with differing strengths: choices of basis set, the approach taken to electron-ion interactions, and the approach to energy minimisation can all vary considerably between different approaches. The Projector Augmented Wave (PAW) method of Blöchl [3] combines two highly-advantageous properties: the accuracy and transferability of all-electron (AE) methods, with the computational efficiency of plane-wave or grid-based pseudopotential methods. PAW proceeds by relating soft pseudo- (PS) wavefunctions, which can be easily represented by plane-waves or on a cartesian grid, to corresponding AE wavefunctions, which vary rapidly near the nuclei. It does this by introducing a linear transformation augmenting the soft part of the wavefunction with partial waves near each atom. PAW is thus an all-electron method in the sense that the all-electron energies and wavefunctions are reproduced (rather than that it calculates all the electronic states explicitly). As well as enabling 'softer' wavefunctions than norm-conserving pseudopotentials, and thus smaller basis sets, PAW enables new types of calculation. Crucially, PAW allows direct access to quantities whose evaluation requires values of wavefunctions, densities and electric fields in the immediate vicinity of the atomic nuclei, which is not possible with norm-conserving pseudopotentials. This occurs as when calculating shielding tensors in Nuclear Magnetic Resonance [4], calculating Electric Field Gradients at nuclei [5], and finding energies of core to conduction transitions in theoretical spectroscopy [6].

All these advantages imply that PAW should be ideal for tackling the significant challenges of theoretical simulation of large-scale models of nanostructured materials. The ability to rigorously predict and explain electronic properties of complex nanoscale systems using PAW-based calculations would be of great value to applications in electronics, photovoltaics, sensing, and elsewhere. However, the majority of the existing, widely-used implementations of DFT using PAW, e.g. VASP [7], PWPAW [8, 9], GPAW [10], PWSCF [11], and ABINIT [12] are all 'traditional' DFT methods, in that they directly work with eigenstates of the Kohn-Sham Hamiltonian, represented in terms of plane-waves or on a grid. Their computational effort thus scales asymptotically as the cube of the system size N. This cost becomes prohibitive much beyond around 1000 atoms with current hardware. Despite the efficiency of PAW, these methods therefore cannot easily be used to simulate the extremely large systems required to study realistically-sized isolated nanostructures.

'Linear-Scaling' (LS) approaches to DFT are designed to overcome this scaling limitation, bypassing the use of Kohn-Sham eigenstates, by working directly with the density matrix (DM) [13]. In LS-DFT, the density matrix is constructed from nonorthogonal local orbitals centered on the atoms, here denoted $\phi_\alpha(\mathbf{r})$, where the index α runs over all orbitals on all sites. It is known [14, 15] that a general flexible form for the DM is:

$$\rho(\mathbf{r}, \mathbf{r'}) = \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta(\mathbf{r'}) \; . \tag{1}$$

where $K^{\alpha\beta}$ is a generalisation of occupancies to nonorthogonal orbitals. Superscript indices denote contravariant quantities throughout this work, and subscript indices covariant quantities. The Einstein convention of summation over repeated Greek indices is employed throughout. Furthermore, in insulators, the 'nearsightedness' of the density matrix [16] means $K^{\alpha\beta}$ is short-ranged, and thus a highly-sparse matrix in large systems. In practice, $K^{\alpha\beta}$ can be truncated for $|\mathbf{R}_\alpha - \mathbf{R}_\beta| > R_K$ for some

range R_K. The total energy and other properties can then be shown to converge systematically with increasing range R_K of the density matrix [17]. The speed at which this convergence occurs is known to be controlled by the magnitude of the energy gap between the occupied and unoccupied states.

To obtain the ground-state energy in such methods, it is expressed in terms of the Kohn-Sham Hamiltonian $\hat{H}_{\mathrm{KS}}$ as

$$E_{\mathrm{T}} = K^{\alpha\beta} H_{\beta\alpha} - E_{\mathrm{dc}}[n] \; , \tag{2}$$

where $H_{\beta\alpha} = \langle \phi_\beta | \widehat{H}_{\mathrm{KS}} | \phi_\alpha \rangle$ and $E_{\mathrm{dc}}[n]$ is the standard double-counting term expressed in terms of the density $n(\mathbf{r}) = \rho(\mathbf{r}, \mathbf{r})$. The total energy E_{T} is thus a functional of $K^{\alpha\beta}$ and $\{\phi_\alpha(\mathbf{r})\}$, and can be minimised with respect to these quantities. This in turn means that the local orbitals $\phi_\alpha(\mathbf{r})$ can be optimised in-situ, allowing a minimal set to be used. The density matrix must obey certain properties to represent a valid DM for an N_e-electron system [14]: it must be correctly normalised, in that $\mathrm{Tr}[\rho] = N_e$, and it must be idempotent, in that $\rho^2 = \rho$, so that it represents a set of orthogonal eigenstates. These constraints can be put in the form of matrix equations involving $K^{\alpha\beta}$. Combined with suitable algorithms for purification and optimisation of the density matrix, these ideas underlie most linear-scaling approaches to DFT.

To date, most practical $O(N)$ methods have supported only norm-conserving pseudopotentials (NCPPs), and have thus not been able to take advantage of the benefits of PAW. This article details a novel linear-scaling approach to PAW, implemented within the ONETEP LS-DFT method [18, 19, 20]. The choice of nomenclature employed closely follows that of Ref. [12], and readers are referred to that work for an introduction to the PAW method as it is applied in traditional DFT.

The ONETEP linear-scaling DFT code has been demonstrated to be suitable for large-scale simulations of nanostructures [21, 22], and scales in parallel to many thousands of cores [23, 24]. High-impact applications of the combined LS+PAW method have already been described to various systems, including transition metal dichalcogenide heterostructures [25], black-phosphorous transistor structures [26], and CdS nanostructures [27], and a range of systems for Electron Energy Loss Spesctroscopy (EELS) [28], but the foundations of this implementation of the method have not previously been described, hence the current work.

2. Linear-Scaling PAW Formalism

ONETEP uses a density matrix expressed as in Eq. 1, where $\{\phi_\alpha(\mathbf{r})\}$ are known as Nonorthogonalised Generalised Wannier Functions (NGWFs), and are expressed in terms of an underlying basis of psinc functions (effectively bandwidth-limited delta functions) on each grid point $\mathbf{r}_i$:

$$\phi_\alpha(\mathbf{r}) = \sum_i c_{i\alpha} D(\mathbf{r} - \mathbf{r}_i) \; . \tag{3}$$

The grid spacing is defined by a cutoff energy E_c, and the radius of the NGWFs is controlled by a strict cutoff R_ϕ beyond which $c_{i\alpha} = 0$. The total energy can be shown to be variational with respect to both parameters, which makes the accuracy of the NGWF representation systematically controllable and asymptotically equivalent to that of plane waves, and thus ideal for PAW.

Since PAW is based around transformation of the eigenstates between AE and PS forms, to use it in a linear-scaling method it must first be reformulated in

terms of a transformation of the density matrix. All the terms in the total energy and Hamiltonian can then be evaluated in linear-scaling computational effort. The evaluation of PAW energies and forces is relatively complex, and this work only describes aspects central to its adaptation to LS-DFT. The reader is referred to Refs. [7, 12] for details overlooked as a result.

The fundamental transformation in traditional PAW [3, 7] is that relating an AE orbital $|\psi_n\rangle$ to a PS orbital $|\widetilde{\psi}_n\rangle$:

$$|\psi_n\rangle = |\widetilde{\psi}_n\rangle + (|\varphi_\nu\rangle - |\widetilde{\varphi}_\nu\rangle)\langle\widetilde{p}^\nu|\widetilde{\psi}_n\rangle \; . \tag{4}$$

This transformation is written in terms of pre-defined projectors $|\widetilde{p}^\nu\rangle$, PS partial waves $|\widetilde{\varphi}_\nu\rangle$ and AE partial waves $|\varphi_\nu\rangle$, localised within a chosen core radius r_c around each atom. The index ν runs over all projectors of all atoms in the system. The projectors are dual to the PS partial waves, hence $\langle\widetilde{p}^\nu|\widetilde{\varphi}_\mu\rangle = \delta_{\nu\mu}$, and within each sphere the projectors and partial waves should ideally form a complete set, giving the completeness relation $\sum_\nu |\widetilde{p}^\nu\rangle\langle\widetilde{\varphi}_\nu| = 1$. This latter relation is only strictly true in the limit of a complete set of partial waves. However, with a suitably-designed PAW potential, it can be shown to be a good approximation when the operator is applied to any state whose energy lies within the range of energies corresponding to valence and low-lying conduction states. Note also that it is implicitly assumed that the projectors associated with different atoms do not overlap: in practice this must be very nearly true for accurate results, but a small amount of volume overlap (around 5%) can be tolerated without any significant devation relative to AE results. In practice, all overlap terms of the form $\langle\widetilde{p}^\nu|\widetilde{\varphi}_\mu\rangle$ where μ and ν originate on different atoms are explicitly neglected.

Within the Linear-Scaling DFT approach, direct access to the eigenstates is not possible, so a transformation equivalent to Eq. 4 must be applied to the AE density matrix. Defining the AE density matrix in terms of the AE orbitals written via the PAW transformation of Eq. 4, one obtains:

$$\begin{aligned}
\rho = & \sum_n \Big(|\widetilde{\psi}_n\rangle + \sum_\nu (|\varphi_\nu\rangle - |\widetilde{\varphi}_\nu\rangle)\langle\widetilde{p}^\nu|\widetilde{\psi}_n\rangle\Big) f_n \Big(\langle\widetilde{\psi}_n| + \sum_\mu (\langle\varphi_\mu| - \langle\widetilde{\varphi}_\mu|)\langle\widetilde{\psi}_n|\widetilde{p}^\mu\rangle\Big) \\
= & \sum_n |\widetilde{\psi}_n\rangle f_n \langle\widetilde{\psi}_n| + \sum_{\nu\mu} (|\varphi_\nu\rangle - |\widetilde{\varphi}_\nu\rangle)(\langle\varphi_\mu| - \langle\widetilde{\varphi}_\mu|) \sum_n \langle\widetilde{p}^\nu|\widetilde{\psi}_n\rangle f_n \langle\widetilde{\psi}_n|\widetilde{p}^\mu\rangle \\
& + \sum_\nu (|\varphi_\nu\rangle - |\widetilde{\varphi}_\nu\rangle)\langle\widetilde{p}^\nu| \sum_n |\widetilde{\psi}_n\rangle f_n \langle\widetilde{\psi}_n| + \sum_n |\widetilde{\psi}_n\rangle f_n \langle\widetilde{\psi}_n|\widetilde{p}^\mu\rangle \sum_\mu (\langle\varphi_\mu| - \langle\widetilde{\varphi}_\mu|)
\end{aligned} \tag{5}$$

At this stage it is possible to identify the PS density matrix $\widetilde{\rho} = \sum_n |\widetilde{\psi}_n\rangle f_n \langle\widetilde{\psi}_n|$, which will become the fundamental quantity with which the calculation works. To proceed further, one can insert identity operators assuming a complete set of partial waves within the PAW spheres, i.e. $\sum_\mu |\widetilde{p}^\mu\rangle\langle\widetilde{\varphi}_\mu| = 1$ and $\sum_\nu |\widetilde{\varphi}_\nu\rangle\langle\widetilde{p}^\nu| = 1$ after the third term and before the fourth:

$$\begin{aligned}
\rho = \widetilde{\rho} & + \sum_{\nu\mu} (|\varphi_\nu\rangle - |\widetilde{\varphi}_\nu\rangle)\langle\widetilde{p}^\nu|\widetilde{\rho}|\widetilde{p}^\mu\rangle(\langle\varphi_\mu| - \langle\widetilde{\varphi}_\mu|) \\
& + \sum_{\nu\mu} (|\varphi_\nu\rangle - |\widetilde{\varphi}_\nu\rangle)\langle\widetilde{p}^\nu|\widetilde{\rho}|\widetilde{p}^\mu\rangle\langle\widetilde{\varphi}_\mu| + \sum_{\nu\mu} |\widetilde{\varphi_\nu}\rangle\langle\widetilde{p}^\nu|\widetilde{\rho}|\widetilde{p}^\mu\rangle(\langle\varphi_\mu| - \langle\widetilde{\varphi}_\mu|)
\end{aligned} \tag{6}$$

It is important to note that at this point the assumption is made that any nonlocal operators will not act between different PAW spheres. This point could require special

attention if the formalism is in future extended to be able to perform exact exchange calculations. Finally, by collecting terms, one can obtain:

$$\rho = \widetilde{\rho} + \sum_{\nu\mu}(|\varphi_\nu\rangle\langle\varphi_\mu| - |\widetilde{\varphi}_\nu\rangle\langle\widetilde{\varphi}_\mu|)\langle\widetilde{p}^\nu|\widetilde{\rho}|\widetilde{p}^\mu\rangle \tag{7}$$

The NGWFs used within a ONETEP calculation, $\{\phi_\alpha(\mathbf{r})\}$, are constructed out of psinc functions as per Eq. 3 and thus display convervence properties similar to plane-waves. They are an ideal means to represent this 'soft' part of the DM, so one can define

$$\widetilde{\rho} = |\phi_\alpha\rangle K^{\alpha\beta}\langle\phi_\beta| \ . \tag{8}$$

In the existing NCPP-based ONETEP method, the NGWFs and the kernel are written in the same basic form, and can be optimised to minimise the total energy subject to the requirement that the kernel be idempotent and correctly normalised: no reference need be made at any point to the eigenstates of the Hamiltonian.

To adapt this for PAW, one must bear in mind that the constraints of idempotency and correct normalisation, $\mathrm{Tr}[\rho] = N_e$ and $\rho^2 = \rho$, must apply to the all electron density DM, not its pseudised form $\tilde{\rho}$. For NCPP case, these two constraints can be expressed as

$$K^{\alpha\beta}S_{\beta\alpha} = N_e \ . \tag{9}$$

and

$$K^{\alpha\gamma}S_{\gamma\delta}K^{\delta\beta} = K^{\alpha\beta} \ . \tag{10}$$

with $S_{\alpha\beta} = \langle\phi_\alpha|\phi_\beta\rangle$ being the overlap matrix of the NGWFs.

In PAW, one can retain the exact same forms, as long as the overlap matrix between NGWFs is redefined to account for the PAW transformation, in the same way as the overlap between eigenstates is redefined in traditional PAW:

$$S_{\alpha\beta} = \langle\phi_\alpha|\Big[1 + |\widetilde{p}^\nu\rangle(\langle\varphi_\nu|\varphi_\mu\rangle - \langle\widetilde{\varphi}_\nu|\widetilde{\varphi}_\mu\rangle)\langle\widetilde{p}^\mu|\Big]|\phi_\beta\rangle \ . \tag{11}$$

Then it can be shown that the same expressions can be retained for the requirements of a normalised, purified kernel, when expressed in terms of $K^{\alpha\beta}$ and the redefined $S_{\alpha\beta}$.

The basic premise of calculating energies in PAW is that the 'atomic' quantities within the spheres centred on each atom are treated on radial grids, while the rest, the 'soft' parts, are treated on a uniform grid. All components of a given AE quantity x are thus broken down into a PS part on the uniform grid, denoted $\widetilde{x}$, an AE atomic part, denoted x^1, and a PS atomic part to be subtracted to cancel the PS part on the grid, denoted $\widetilde{x}^1$. In order to calculate the all-electron energy using these NGWFs, the total valence electron density is first broken down into the components:

$$n_v(\mathbf{r}) = \rho(\mathbf{r},\mathbf{r}) = \widetilde{n}(\mathbf{r}) + n^1(\mathbf{r}) - \widetilde{n}^1(\mathbf{r}) \ , \tag{12}$$

where, from Eqs. 7 and 8,

$$\widetilde{n}(\mathbf{r}) \ = \phi_\alpha(\mathbf{r})K^{\alpha\beta}\phi_\beta(\mathbf{r}) \ , \tag{13}$$

$$n^1(\mathbf{r}) = \varphi_\nu(\mathbf{r})\rho^{\nu\mu}\varphi_\mu(\mathbf{r}) \ , \tag{14}$$

$$\widetilde{n}^1(\mathbf{r}) = \widetilde{\varphi}_\nu(\mathbf{r})\rho^{\nu\mu}\widetilde{\varphi}_\mu(\mathbf{r}) \ . \tag{15}$$

The crucial matrix $\rho^{\nu\mu}$ is the projection of the density matrix in the PAW spheres:

$$\rho^{\nu\mu} = \langle\widetilde{p}^\nu|\phi_\alpha\rangle K^{\alpha\beta}\langle\phi_\beta|\widetilde{p}^\mu\rangle \ . \tag{16}$$

Because both sets of functions ϕ_α and $\widetilde{p}_\nu$ are strictly localised, $\langle\phi_\alpha|\widetilde{p}^\nu\rangle$ is very sparse in large systems, so $\rho^{\nu\mu}$ can be rapidly evaluated from $K^{\alpha\beta}$ with efficient sparse matrix algebra [23, 29]. It is a 'block-diagonal' matrix, in that only blocks for which ν and μ are on the same atom are nonzero.

One of the central aims of PAW is to avoid needing to treat explicitly the interaction of spatially rapidly-varying quantities such as core densities between different atoms. Therefore a key susbstition is made to enable the density-density interactions to be calculated efficiently. The total density is further decomposed in the manner of Refs [7, 12] into soft pseudo density plus a compensation density term (sometimes referred to as augmentation density). This also occurs for other AE terms which need to be represented on the regular cartesian grid. A soft pseudo-density for the nucleus and core electrons $\widetilde{n}_{\mathrm{Zc}}$ replaces the AE version n_{Zc}: this is used to generate the Hartree potential resulting from the nucleus and any core electrons. Similarly, a soft core density $\widetilde{n}_{\mathrm{c}}$ replaces the AE n_{c}, which is used to calculate the nonlinear core-correction terms (i.e. any nonlinearity in the exchange-correlation potential resulting from overlap between core and valence electron density).

The compensation density is defined by an expansion of the form:

$$\widehat{n}(\mathbf{r}) = \sum_{LM} \rho^{\nu\mu} \widehat{Q}^{LM}_{\nu\mu}(\mathbf{r}) \;, \tag{17}$$

where $\widehat{Q}^{LM}_{\nu\mu}(\mathbf{r})$ are 'soft' functions designed to reproduce the L, M moments of the difference between the AE and PS densities (See Ref. [12], Eqs. 12-17, for further details). The notation $\widehat{x}$ refers to quantities associated with this compensation density.

The total energy is divided up as $E = \widetilde{E} + E^1 + \widetilde{E}^1$, with

$$\widetilde{E} \;= K^{\alpha\beta}\langle\phi_\beta|\frac{\text{-}1}{2}\nabla^2|\phi_\alpha\rangle + E_{\mathrm{xc}}[\widetilde{n}+\widetilde{n}_{\mathrm{c}}] + E_{\mathrm{H}}[\widetilde{n}+\widehat{n}] + \int v_{\mathrm{H}}[n_{\mathrm{Zc}}](\widetilde{n}+\widehat{n})d\mathbf{r} + E_{\mathrm{II}} \tag{18}$$

$$E^1 = \rho^{\nu\mu}\langle\varphi_\mu|\frac{\text{-}1}{2}\nabla^2|\varphi_\nu\rangle + E_{\mathrm{xc}}[n^1+n_{\mathrm{c}}] + E_{\mathrm{H}}[n^1] + \int v_{\mathrm{H}}[n_{\mathrm{Zc}}]n^1 d\mathbf{r} \tag{19}$$

$$\widetilde{E}^1 = \rho^{\nu\mu}\langle\widetilde{\varphi}_\mu|\frac{\text{-}1}{2}\nabla^2|\widetilde{\varphi}_\nu\rangle + E_{\mathrm{xc}}[\widetilde{n}^1+\widetilde{n}_{\mathrm{c}}] + E_{\mathrm{H}}[\widetilde{n}^1+\widehat{n}] + \int v_{\mathrm{H}}[\widetilde{n}_{\mathrm{Zc}}](\widetilde{n}^1+\widehat{n})d\mathbf{r} \tag{20}$$

Note that the compensation density $\widehat{n}$ is generally not included in the XC terms: as its shape is unphysical, it can have unwanted effects on the XC potential [30].

Taking the derivative of the full expression for E with respect to the pseudo-density operator $\widetilde{\rho}$ gives the Hamiltonian $\widetilde{H}$

$$\widetilde{H} = -\frac{1}{2}\nabla^2 + \widetilde{v}_{\mathrm{eff}} + |\widetilde{p}^\nu\rangle D_{\nu\mu}\langle\widetilde{p}^\mu| \;, \tag{21}$$

in which each of the three terms represent kinetic, local effective potential and nonlocal effective potential energy contributions. The nonlocal energies $D_{\nu\mu}$ will be defined below, and $\widetilde{v}_{\mathrm{eff}}$ is an effective potential defined on a uniform cartesian grid in terms of Hartree, XC and ionic parts,

$$\widetilde{v}_{\mathrm{eff}} = \widetilde{v}_{\mathrm{H}}[\widetilde{n}+\widehat{n}] + \widetilde{v}_{\mathrm{H}}[\widetilde{n}_{\mathrm{Zc}}] + \widetilde{v}_{\mathrm{xc}}[\widetilde{n}+\widetilde{n}_{\mathrm{c}}] \;. \tag{22}$$

Because they involve no AE quantities evaluated outside the spheres around each atom, all three of these terms are well-behaved on the uniform cartesian grid. They can thus be evaluated numerically with the standard methods already implemented in grid-based DFT codes.

The nonlocal energies are similarly decomposed into three parts, collecting terms resulting from augmentation densities, AE partial waves and PS partial waves respectively:

$$D_{\nu\mu} = \widehat{D}_{\nu\mu} + D^1_{\nu\mu} - \widetilde{D}^1_{\nu\mu} \tag{23}$$

with

$$\widehat{D}_{\nu\mu} = \sum_{LM} \int \widetilde{v}_{\mathrm{H}}[\widetilde{n} + \widehat{n} + \widetilde{n}_{\mathrm{Zc}}](\mathbf{r}) \widehat{Q}^{LM}_{\nu\mu}(\mathbf{r}) d\mathbf{r} \ , \tag{24}$$

$$D^1_{\nu\mu} = \langle \varphi_\nu | \frac{-1}{2} \nabla^2 + v^1_{\mathrm{eff}} | \varphi_\mu \rangle \ , \tag{25}$$

$$\widetilde{D}^1_{\nu\mu} = \langle \widetilde{\varphi}_\nu | \frac{-1}{2} \nabla^2 + \widetilde{v}^1_{\mathrm{eff}} | \widetilde{\varphi}_\mu \rangle \tag{26}$$

$$+ \sum_{LM} \int \widetilde{v}_{\mathrm{H}}[\widetilde{n}^1 + \widehat{n} + \widetilde{n}_{\mathrm{Zc}}](\mathbf{r}) \widehat{Q}^{LM}_{\nu\mu}(\mathbf{r}) \mathrm{d}\mathbf{r} \ , \tag{27}$$

and

$$v^1_{\mathrm{eff}} = v_{\mathrm{H}}[n^1] + v_{\mathrm{H}}[n_{\mathrm{Zc}}] + v_{\mathrm{xc}}[n^1 + n_{\mathrm{c}}] \ , \tag{28}$$

$$\widetilde{v}^1_{\mathrm{eff}} = v_{\mathrm{H}}[\widetilde{n}^1 + \widehat{n}] + v_{\mathrm{H}}[\widetilde{n}_{\mathrm{Zc}}] + v_{\mathrm{xc}}[\widetilde{n}^1 + \widetilde{n}_{\mathrm{c}}] \ . \tag{29}$$

Calculation of the bandstructure energy $E_{\mathrm{bs}} = K^{\alpha\beta} \widetilde{H}_{\beta\alpha}$ in $O(N)$ computational effort requires that matrix elements in the NGWF representation, $\langle \phi_\beta | \widetilde{H} | \phi_\alpha \rangle$ can each be calculated in $O(1)$. This can be achieved for the first two terms in Eq. 21 with the 'FFT box' approach [31], in which a box large enough to encompass one function $\phi_\alpha(\mathbf{r})$ and all its overlapping neighbours $\phi_\beta(\mathbf{r})$ is used to evaluate matrix elements such as $\langle \phi_\beta | -\frac{1}{2}\nabla^2 | \phi_\alpha \rangle$ and $\langle \phi_\beta | \widetilde{v}_{\mathrm{eff}} | \phi_\alpha \rangle$.

The nonlocal contribution to $\widetilde{H}_{\beta\alpha}$ is

$$V^{\mathrm{nl}}_{\beta\alpha} = \langle \phi_\beta | \widetilde{p}^\nu \rangle D_{\nu\mu} \langle \widetilde{p}^\mu | \phi_\alpha \rangle \ . \tag{30}$$

To find this, the NGWF-projector overlaps $\langle \phi_\beta | \widetilde{p}^\nu \rangle$ are constructed by building each projector $\widetilde{p}^\nu(\mathbf{r})$ in reciprocal space in the FFT box, transforming it to real space and finding its overlap with each NGWF, as described in Ref. [29].

The FFT box approach is also used for constructing the screened nonlocal energy term, $\widehat{D}_{\nu\mu}$. Each of the contributing integrals $\int \widetilde{v}_{\mathrm{H}}(\mathbf{r}) \widehat{Q}^{LM}_{\nu\mu}(\mathbf{r}) d\mathbf{r}$, nominally needs to be taken on the whole uniform grid of the simulation cell. However, it is clear from the limited range of $\widehat{Q}^{LM}_{\nu\mu}(\mathbf{r})$ that in fact it can be performed on a small 'augmentation box'. This box is set at a size determined by twice the largest PAW radius r_c in the system, since $\widehat{Q}^{LM}_{\nu\mu} = 0$ outside the PAW region of a given atom. The atom-centered terms in $D^1_{\nu\mu}$ and $\widetilde{D}^1_{\nu\mu}$ lend themselves naturally to $O(1)$ evaluation, as they are evaluated on a radial logarithmic grid for each atom, using the techniques described in Eqs. 44-50 of Ref. [12]. The XC energies and potentials on the radial grids are evaluated using an expansion over the moments of $n(\mathbf{r})$ and $v_{\mathrm{xc}}[n](\mathbf{r})$ as described in Eqs. 56-69 of Ref. [12].

The ingredients are thus available for the construction of the sparse matrix $\widetilde{H}_{\alpha\beta}$ from $K^{\alpha\beta}$ and $\{\phi_\alpha(\mathbf{r})\}$ in $O(N)$ computational effort. First, the PS density $\widetilde{n}$ is constructed on the full grid (Eq. 13), then the sphere-projected density kernel $\rho^{\nu\mu}$ is calculated (Eq. 16) and used to construct $\widehat{n}$ on the full grid, again employing the 'augmentation box', and thus to construct $\widetilde{v}_{\mathrm{eff}}$. $\rho^{\nu\mu}$ is re-used to find $\widehat{n}$, n^1 and $\widetilde{n}^1$ for each atom, and these enable construction of v^1_{eff} and $\widetilde{v}^1_{\mathrm{eff}}$ and thus $D^1_{\nu\mu}$ and $\widetilde{D}^1_{\nu\mu}$ (Eqs. 24-29). Finally, the nonlocal energies are screened using Eq. 23 to produce $D_{\nu\mu}$.

3. Full Calculations

The practical scheme for calculations with this method is as follows: the NGWFs are first initialised to suitably-chosen pseudoatomic orbitals, obtained by solving the isolated pseudoatom within a sphere of radius R_α for each species present [32]. For example, for first-row elements one s-like PAO and three p-like PAOs are taken as the starting point for NGWF optimisation, with further valence shells added as required for heavier elements. Initialisation of the kernel $K^{\alpha\beta}$ describing the density matrix is achieved by constructing an initial trial hamiltonian on the basis of a superposition of the pseudoatomic densities, and applying Palser-Manolopoulos canonical purification [33].

Minimisation of the total energy $E = K^{\alpha\beta}\widetilde{H}_{\beta\alpha} - E_{\rm dc}[n]$ is then achieved using a two-level loop as described in Ref. [19]. The inner loop minimises the energy with respect to elements of the density kernel $K^{\alpha\beta}$ using a variant [34] of the Li-Nunes-Vanderbilt method [35]. The purified kernel is represented in terms of an auxiliary kernel $L^{\alpha\beta}$ as $K^{\alpha\beta} = (3LSL - 2LSLSL)^{\alpha\beta}$ (i.e. the result of one iteration of the McWeeny purification transformation). Consistent normalisation is ensured by replacing $K^{\alpha\beta}$ with $K_n^{\alpha\beta} = \frac{nK^{\alpha\beta}}{K^{\kappa\lambda}S_{\lambda\kappa}}$, which is a rescaled version which is defined to be normalised to n electrons. Minimisation of E subject to the appropriate constraints on $K^{\alpha\beta}$ is thus in fact obtained by minimising instead with respect to the elements $L^{\alpha\beta}$.

The outer optimisation loop optimises the coefficients $c_{i\alpha}$ of Eq. 3 which describe the NGWFs $\phi_\alpha(\mathbf{r})$. For this, it is necessary to obtain the gradient of the total energy with respect to the expansion coefficients of the NGWFs in the psinc basis. This can be found by taking derivatives of the total energy expression with respect to the value of specific NGWF α at a point $\mathbf{r}$, $\phi_\alpha(\mathbf{r})$, ensuring all terms with some dependence on $\phi_\alpha(\mathbf{r})$ are propagated appropriately. The result is:

$$\frac{\partial E}{\partial \phi_\alpha(\mathbf{r})} = K_n^{\alpha\beta}[\tilde{H}\phi_\beta](\mathbf{r}) + \phi_\beta(\mathbf{r})\tilde{Q}^{\alpha\beta} + \sum_{\mu\nu}\tilde{p}^\mu(\mathbf{r})O_{\mu\nu}\langle\tilde{p}^\nu|\phi_\alpha\rangle\tilde{Q}^{\alpha\beta}\ , \quad (31)$$

where

$$\tilde{Q}^{\alpha\beta} = \frac{n}{K^{\kappa\lambda}S_{\lambda\kappa}}\Big(3L\tilde{H}'L - 2L\tilde{H}'LSL - 2LSL\tilde{H}'L - \mu K\Big)^{\alpha\beta} \quad (32)$$

with $\tilde{H}' = \tilde{H} - \mu S$ and μ given by

$$\mu = \frac{(3LSL - 2LSLSL)^{\alpha\beta}\tilde{H}_{\beta\alpha}}{(3LSL - 2LSLSL)^{\kappa\lambda}S_{\lambda\kappa}}\ . \quad (33)$$

The first term in Eq. 31 can be expanded as:

$$K_n^{\alpha\beta}[\tilde{H}\phi_\beta](\mathbf{r}) = -\frac{1}{2}\nabla^2[K_n^{\alpha\beta}\phi_\beta](\mathbf{r}) + \tilde{v}_{\rm eff}(\mathbf{r})[K_n^{\alpha\beta}\phi_\beta(\mathbf{r})] + \sum_{\mu\nu}\tilde{p}^\mu(\mathbf{r})\left(\hat{D}_{\mu\nu} + D^1_{\mu\nu} - \tilde{D}^1_{\mu\nu}\right)\langle\tilde{p}^\nu|\phi_\alpha\rangle\ . \quad (34)$$

Eqs. 31-34 can all be evaluated in linear-scaling effort via the methods already discussed, namely a combination of sparse matrix multiplication, FFT-box based function summation, and overlap integrals of localised functions.

Both optimisation loops procede via a conjugate-gradients procedure using the steepest-descents energy gradients $\partial E/\partial L^{\alpha\beta}$ and $\partial E/\partial c_i$ as initial search directions

and then updating the search directions using the Fletcher-Reeves CG scheme thereafter.

Once convergence with respect to both quantities is achieved, a variational minimum energy for a given basis quality (as defined by $E_{\rm cut}$, R_ϕ and R_K) is obtained. In practice, arbitrarily high accuracy can be obtained in the optimisation with respect to $L^{\alpha\beta}$ for a given set of NGWFs, while due to the competing demands of localisation and energy minimisation, it is only possible to optimise the NGWFs to a certain threshold RMS value determined by the localisation radius. Typically, setting this RMS threshold to $10^{-6}\,\mathrm{Ha}\,a_0^{-3/2}$ is sufficient to allow energy and force calculations to a convergence below 1 meV/atom.

For calculations utilising norm-conserving pseudopotentials, the highest magnitude of G-vectors required to describe the cartesian grid on which the density is expressed is exactly twice the magnitude of that describing the wavefunctions. There is thus no reason to use any finer a grid than $2\times$ the wavefunction grid in such calculations. By contrast, it is important to note that in PAW, the convergence with respect to the spacing of the cartesian grid on which the density is described can be rather slower than this. This is because the features of the shape functions and/or the pseudised core density functions can be rather more strongly peaked than the wavefunctions themselves. It is often necessary to use a larger multiplier than $2\times$ the underlying cutoff energy. In the tests in this article, grids of $4\times$ the underlying wavefunction grid are used to ensure complete convergence.

4. Atomic Forces

As described in [20] and [32], the ionic forces in LS-DFT can be determined from the Hellmann-Feynman theorem. Normally, this results in force terms due only to those parts of the total energy which depend explicitly on ionic coordinates: the ion-ion term, the local potential term, the PS density nonlinear XC term, the compensation density term, and the nonlocal term:

$$\mathbf{F}_I = -\frac{dE}{d\mathbf{R}_I} = \mathbf{F}_I^{\rm ion-ion} + \mathbf{F}_I^{\rm loc} + \mathbf{F}_I^{\rm nlcc} + \widehat{\mathbf{F}}_I + \mathbf{F}_I^{\rm nl}\ . \tag{35}$$

In the nonlocal term $\mathbf{F}_I^{\rm nl}$ in PAW there is, however, the added complication that the overlap operator changes with the positions of the ions. Therefore, in addition to the explicit dependence of the energy on $\mathbf{R}_I$, there is also a further term entering via the dependence of the orthonormalisation condition on the location of the ions. To render this into a form that can be evaluated, the identity $\frac{\partial}{\partial \mathbf{R}_I}\left[K^{\alpha\beta}\langle\phi_\beta|\hat{S}|\phi_\alpha\rangle\right] = 0$ is required. The total force is thus written as:

$$\begin{aligned}
\mathbf{F}_I =& -\frac{\partial E}{\partial \mathbf{R}_I} - \left\{\frac{\partial E}{\partial|\phi_\alpha\rangle}\frac{d|\phi_\alpha\rangle}{d\mathbf{R}_I} + \frac{\partial E}{\partial\langle\phi_\alpha|}\frac{d\langle\phi_\alpha|}{d\mathbf{R}_I}\right\} \\
=& -\frac{\partial E}{\partial \mathbf{R}_I} - \left\{K^{\alpha\beta}\langle\phi_\beta|\hat{H}\frac{d|\phi_\alpha\rangle}{d\mathbf{R}_I} + \frac{d\langle\phi_\alpha|}{d\mathbf{R}_I}\hat{H}|\phi_\beta\rangle K^{\beta\alpha}\right\} \\
=& -\frac{\partial E}{\partial \mathbf{R}_I} + K^{\alpha\delta}H_{\delta\gamma}S^{\gamma\beta}\langle\phi_\beta|\frac{d\hat{S}}{d\mathbf{R}_I}|\phi_\alpha\rangle\ ,
\end{aligned} \tag{36}$$

where the identities $\hat{1} = |\phi_\gamma\rangle S^{\gamma\delta}\langle\phi_\delta|\hat{S}$ has been employed in obtaining the final expression.

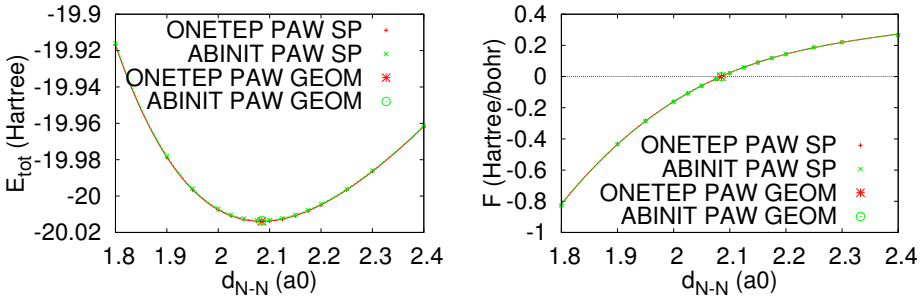

Figure 1. Total energies and forces as a function of bondlength for N_2, O_2 (triplet), CO and TiO_2 molecules. The curves shown are for $E(d)$ are a quartic polynomial fit to the data, while those for $F(d)$ are the derivative of this fit. The calculated force $F(x)$ is thus seen to agree very well with $-dE/dx$.

After some manipulation, the different force terms listed above evaluate to:

$$\begin{aligned}
\mathbf{F}_I^{\mathrm{loc}} &= -\int (\widetilde{n}(\mathbf{r})+\widehat{n}(\mathbf{r}))\frac{\partial v_{\mathrm{H}}[\widetilde{n}_{\mathrm{Zc}}](\mathbf{r})}{\partial \mathbf{R}_I}\,\mathrm{d}\mathbf{r}\,, \qquad (37)\\
\mathbf{F}_I^{\mathrm{nlcc}} &= -\int v_{\mathrm{xc}}[\widetilde{n}+\widetilde{n}_{\mathrm{c}}]\frac{\partial \widetilde{n}_{\mathrm{c}}(\mathbf{r})}{\partial \mathbf{R}_I}\,\mathrm{d}\mathbf{r}\\
\widehat{\mathbf{F}}_I &= -\int \widetilde{v}_{\mathrm{H}}[\widetilde{n}+\widehat{n}+\widetilde{n}_{\mathrm{Zc}}](\mathbf{r})\frac{\partial \widehat{n}(\mathbf{r})}{\partial \mathbf{R}_I}\,\mathrm{d}\mathbf{r}\\
\mathbf{F}_I^{\mathrm{nl}} &= \sum_{\nu\in I}\Big[-D_{\nu\mu}\langle\widetilde{p}^{\mu}|\phi_\alpha\rangle K^{\alpha\beta}\left\langle\phi_\beta|\frac{\partial\widetilde{p}^{\nu}}{\partial\mathbf{R}_I}\right\rangle\\
&\quad -\left\langle\frac{\partial\widetilde{p}^{\nu}}{\partial\mathbf{R}_I}|\phi_\alpha\right\rangle K^{\alpha\beta}\langle\phi_\beta|\widetilde{p}^{\mu}\rangle D_{\mu\nu}\\
&\quad +O_{\nu\mu}\langle\widetilde{p}^{\mu}|\phi_\alpha\rangle K^{\alpha\gamma}H_{\gamma\delta}S^{\delta\beta}\left\langle\phi_\beta|\frac{\partial\widetilde{p}^{\nu}}{\partial\mathbf{R}_I}\right\rangle\\
&\quad +\left\langle\frac{\partial\widetilde{p}^{\nu}}{\partial\mathbf{R}_I}|\phi_\alpha\right\rangle K^{\alpha\gamma}H_{\gamma\delta}S^{\delta\beta}\langle\phi_\beta|\widetilde{p}^{\mu}\rangle O_{\mu\nu}\Big]\,.
\end{aligned}$$

The notation $\nu \in I$ means that the sum runs only over projectors ν on atom I. Construction of $\mathbf{F}_I^{\mathrm{loc}}$, $\mathbf{F}_I^{\mathrm{nlcc}}$ and $\langle\phi_\beta|\frac{\partial\widetilde{p}^{\nu}}{\partial\mathbf{R}_I}\rangle$ is computationally equivalent to the evaluation of corresponding terms in the case of norm conserving potentials, which is described in Ref. [20]. Construction of $\widehat{\mathbf{F}}_I$ again employs the 'augmentation box': the gradient of the augmentation density is calculated by an FFT for the density associated with each box, multiplication by the **G**-vectors of the box, and a transform back to obtain $-\frac{\partial\widehat{n}(\mathbf{r})}{\partial\mathbf{R}_I}$. The final term, $\mathbf{F}_I^{\mathrm{nl}}$, is obtained via a series of sparse matrix operations on the corresponding matrices.

As discussed in [32], in cases where it is not feasible to optimise the NGWFs sufficiently that their gradient is everywhere effectively zero, one can instead, with minimal computational effort, calculate the residual Pulay terms in terms of the NGWF gradients. This involves no extra complications in PAW as the NGWF gradient already contains all relevant contributions. The resulting analytic forces can be used as inputs for tasks such as geometry optimisation and molecular dynamics, and agree to very high precision with the gradient of the energy landscape.

5. Demonstration and Applications of the Method

The purpose of this article is to outline the novel aspects of the methodology and describe its computational implementation. A number of recent papers have showcased applications of the method to realistic applications, so an example application will not be included in this work. Instead, benchmark results are presented for selected small systems that seek to demonstrate the accuracy of the method. Firstly, the N_2 dimer is examined carefully, so as to demonstrate the exact equivalence of the method to the implementation of PAW method in the plane-wave DFT code ABINIT, which itself has been exhaustively benchmarked against all-electron methods and shown to be comparably accurate [36].

Figure 1 shows the agreement between two calculation methodologies: plane-wave DFT with a traditional $O(N^3)$ PAW implementation, performed using the ABINIT code (green), and the present $O(N)$ PAW method implemented in ONETEP (red). PAW datasets for each atomic species are generated with the standard recipes supplied with the AtomPAW code [8], employing the PW92 formulation of the Local Density Approximation. Figure 1 shows that for small dimer systems, the new LS-PAW method produces total energies in very precise agreement with traditional DFT results obtained by using the same PAW datasets in ABINIT, and furthermore that the calculated forces agree very precisely with the gradient of the energy with respect to atomic positions, such that the optimised geometries are in precise agreement. In these tests, well-converged parameters are used for the NGWF radii and cutoff energies: R_α =10.0a_0, E_c=1000 eV respectively.

Next, convergence is investigated for Perylenetetracarboxylic dianhydride (PTCDA), an organic dye molecule which is of current interest as an organic semiconductor. It is a medium-sized organic system (38 atoms) which contains oxygen, hence when simulated with traditional norm-conserving PSPs requires quite a high cutoff energy for good convergence. Here PAW datasets from the JTH library [37] are compared with standard NCPP datasets supplied with ONETEP. The molecules is modelled in a cubic box of side 28 Å. Geometry relaxation is performed with the PBE functional beforehand with highly-converged parameters (1000eV, 10a_0 NGWF cutoff radius) to provide an accurate starting geometry.

In general, convergence results for total energy differences between similar structures are more meaningful than convergence of total energies themselves, and it is not necessary to run calculations with parameters where complete numerical convergence with respect to the plane wave cutoff is achieved. Indeed, if total energy convergence is required, the cutoff energy required for binding energy convergence will often be significantly exceeded. However, if good total energy convergence is achieved, then good convergence of derived properties such as binding energy must have already been achieved and the former therefore represents a stronger criterion. Good overall convergence of the total energy is also highly desirable from the point of view of comparability between calculations run with different supercell sizes.

Figure 2 demonstrates that with the PAW approach, total energy can be reached at considerably lower cutoff energies than is the case with the traditional NCPPs. Here, convergence to within 0.1eV/atom is reached around 600 eV in PAW, whereas 1000 eV is required before the same level of convergence is possible with NCPPs, which results in a considerable saving of computational effort. It is important to note (see insets) that the rate of total energy convergence with respect to NGWF radius is, as expected, very close to exactly equivalent in both methods, since this behaviour is

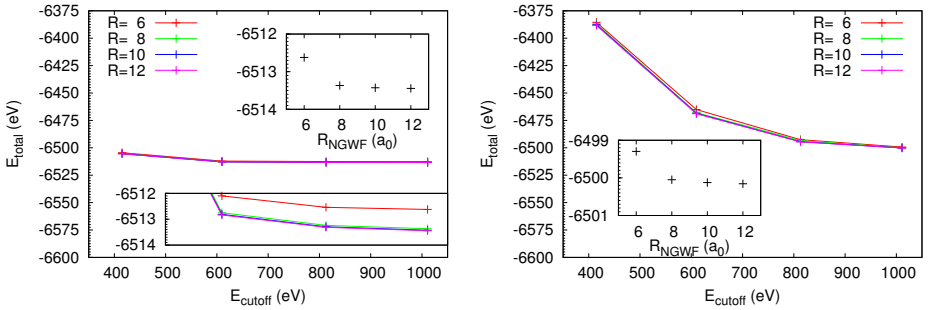

Figure 2. Total energy convergence for a model of perylenetetracarboxylic dianhydride (PTCDA). Left: total energy convergence with respect to cutoff energy for the current PAW method. Lower inset reproduces a zoomed-in region of the tail of the plot. Upper inset shows convergence with respect to NGWF radius at 1000eV cutoff energy. Right: total energy convergence for standard norm-conserving pseudopotentials. Total energy convergence of under 0.1 eV/atom is achieved with respect to NGWF radius by around $10a_0$ in both methods, but in PAW total energy convergence with respect to the energy cutoff defining the grid is achieved much sooner: by around 600eV for this system.

related to the details of the density matrix in regions outside of both the PAW spheres and the norm-conserving pseudopotential radii.

It is next demonstrated that overall linear-scaling of the computational effort, which has been extensively demonstrated in previous work with the ONETEP code, is retained in PAW. Trial systems comprising slabs of TiO_2 exposing the [101] surface, of increasing size, are simulated on a fixed number of parallel cores. The prototype slab is 4 primitive cells in depth, and comprises 5×3 units of the surface unit cell, for a total of 720 atoms. Larger models can then be constructed by repeating this cell $n = 1$, 2, 3, 4 and 5 times along the x-direction, for cells ranging from 720 to 3600 atoms, with $N = 720n$. Extending the cell in just one direction creates rather more skewed unit cells than might be used in practical calculations, but allows a regime to be reached where all relevant matrices are strictly sparse, so provides a good demonstration of the linear-scaling behaviour. A psinc cutoff energy of 600eV, an NGWF radius of $8a_0$, and a kernel cutoff of $40a_0$ are used for all calculations. The PAW dataset for Ti from the JTH library [37] includes $3s$ and $3p$ semicore states, hence a total of 13 NGWFs per Ti atom are required, and 4 NGWFs per O atom, making this a rather challenging test system.

Figure 3(a) shows the total wall time for the execution of a complete NGWF and kernel optimisation cycle (one total energy calculation from scratch) for $n = 1, \ldots, 5$ repetitions of the prototype slab, run on 720 cores of the ARCHER Cray XC30 supercomputer. It is clear that beyond around 1000 atoms the wall time scales as $O(N)$ in this regime.

That this overall $O(N)$ scaling must be carried over from calculations with NCPPs can also be seen from the scaling data in Figure 3(b). This shows total wall time for total energy calculations on a van der Waals heterostructure, comprising a layer of hBN, a layer of black phosphorus, and another layer of hBN, for a total of 1824 atoms. The number of cores is varied from 240 to 5184, using a hybrid OpenMP-MPI paralellisation scheme with 6 OpenMP threads per MPI process. Overall scaling is

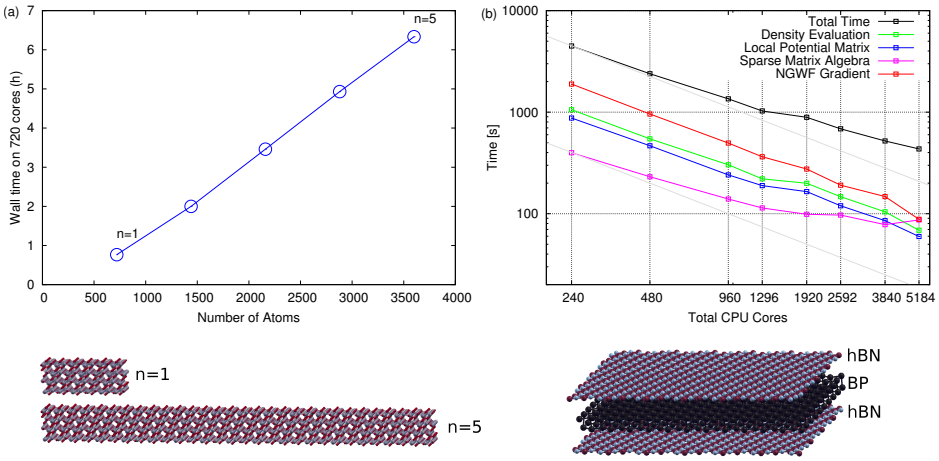

Figure 3. (a) Scaling of total computational effort, measured by wall time when run on 720 cores of a Cray XC30 machine (ARCHER). The system simulated is a series of multiples of a 720-atom surface slab of anatase-structure TiO_2, exposing the [101] surface, which is of interest in photocatalytic applications. (b) Strong-scaling of wall time with respect to number of cores, for a calculation on a van der Waals heterostructure, comprising a layer of hBN, a layer of black phosphorus, and another layer of hBN, for a total of 1824 atoms. Ideal scaling would be parallel to the grey lines. Overall scaling is good until around 1920 cores, and the majority of the computational effort is accounted for by the four main routines relating to density, local potential, NGWF gradient and sparse matrix algebra, with very similar balance as for NCPP calculations.

seen to be rather good until around 1920 cores, and improvement is maintained all the way to 5184 cores. A ballpark figure of 1 atom per core for calculations on transition metals is generally a good guide to the limit of highly-efficient parallel scaling for most calculations. In all these runs, the vast majority of the time is seen to be spent in routines which are related to the evaluation of quantities involving the NGWFs and the sparse matrices, all of which scale the same way when using PAW as they do in NCPPs. Explicit 'overhead' associated with PAW on-site quantities, which are in any case evaluated atom-by-atom and are thus inherently $O(N)$, is not a significant contributor to the total computational effort.

As a final demonstration of the method, recent applications of this methodology to problems in the area of nanomaterials simulation will be briefly summarised. In these applications, the full power of the linear-scaling methodology was demonstrated, with very large-scale simulations of up to 2000 atoms. It should be noted that previous benchmark calculations have showed that the ONETEP code is capable of scaling to calculations of up to 100,000 atoms on upwards of 10,000 cores [23, 24], and the current PAW methodology retains this scaling in full.

Constantinescu *et al.* [25] used the approach in a 2015 investigation of the energetic and electronic properties of bilayer heterostructures of transition metal dichalcogenide materials, specifically MoS_2 and $MoSe_2$. We were able to show that the energy landscape was very flat under rotation of the layers with respect to each other, with no strongly preferred alignment: this means almost any stacking can be envisaged. However, there were strong effects of the heterostructure construction on

the electronic properties: by unfolding the spectral function into the primitive cells, we showed that the electronic band structure, specifically the band offsets and the carrier effective masses, of each layer is tunable through rotation. This work would have been very much more expensive without the benefit of the PAW formalism described here. Further investigations as to the convergence properties of the ONETEP PAW method applied to MoS_2 and $MoSe_2$ can be found in the supplementary material to Ref. [25]. Further work on layered material heterostructures by Constantinescu *et al.* [26] involved an investigation of multi-layer stacks comprising black phosphorus (BP) and hexagonal boron nitride (hBN). In this work we were able to show that spacer layers of hBN were capable of encapsulating the BP and preserving its electronic properties, and that the devices modelled showed considerable promise for application in tunnelling field effect transistors. Once again, such large simulations were made considerably more feasible via the use of the ONETEP PAW methodology.

More recently, Tait *et al.* showed that the PAW method within ONETEP can be used to implement methodology for Electronic Energy Loss Spectroscopy (EELS) [28]. The simulated EELS approach has been previously implemented successfully with plane-wave methods, but in the case of LS-DFT it enables EELS calculations on very large model systems such as grain boundaries and surfaces. Finally, recent work has focussed on combination of PAW with approaches for Linear-Response TDDFT in the LS-DFT framework [38, 39], resulting in applications of the combined approach to study realistic seminconductor nanocrystals of CdS [27]. The results indicate that PAW is highly appropriate to studying the properties of inorganic nanomaterials systems incorporating transition metals such as Cd, which is challenging with traditional methods for a wide variety of reasons.

6. Conclusion

An adaptation of the PAW method to a Linear-Scaling DFT framework used in the ONETEP code has been presented. Its equivalence to other PAW implementations and its considerably improved convergence properties have been demonstrated, particularly for model systems which would present considerable difficulties for standard cubic-scaling, norm-conserving pseudopotential approaches. In particular, for systems containing transition metal ions, it is considerably easier to construct highly-accurate transferrable potentials which require only relatively low cutoff energies. Its scaling with respect to number of atoms and number of parallel cores has been demonstrated for large inorganic systems. In the other works discussed which utilise this methodology, it has been enabled production calculations of properties of systems up to 2000 atoms. The combination of ONETEP LS-DFT with the PAW approach thus shows considerable promise for future nanomaterials simulations.

Acknowledgments

N.D.M.H. acknowledges the Engineering and Physical Sciences Research Council (Grant EP/G05567X/1) for postdoctoral funding from 2009-2012, during which time the code implementing this methodology was written. This work used the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk). Imperial College HPC service and the Centre for Scientific Computing at the University of Warwick are also thanked for computing resources on which the methodology was tested. The ONETEP Developers' Group (P. Haynes, A. Mostofi, C.-K. Skylaris and M.C. Payne)

and David O'Regan are thanked for valuable discussions. Gabriel Constantinescu, Edward Tait, and Jose-Maria Escartín are thanked for their contribution to the ongoing development of methodology for Projector Augmented Wave methodology.

References

[1] Hohenberg P and Kohn W 1964 *Phys. Rev.* **136** B864
[2] Kohn W and Sham L J 1965 *Phys. Rev.* **140** A1133
[3] Blöchl P E 1994 *Phys. Rev. B* **50** 17953
[4] Pickard C J and Mauri F 2001 *Phys. Rev. B* **63** 245101
[5] Petrilli H M, Blöchl P E, Blaha P and Schwarz K 1998 *Phys. Rev. B* **57** 14690–14697
[6] Jayawardane D N, Pickard C J, Brown L M and Payne M C 2001 *Phys. Rev. B* **64** 115107
[7] Kresse G and Joubert D 1999 *Phys. Rev. B* **59** 1758
[8] Holzwarth N A W, Tackett A R and Matthews G E 2001 *Comput. Phys. Commun.* **135** 329–347 ISSN 0010-4655
[9] Tackett A, Holzwarth N and Matthews G 2001 *Comput. Phys. Commun.* **135** 348–376 ISSN 0010-4655
[10] Mortensen J J, Hansen L B and Jacobsen K W 2005 *Phys. Rev. B* **71** 035109
[11] Giannozzi P, Baroni S, Bonini N, Calandra M, Car R, Cavazzoni C, Ceresoli D, Chiarotti G L, Cococcioni M, Dabo I, Dal Corso A, de Gironcoli S, Fabris S, Fratesi G, Gebauer R, Gerstmann U, Gougoussis C, Kokalj A, Lazzeri M, Martin-Samos L, Marzari N, Mauri F, Mazzarello R, Paolini S, Pasquarello A, Paulatto L, Sbraccia C, Scandolo S, Sclauzero G, Seitsonen A P, Smogunov A, Umari P and Wentzcovitch R M 2009 *J. Phys.: Cond. Mat.* **21** 395502 ISSN 0953-8984, 1361-648X
[12] Torrent M, Jollet F, Bottin F, Zérah G and Gonze X 2008 *Comput. Mater. Sci.* **42** 337–351 ISSN 0927-0256
[13] Bowler D R and Miyazaki T 2012 *Rep. Prog. Phys.* **75** 036503 ISSN 0034-4885
[14] McWeeny R 1960 *Rev. Mod. Phys.* **32** 335–369
[15] Hernandez E and Gillan M J 1995 *Phys. Rev. B* **51** 10157–10160
[16] Prodan E and Kohn W 2005 *Proc. Natl. Acad. Sci. U. S. A.* **102** 11635–11638
[17] Skylaris C K and Haynes P D 2007 *J. Chem. Phys.* **127** 164712 ISSN 00219606 URL `http://link.aip.org/link/JCPSA6/v127/i16/p164712/s1&Agg=doi`
[18] Skylaris C K, Mostofi A A, Haynes P D, Dieguez O and Payne M C 2002 *Phys. Rev. B* **66** 035119
[19] Skylaris C K, Haynes P D, Mostofi A A and Payne M C 2005 *J. Chem. Phys.* **122** 084119
[20] Hine N D M, Robinson M, Haynes P D, Skylaris C K, Payne M C and Mostofi A A 2011 *Phys. Rev. B* **83** 195102
[21] Heiss M, Fontana Y, Gustafsson A, Wust G, Magen C, O'Regan D D, Luo J W, Ketterer B, Conesa-Boj S, Kuhlmann A V, Houel J, Russo-Averchi E, Morante J R, Cantoni M, Marzari N, Arbiol J, Zunger A, Warburton R J and Fontcuberta i Morral A 2013 *Nature Mat.* **12** 439–444
[22] Corsini N R C, Zhang Y, Little W R, Karatutlu A, Ersoy O, Haynes P D, Molteni C, Hine N D M, Hernandez I, Gonzalez J, Rodriguez F, Brazhkin V V and Sapelkin A 2015 *Nano Letters* **15** 7334–7340 pMID: 26457875
[23] Hine N D M, Haynes P D, Mostofi A A, Skylaris K and Payne M C 2009 *Comput. Phys. Commun.* **180** 1041–1053
[24] Wilkinson K A, Hine N D and Skylaris C K 2014 *J. Chem. Theory Comput.*
[25] Constantinescu G C and Hine N D M 2015 *Phys. Rev. B* **91** 195416
[26] Constantinescu G C and Hine N D M 2016 *Nano Lett.* **16** 2586–2594 ISSN 1530-6984
[27] Corsini N R C, Hine N D M, Haynes P D and Molteni C 2016 *Under review*
[28] Tait E W, Ratcliff L E, Payne M C, Haynes P D and Hine N D M 2016 *J. Phys.: Condens. Matter* **28** 195202 ISSN 0953-8984
[29] Hine N D M, Haynes P D, Mostofi A A and Payne M C 2010 *J. Chem. Phys.* **133** 114111 ISSN 00219606
[30] Torrent M, Holzwarth N, Jollet F, Harris D, Lepley N and Xu X 2010 *Comput. Phys. Commun.* **181** 1862–1867 ISSN 0010-4655
[31] Skylaris C K, Mostofi A A, Haynes P D, Pickard C J and Payne M C 2001 *Comput. Phys. Commun.* **140** 315–322
[32] Ruiz-Serrano A, Hine N D M and Skylaris C K 2012 *J. Chem. Phys.* **136** 234101–234101–9 ISSN 00219606

[33] Palser A H R and Manolopoulos D E 1998 *Phys. Rev. B* **58** 12704
[34] Haynes P D, Skylaris C K, Mostofi A A and Payne M C 2008 *J. Phys. Cond. Mat.* **20** 294207
[35] Li X P, Nunes R W and Vanderbilt D 1993 *Phys. Rev. B* **47** 10891–10894
[36] Lejaeghere K, Bihlmayer G, Björkman T, Blaha P, Blügel S, Blum V, Caliste D, Castelli I E, Clark S J, Dal Corso A, de Gironcoli S, Deutsch T, Dewhurst J K, Di Marco I, Draxl C, Dułak M, Eriksson O, Flores-Livas J A, Garrity K F, Genovese L, Giannozzi P, Giantomassi M, Goedecker S, Gonze X, Grånäs O, Gross E K U, Gulans A, Gygi F, Hamann D R, Hasnip P J, Holzwarth N A W, Iuşan D, Jochym D B, Jollet F, Jones D, Kresse G, Koepernik K, Küçükbenli E, Kvashnin Y O, Locht I L M, Lubeck S, Marsman M, Marzari N, Nitzsche U, Nordström L, Ozaki T, Paulatto L, Pickard C J, Poelmans W, Probert M I J, Refson K, Richter M, Rignanese G M, Saha S, Scheffler M, Schlipf M, Schwarz K, Sharma S, Tavazza F, Thunström P, Tkatchenko A, Torrent M, Vanderbilt D, van Setten M J, Van Speybroeck V, Wills J M, Yates J R, Zhang G X and Cottenier S 2016 *Science* **351** ISSN 0036-8075
[37] Jollet F, Torrent M and Holzwarth N 2014 *Comput. Phys. Commun.* **185** 1246 – 1254 ISSN 0010-4655
[38] Zuehlsdorff T J, Hine N D M, Spencer J S, Harrison N M, Riley D J and Haynes P D 2013 *J. Chem. Phys.* **139** 064104 ISSN 0021-9606, 1089-7690
[39] Zuehlsdorff T J, Hine N D M, Payne M C and Haynes P D 2015 *J. Chem. Phys.* **143** 204107 ISSN 0021-9606, 1089-7690