ELSEVIER

Computer Physics Communications 147 (2002) 788–802

Computer Physics Communications

www.elsevier.com/locate/cpc

Total-energy calculations on a real space grid with localized functions and a plane-wave basis

Arash A. Mostofi *, Chris-Kriton Skylaris, Peter D. Haynes, Mike C. Payne

Theory of Condensed Matter, Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK

Received 19 December 2001; accepted 1 February 2002

Abstract

We present a novel real space formalism for ab initio electronic structure calculations. We use localized non-orthogonal functions that are expressed in terms of a basis set that is equivalent to a plane-wave basis. As a result, advantages of the plane-wave approach also apply to our method: its applicability to any lattice symmetry, and systematic basis set improvement via the kinetic energy cut-off parameter. The localization of our functions enables the use of fast Fourier transforms over small regions of the simulation cell to calculate the total energy with efficiency and accuracy. With just one further variational approximation, namely the truncation of the density matrix, the calculation may be performed with a cost that scales linearly with system size for insulating systems.

PACS: 71.15.-m; 31.15.-p; 31.15.Ar

Keywords: Electronic structure; Density functional theory; Linear-scaling; FFT; Systematic basis set

1. Introduction

Density functional theory (DFT) in conjunction with the plane-wave pseudopotential method has been widely used as a theoretical tool for studying a variety of condensed matter systems [1]. The original formulation of Kohn and Sham [2] in terms of extended orthonormal wavefunctions has a computational cost that scales asymptotically as the cube of the system size. This bottleneck restricts the approach to the study of no more than a few hundred atoms, even on parallel supercomputers. As a consequence, over the last decade there has been much effort devoted to the development of methods whose computational cost scales linearly with system size [3]. Such schemes often rely on a reformulation of the problem in terms of localized functions in real space [4]. One such set of functions are the orthogonal Wannier functions: a unitary transformation of the extended Bloch wavefunctions [5]. These (and as a result the density matrix) are known to be exponentially localized in insulators, with the degree of localization determined by the band gap [6–9]. Furthermore, it is known that an essentially equivalent representation that is non-orthogonal can be better localized [10] and hereafter we shall refer to such localized functions as non-orthogonal generalized Wannier functions (NGWFs).

* Corresponding author.
E-mail address: aam24@phy.cam.ac.uk (A.A. Mostofi).

0010-4655/02/$ – see front matter
PII: S0010-4655(02)00461-7

In our approach we formulate the total energy of the system in terms of NGWFs that are represented on a real space grid subject to periodic boundary conditions. The localization properties of the NGWFs are set *a priori*, i.e. each NGWF (usually atom centered) is confined within a spherical localization region (LR) whose radius is pre-defined. In general, each NGWF may have a different localization radius. Our basis set consists of a mesh of cell periodic functions which are a unitary transformation of an equivalent plane-wave basis. As a result, we can use fast Fourier transform (FFT) methods familiar from plane-wave DFT calculations to switch easily between real and reciprocal space representations.

Real space techniques have been used by many authors for performing DFT calculations. In particular, there are methods which use functions that are strictly localized within spherical regions on real space grids [11,12]. These approaches have given rise to some important methodological developments. However, it would be desirable to have a method with a basis set that can be improved systematically [13,14], just as in plane-wave DFT. Our approach uses a basis that is directly comparable to a plane-wave basis.

We begin by describing our basis set and its properties in Section 2. In Section 3 we describe how to calculate the total energy of a system with a set of localized functions and in Section 4 we introduce our novel FFT box technique and demonstrate how it can be used to calculate the total energy with a computational cost that scales linearly with system size. Section 5 describes how we minimize the total energy, and we present our results and conclusions in Sections 6 and 7, respectively.

2. Basis set

We consider a unit cell (which we shall refer to as the simulation cell) with primitive lattice vectors $\mathbf{A}_i$ ($i = 1, 2, 3$), volume $V = |\mathbf{A}_1 \cdot (\mathbf{A}_2 \times \mathbf{A}_3)|$, and $N_i = 2J_i + 2$ grid points along direction i, where the J_i are integers. We define our basis functions to be the cell periodic, bandwidth limited Dirac delta functions (Fig. 1) given by,

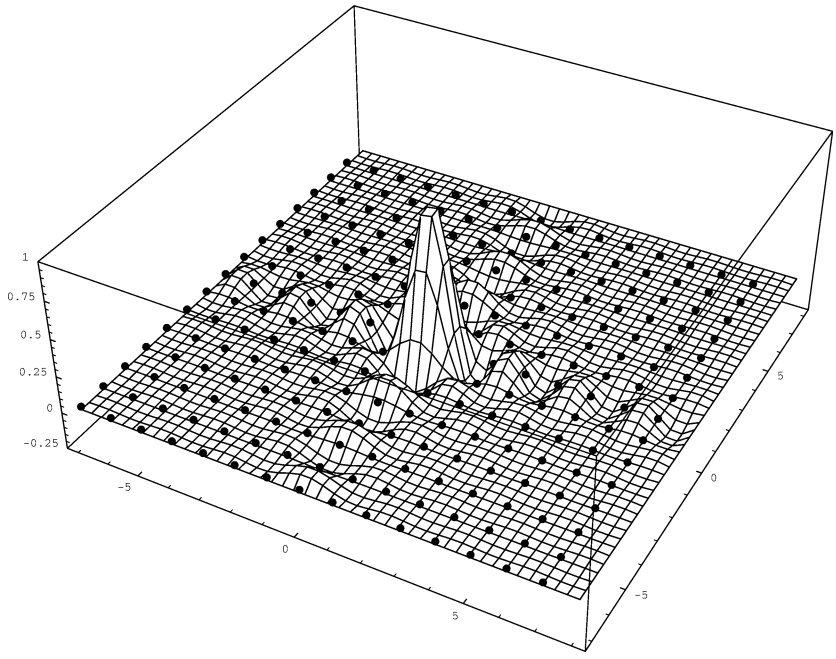

Fig. 1. The form of a basis function, $D(\mathbf{r})$, in two-dimensions.

$$D_{KLM}(\mathbf{r}) \equiv D(\mathbf{r}-\mathbf{r}_{KLM}) = \frac{1}{N_1N_2N_3}\sum_{p=-J_1}^{J_1+1}\sum_{q=-J_2}^{J_2+1}\sum_{s=-J_3}^{J_3+1} \mathrm{e}^{\mathrm{i}(p\mathbf{B}_1+q\mathbf{B}_2+s\mathbf{B}_3)\cdot(\mathbf{r}-\mathbf{r}_{KLM})}, \tag{1}$$

where p, q and s are integers, and the $\mathbf{B}_i$ are the reciprocal lattice vectors:

$$\mathbf{B}_1 = \frac{2\pi(\mathbf{A}_2\times\mathbf{A}_3)}{\mathbf{A}_1\cdot(\mathbf{A}_2\times\mathbf{A}_3)}, \quad \text{etc.} \tag{2}$$

The NGWFs $\{\phi_\alpha\}$ are expanded in terms of our basis $D(\mathbf{r})$,

$$\phi_\alpha(\mathbf{r}) = \sum_{KLM} C_{KLM,\alpha} D(\mathbf{r}-\mathbf{r}_{KLM}), \tag{3}$$

where the $C_{KLM,\alpha}$ are the expansion coefficients of $\phi_\alpha(\mathbf{r})$ in the basis $D(\mathbf{r})$ and the sum is over all the grid points of the simulation cell,

$$\mathbf{r}_{KLM} = \frac{K}{N_1}\mathbf{A}_1 + \frac{L}{N_2}\mathbf{A}_2 + \frac{M}{N_3}\mathbf{A}_3, \tag{4}$$

where K, L, and M are integers.

There is one basis function centered on each grid point of the simulation cell. They have the property that they are non-zero at the grid point on which they are centered and zero at all other grid points (A.1). This basis spans the same Hilbert space as the basis of plane-waves that can be represented by the real space grid of our simulation cell: a unitary transformation relates the two. Further properties are derived in Appendix A.

Due to the localization of the NGWFs, the expansion coefficient, $C_{KLM,\alpha}$, of a particular NGWF, ϕ_α, is equal to zero if the grid point $\mathbf{r}_{KLM}$ does not fall within the LR of ϕ_α. Consequently, because the size of each LR is independent of system size, each NGWF is expanded in terms of a number of basis functions that is independent of system size.

3. Total energy

In terms of the set of NGWFs $\{\phi_\alpha\}$, the single-particle density matrix in the co-ordinate representation is expressed as

$$\rho(\mathbf{r},\mathbf{r}') = \phi_\alpha(\mathbf{r})K^{\alpha\beta}\phi_\beta(\mathbf{r}'), \tag{5}$$

where the density kernel, $K^{\alpha\beta}$, is the matrix representation of the density matrix in terms of the contravariant duals of the NGWFs [15,16]. The NGWFs are real, as we are concerned only with calculations at the Γ-point. Summation over repeated Greek indices is assumed throughout.

The charge density is given by,

$$n(\mathbf{r}) = 2\rho(\mathbf{r},\mathbf{r}) = 2K^{\alpha\beta}\phi_\alpha(\mathbf{r})\phi_\beta(\mathbf{r}) \equiv 2K^{\alpha\beta}\rho_{\alpha\beta}(\mathbf{r}), \tag{6}$$

where $\rho_{\alpha\beta}(\mathbf{r})$ is defined as,

$$\rho_{\alpha\beta}(\mathbf{r}) \equiv \phi_\alpha(\mathbf{r})\phi_\beta(\mathbf{r}) = \sum_{X=0}^{2N_1-1}\sum_{Y=0}^{2N_2-1}\sum_{Z=0}^{2N_3-1} \rho_{\alpha\beta}(\mathbf{r}_{XYZ})B_{XYZ}(\mathbf{r}). \tag{7}$$

This quantity is expanded in terms of the basis functions of the fine grid, $B_{XYZ}(\mathbf{r})$ (see Appendix A), because the density needs to be represented in terms of a basis with twice the cut-off frequency of the NGWFs.

Neglecting the contribution due to the fixed background of ionic charges, the total energy, $E[n]$, of the system is the sum of the kinetic energy ($E_{\rm K}$), the Hartree energy ($E_{\rm H}$), the local pseudopotential energy ($E_{\rm loc}$), the non-local pseudopotential energy ($E_{\rm nl}$), and the exchange-correlation energy ($E_{\rm xc}$).

The kinetic energy is given by

$$E_{\rm K}[n] = -\int_V {\rm d}\mathbf{r}' \left[\nabla^2_{\mathbf{r}}\rho(\mathbf{r},\mathbf{r}')\right]_{\mathbf{r}=\mathbf{r}'} = 2K^{\alpha\beta}\langle\phi_\beta|\widehat{T}|\phi_\alpha\rangle, \tag{8}$$

where $\widehat{T}$ is related to the Laplacian operator by $\widehat{T} = -\frac{1}{2}\nabla^2$. This is most conveniently applied in reciprocal space, where it is diagonal. We see that the kinetic energy involves matrix elements of the form $\langle\phi_\beta|\widehat{T}|\phi_\alpha\rangle$. Thus, one way of proceeding would be to FFT ϕ_α in the whole simulation cell, apply the Laplacian, FFT back and perform the integral over real space as a summation over grid points as suggested by Eq. (A.5). If N is the number of grid points in the simulation cell (which is proportional to system size), this procedure has a computational cost that scales as $\mathrm{O}(N\log N)$ for each NGWF and hence as $\mathrm{O}(N^2\log N)$ overall. We will see later how linear-scaling can be achieved for all the terms that compose $E[n]$.

The Hartree energy is given by,

$$\begin{aligned} E_{\rm H}[n] &= \frac{1}{2}\iint {\rm d}\mathbf{r}\,{\rm d}\mathbf{r}'\,\frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} = \frac{1}{2}\int {\rm d}\mathbf{r}\,V_{\rm H}(\mathbf{r})n(\mathbf{r}) \\ &= K^{\alpha\beta}\langle\phi_\beta|V_{\rm H}|\phi_\alpha\rangle, \end{aligned} \tag{9}$$

where the Hartree potential $V_{\rm H}(\mathbf{r})$ is,

$$V_{\rm H}(\mathbf{r}) = \int {\rm d}\mathbf{r}'\,\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} = \sum_{X=0}^{2N_1-1}\sum_{Y=0}^{2N_2-1}\sum_{Z=0}^{2N_3-1} V_{\rm H}(\mathbf{r}_{XYZ})B_{XYZ}(\mathbf{r}). \tag{10}$$

We see that $V_{\rm H}$ is the result of a convolution between the charge density and the Coulomb potential, and is thus represented by the fine grid basis functions.

$V_{\rm H}$ is calculated on the fine grid in reciprocal space, where the real space convolution becomes a simple product, and fast Fourier transformed back to real space. ϕ_α is Fourier interpolated onto the fine grid and its product with $V_{\rm H}$ is taken. The result is Fourier filtered to the standard grid, and the matrix elements $\langle\phi_\beta|[V_{\rm H}\phi_\alpha]_D\rangle$ calculated by summation over the grid points of the *standard* grid, where the subscript D shows that we only need to consider frequency components represented by the basis functions $\{D(\mathbf{r})\}$ as shown in Eq. (A.5).

The local pseudopotential energy is given by the integral

$$E_{\rm loc}[n] = \int {\rm d}\mathbf{r}\,V_{\rm loc}(\mathbf{r})n(\mathbf{r}) = \int {\rm d}\mathbf{r}\left[V_{\rm loc}(\mathbf{r})\right]_B n(\mathbf{r}), \tag{11}$$

where the second equality makes use of Eq. (A.9), and the subscript B shows that only frequency components represented by the fine grid basis functions, $B_{XYZ}(\mathbf{r})$, need to be considered. Thus, it may be calculated in the same way as the Hartree energy:

$$E_{\rm loc}[n] = 2K^{\alpha\beta}\langle\phi_\beta|[V_{\rm loc}]_B|\phi_\alpha\rangle = 2K^{\alpha\beta}\langle\phi_\beta|\left([V_{\rm loc}]_B\phi_\alpha\right)_D\rangle. \tag{12}$$

The non-local pseudopotential energy is given by

$$E_{\rm nl} = 2K^{\alpha\beta}\langle\phi_\beta|\widehat{V}_{\rm nl}|\phi_\alpha\rangle, \tag{13}$$

where the non-local pseudopotential operator $\widehat{V}_{\rm nl}$ that we use is in Kleinman–Bylander form [17],

$$\widehat{V}_{\rm nl} = \sum_I \sum_{lm(I)} \frac{|\delta\widehat{V}_l^{(I)}\Psi_{lm}^{(I)}\rangle\langle\Psi_{lm}^{(I)}\delta\widehat{V}_l^{(I)}|}{\langle\Psi_{lm}^{(I)}|\delta\widehat{V}_l^{(I)}|\Psi_{lm}^{(I)}\rangle}, \tag{14}$$

where the first summation is over the atoms I, and the second is over the angular momentum components for each atom. $\delta\widehat{V}_l^{(I)}$ is the angular momentum dependent part of the atomic pseudopotential and the $\Psi_{lm}^{(I)}$ are its associated pseudo-orbitals. We see that the matrix elements of the non-local potential involve the calculation of integrals of the form $\langle\phi_\beta|\delta\widehat{V}_l^{(I)}\Psi_{lm}^{(I)}\rangle = \langle\phi_\beta|[\delta\widehat{V}_l^{(I)}\Psi_{lm}^{(I)}]_D\rangle$, which may be performed by summation over the standard real space grid.

The exchange-correlation energy within the LDA is given by approximating the integral over the exchange-correlation energy density, $\epsilon_{\rm xc}(n(\mathbf{r}))$, by a summation over the points of the fine grid:

$$E_{\rm xc}[n] = \int_V d\mathbf{r}\,\epsilon_{\rm xc}\big(n(\mathbf{r})\big)n(\mathbf{r}) \simeq \frac{V}{8N_1N_2N_3}\sum_{X=0}^{2N_1-1}\sum_{Y=0}^{2N_2-1}\sum_{Z=0}^{2N_3-1}\epsilon_{\rm xc}\big(n(\mathbf{r}_{XYZ})\big)n(\mathbf{r}_{XYZ}). \tag{15}$$

This is approximate as the fine grid cannot represent the exchange-correlation energy density without some aliasing. The errors associated with this approximation are expected to be similar to those encountered in traditional plane-wave codes [1].

We may write the total energy as,

$$E[n] = 2K^{\alpha\beta}H_{\beta\alpha} - E_{\rm DC}[n], \tag{16}$$

where the matrix elements of the Kohn–Sham Hamiltonian are given by,

$$H_{\beta\alpha} = \langle\phi_\beta|\big[\widehat{T} + V_{\rm H}(\mathbf{r}) + V_{\rm loc}(\mathbf{r}) + \widehat{V}_{\rm nl} + V_{\rm xc}(\mathbf{r})\big]|\phi_\alpha\rangle, \tag{17}$$

$V_{\rm xc}(\mathbf{r}) = \frac{\delta E_{\rm xc}[n]}{\delta n(\mathbf{r})}$ is the exchange-correlation potential, and $E_{\rm DC}[n]$ is the double-counting correction,

$$E_{\rm DC}[n] = \frac{1}{2}\int d\mathbf{r}\,n(\mathbf{r})V_{\rm H}(\mathbf{r}) + \int d\mathbf{r}\,n(\mathbf{r})V_{\rm xc}(\mathbf{r}) - E_{\rm xc}[n]. \tag{18}$$

As we are dealing with a sparse system, the number of non-zero matrix elements will be proportional to the system size in the limit of large systems. The procedure outlined above for calculating each of these matrix elements has a computational cost that scales as $\mathrm{O}(N\log N)$ for each element. This is because FFTs of NGWFs are performed over the entire simulation cell. Thus the cost of computing all non-zero matrix elements scales as $\mathrm{O}(N^2\log N)$.

4. Total energy using the FFT box technique

4.1. The FFT box

Linear-scaling can be achieved by restricting the number of plane waves that each basis function is comprised of so that it is independent of system-size. In other words, because the NGWFs are strictly zero outside their respective localization regions, we need not use the entire simulation cell to perform FFTs on them, but may use a much smaller region. This must be done in such a way as to preserve the Hermiticity of operators, i.e. so that

$$O_{\alpha\beta} \equiv \langle\phi_\alpha|\widehat{O}|\phi_\beta\rangle = O_{\beta\alpha}. \tag{19}$$

Also, it is important that the representation of an operator in this new, contracted basis is consistent throughout the calculation: when we compute the two matrix elements $O_{\alpha\beta}$ and $O_{\gamma\beta}$, in both cases we require the quantity $\widehat{O}|\phi_\beta\rangle$, and in both cases this quantity should be the same.

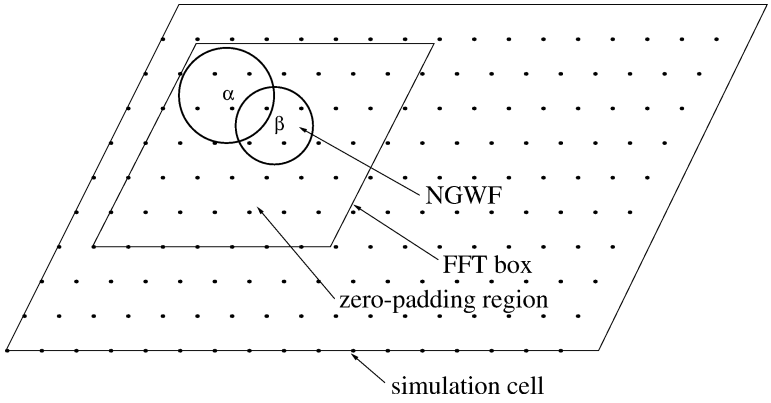

Fig. 2. The simulation cell and FFT box for a pair of overlapping NGWFs, $|\phi_\alpha\rangle$ and $|\phi_\beta\rangle$.

We investigated these matters in detail in the context of the kinetic energy in an earlier communication [18], and we found an efficient and accurate solution that we call the 'FFT box' technique, which involves performing computations over *restricted regions* of the simulation cell, using the fact that the NGWFs are localized in real space. First we shall define the FFT box and then show in detail how it is used to compute the matrix elements of each component of the Hamiltonian in $\mathrm{O}(N)$ operations.

We define the FFT box to be a miniature and commensurate version of the simulation cell whose size is such that it can contain any pair of NGWFs that exhibit any degree of overlap. Its dimensions and shape are determined at the start of a calculation and are universal throughout that calculation. It should have the same spacing of grid points in each lattice vector direction as the simulation cell, and its origin (which is in general different for the calculation of each matrix element) should coincide with a particular grid point of the simulation cell (Fig. 2). Treating the FFT box as a miniature simulation cell with $n_i = 2j_i + 2$ points along lattice vector $\mathbf{a}_i$ (where the j_i are integers), and with volume $v = |\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)|$, we may define a set of basis functions, $\{d_{klm}(\mathbf{r})\}$, as we did for the whole simulation cell in Eq. (1), as follows,

$$d_{klm}(\mathbf{r}) = \frac{1}{n_1 n_2 n_3} \sum_{p=-j_1}^{j_1+1} \sum_{q=-j_2}^{j_2+1} \sum_{s=-j_3}^{j_3+1} \mathrm{e}^{\mathrm{i}(p\mathbf{b}_1+q\mathbf{b}_2+s\mathbf{b}_3)\cdot(\mathbf{r}-\mathbf{r}_{klm})}, \tag{20}$$

where the sum runs over the reciprocal lattice vectors, $\{\mathbf{b}_i\}$, of the FFT box (i.e. a number of plane-waves that is now *independent* of system size), and

$$\mathbf{r}_{klm} = \frac{k}{n_1}\mathbf{a}_1 + \frac{l}{n_2}\mathbf{a}_2 + \frac{m}{n_3}\mathbf{a}_3 \tag{21}$$

for integer k, l, and m. These basis functions have the periodicity of the FFT box, and all the analytic properties derived for the basis functions of the simulation cell in Appendix A carry over to these functions by making the replacements

$$\begin{aligned} V &\longrightarrow v \\ N_i &\longrightarrow n_i \\ \mathbf{A}_i &\longrightarrow \mathbf{a}_i \\ \mathbf{B}_i &\longrightarrow \mathbf{b}_i \\ D_{KLM}(\mathbf{r}) &\longrightarrow d_{klm}(\mathbf{r}) \\ B_{XYZ}(\mathbf{r}) &\longrightarrow b_{xyz}(\mathbf{r}). \end{aligned} \tag{22}$$

4.2. Projection operators

We must also introduce an operator, $\widehat{P}(\alpha\beta)$, that is required to project NGWFs between their representation in terms of the plane-waves of the entire simulation cell and those of the FFT box. This operator has arguments α and β as it is dependent upon the exact location of the FFT box within the simulation cell, which in turn depends upon the particular matrix element that is being calculated. We thus define for the pair of NGWFs ϕ_α and ϕ_β:

$$\widehat{P}(\alpha\beta) = \frac{1}{\Omega} \sum_{k=0}^{n_1-1} \sum_{l=0}^{n_2-1} \sum_{m=0}^{n_3-1} |d_{klm}\rangle \big\langle D_{(k+K_{\alpha\beta})(l+L_{\alpha\beta})(m+M_{\alpha\beta})} \big|, \tag{23}$$

where $K_{\alpha\beta}$, $L_{\alpha\beta}$ and $M_{\alpha\beta}$ are integers denoting the grid point of the simulation cell at which the origin of the FFT box ($k = l = m = 0$) is located, and Ω is the volume per grid point.

When this operator acts upon a function with the periodicity of the simulation cell, i.e. a function given by Eq. (3), it maps it onto the 'same' function with the periodicity of the FFT box:

$$\widehat{P}(\alpha\beta)\phi_\alpha(\mathbf{r}) = \sum_{k=0}^{n_1-1} \sum_{l=0}^{n_2-1} \sum_{m=0}^{n_3-1} c_{klm,\alpha} d_{klm}(\mathbf{r}), \tag{24}$$

where $c_{klm,\alpha} \equiv C_{(k+K_{\alpha\beta})(l+L_{\alpha\beta})(m+M_{\alpha\beta}),\alpha}$.

Similarly, we define a supplementary operator, $\widehat{Q}(\alpha\beta)$, that performs the same task, but for the fine grid representation:

$$\widehat{Q}(\alpha\beta) = \frac{8}{\Omega} \sum_{x=0}^{2n_1-1} \sum_{y=0}^{2n_2-1} \sum_{z=0}^{2n_3-1} |b_{xyz}\rangle \big\langle B_{(x+2K_{\alpha\beta})(y+2L_{\alpha\beta})(z+2M_{\alpha\beta})} \big|. \tag{25}$$

The operators $\widehat{P}^\dagger(\alpha\beta)$ and $\widehat{Q}^\dagger(\alpha\beta)$ map functions from the FFT box back to the simulation cell on the standard and fine grids, respectively.

4.3. Kinetic energy

We want to calculate matrix elements of the form

$$T_{\alpha\beta} = \langle \phi_\alpha | \widehat{T} | \phi_\beta \rangle. \tag{26}$$

The NGWFs are localized in real space, so we need only consider those elements $T_{\alpha\beta}$ for which the localization regions of ϕ_α and ϕ_β overlap, as other contributions will be effectively zero. Once we have established that there is an overlap, we imagine the pair of NGWFs as being enclosed within the FFT box that has been defined for the calculation, as shown in Fig. 2. We then apply the operator $\widehat{P}(\alpha\beta)$ to $|\phi_\beta\rangle$ to give it the periodicity of the FFT box. $\widehat{P}(\alpha\beta)|\phi_\beta\rangle$ may be then fast Fourier transformed to reciprocal space with a computational cost that scales as $O(N_{\text{box}} \log N_{\text{box}})$, where N_{box} is the number of grid points in the FFT box and is *independent of system-size*. $\widehat{T}$ is applied to $\widehat{P}(\alpha\beta)|\phi_\beta\rangle$ in reciprocal space by multiplying by $|\mathbf{k}|^2/2$ at each reciprocal lattice point, $\mathbf{k}$, in the FFT box. Performing another FFT, we obtain $\widehat{T}\widehat{P}(\alpha\beta)|\phi_\beta\rangle$ in the real space FFT box, again with a cost of $O(N_{\text{box}} \log N_{\text{box}})$. We may then apply $\widehat{P}^\dagger(\alpha\beta)$ to this to map it back into the simulation cell where the matrix element $T^{\text{box}}_{\alpha\beta}$, given by

$$\begin{aligned} T^{\text{box}}_{\alpha\beta} &= \big\langle \phi_\alpha \big| \widehat{P}^\dagger(\alpha\beta)\widehat{T}\widehat{P}(\alpha\beta) \big| \phi_\beta \big\rangle \\ &= \Omega \sum_{K,L,M \in \text{LR}^\alpha} C_{KLM,\alpha} \big[\widehat{P}^\dagger(\alpha\beta)\widehat{T}\widehat{P}(\alpha\beta)\phi_\beta \big](\mathbf{r}_{KLM}) \end{aligned} \tag{27}$$

is calculated by summation over the grid points that lie within the localization region of ϕ_α (LR^α). The superscript 'box' signifies that a quantity has been calculated using the FFT box technique. We will show in Section 6 that $T_{\alpha\beta}^{\mathrm{box}}$ is an accurate approximation to $T_{\alpha\beta}$. This is because the FFT box technique is essentially a method of coarse-sampling the frequency components of the NGWFs in reciprocal space. As the NGWFs are truly localized in real space, we expect them to be very smooth in reciprocal space, and thus amenable to coarse-sampling.

For a single matrix element, the FFT box method makes the cost of calculation independent of system size. Thus, for all non-zero matrix elements the cost scales as $\mathrm{O}(N)$.

4.4. Non-local pseudopotential energy

For clarity of notation, we first rewrite the non-local pseudopotential of Eq. (14) as

$$\widehat{V}_{\mathrm{nl}} \equiv \sum_I \sum_{lm(I)} \left|\chi_{lm}^{(I)}\right\rangle\left\langle\chi_{lm}^{(I)}\right|, \tag{28}$$

which defines the *projectors* $|\chi_{lm}^{(I)}\rangle$.

Robust real space methods to calculate the non-local pseudopotential energy exist in the context of traditional plane-wave DFT through the work of King-Smith et al. [19]. We manage to avoid entirely the complications involved in their method through use of the FFT box.

We need to calculate matrix elements such as

$$V_{\mathrm{nl},\alpha\beta} = \langle\phi_\alpha|\widehat{V}_{\mathrm{nl}}|\phi_\beta\rangle = \sum_I \sum_{lm(I)} \left\langle\phi_\alpha\middle|\chi_{lm}^{(I)}\right\rangle\left\langle\chi_{lm}^{(I)}\middle|\phi_\beta\right\rangle, \tag{29}$$

which is just a matter of computing quantities like $\langle\phi_\alpha|\chi_{lm}^{(I)}\rangle$. There will only be a contribution to a particular matrix element $V_{\mathrm{nl},\alpha\beta}$ if there is at least one projector that overlaps with both $|\phi_\alpha\rangle$ and $|\phi_\beta\rangle$ (see Fig. 3). We begin with the radial part of the projectors, $\zeta_l^{(I)}(k)$, on a reciprocal space, linear, radial grid, up to arbitrarily high wavevector. Given that the overlap condition with the NGWFs is satisfied, we then continue to calculate each overlap in turn. For example, to compute $\langle\phi_\alpha|\chi_{lm}^{(I)}\rangle$, we envisage a real space FFT box that contains the NGWF $|\phi_\alpha\rangle$ and the projector $|\chi_{lm}^{(I)}\rangle$. The reciprocal representation of the projector is interpolated onto the corresponding reciprocal space FFT box using

$$\chi_{lm}^{(I)}(\mathbf{k}) = \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}_{(I)}} 4\pi(-i)^l Z_{lm}(\Omega_{\mathbf{k}})\zeta_l^{(I)}(k), \tag{30}$$

where $\mathbf{R}_{(I)}$ is the position vector of atom I, and Z_{lm} are real spherical harmonics.

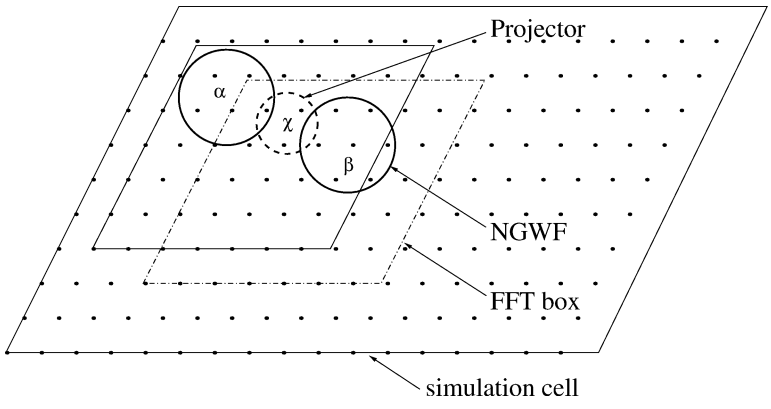

Fig. 3. An example of a typical contribution to the non-local potential matrix element $V_{\mathrm{nl},\alpha\beta}$. The overlap of the projector, $|\chi\rangle$, with $|\phi_\alpha\rangle$ is calculated using the FFT box with solid outline, and that with $|\phi_\beta\rangle$ is done using the FFT box with dashed outline.

Performing a FFT on the projector, we obtain it in the real space FFT box, project it into the simulation cell, and take the overlap with the NGWF by summation over the grid points that lie within the localization region of $|\phi_\alpha\rangle$. In terms of the projection operator, $\widehat{P}$, this process is represented by

$$V^{\text{box}}_{\text{nl},\alpha\beta} = \sum_I \sum_{lm(I)} \langle \phi_\alpha | \widehat{P}^\dagger | \chi^{(I)}_{lm} \rangle \langle \chi^{(I)}_{lm} | \widehat{P} | \phi_\beta \rangle. \tag{31}$$

Since the FFTs are performed only on a restricted region of the simulation cell, namely the FFT box, and because the NGWFs and atom cores are strictly localized, the cost of calculation of the non-local pseudopotential matrix in this way scales as $\mathrm{O}(N)$.

4.5. Charge density

The charge density of Eq. (6) is a quantity that extends over the entire simulation cell. The individual contributions to the charge density, i.e. the $\rho_{\alpha\beta}$ of Eq. (7), however, are localized in real space and may thus be calculated using the FFT box with a cost that is independent of system-size. For a given pair of overlapping NGWFs, $|\phi_\alpha\rangle$ and $|\phi_\beta\rangle$, we project them from the simulation cell into their FFT box. Both are then interpolated onto the fine grid of the FFT box using fast Fourier transforms and zero-padding in reciprocal space. The cost associated with this procedure is system-size independent as it is done over the grid points of the FFT box only. The interpolated NGWFs are multiplied together on the fine grid points of the FFT box and the result projected back onto the fine grid of the simulation cell. In terms of our projection operators, this becomes

$$\rho^{\text{box}}_{\alpha\beta}(\mathbf{r}) = \widehat{Q}^\dagger(\alpha\beta)\big[\big(\widehat{P}(\alpha\beta)\phi_\alpha(\mathbf{r})\big)\big(\widehat{P}(\alpha\beta)\phi_\beta(\mathbf{r})\big)\big]. \tag{32}$$

The total charge density is then built up by summing all the contributions from the FFT boxes of pairs of overlapping NGWFs according to

$$n(\mathbf{r}) = 2K^{\alpha\beta}\rho^{\text{box}}_{\alpha\beta}(\mathbf{r}). \tag{33}$$

4.6. Hartree, local pseudopotential, and exchange-correlation energy

The matrix elements of the Hartree, the local pseudopotential and the exchange-correlation potential may be treated together:

$$\begin{aligned} V_{\text{Hlxc},\alpha\beta} &= \langle \phi_\alpha | \big[V_{\text{H}}(\mathbf{r}) + V_{\text{loc}}(\mathbf{r})\big]_B + V_{\text{xc}}(\mathbf{r}) | \phi_\beta \rangle \\ &\simeq \langle \phi_\alpha | \big(\big[V_{\text{H}}(\mathbf{r}) + V_{\text{loc}}(\mathbf{r}) + V_{\text{xc}}(\mathbf{r})\big]_B \phi_\beta\big)_D \rangle, \end{aligned} \tag{34}$$

where the approximation is due to the inability to faithfully represent the exchange-correlation energy-density on the fine grid. The operator defined by

$$\big[V_{\text{Hlxc}}(\mathbf{r})\big]_B \equiv \big[V_{\text{H}}(\mathbf{r}) + V_{\text{loc}}(\mathbf{r}) + V_{\text{xc}}(\mathbf{r})\big]_B \tag{35}$$

extends over the fine grid of the whole simulation cell. We calculate matrix elements $\langle \phi_\alpha | (V_{\text{Hlxc}})_B | \phi_\beta \rangle$, for a pair of overlapping NGWFs, $|\phi_\alpha\rangle$ and $|\phi_\beta\rangle$, by projecting $|\phi_\beta\rangle$ onto the FFT box that encloses the pair of functions. This NGWF is then interpolated onto the fine grid of the FFT box. $[V_{\text{Hlxc}}(\mathbf{r})]_B$ is projected onto the fine grid of the FFT box and its product taken with $|\phi_\beta\rangle$ on the grid points. The result is Fourier filtered onto the standard grid of the FFT box, thus keeping only frequency components represented through the standard grid basis functions, $D(\mathbf{r})$, and projected back onto the standard grid of the simulation cell. The matrix element is then obtained by computing the overlap with $|\phi_\alpha\rangle$ by summation over the grid points enclosed within its localization region. This procedure may be represented as

$$V^{\text{box}}_{\text{Hlxc},\alpha\beta} = \langle \phi_\alpha | \big(V^{\text{box}}_{\text{Hlxc}}\big)_B | \phi_\beta \rangle = \langle \phi_\alpha | \widehat{P}^\dagger(\alpha\beta)\big[\widehat{Q}(\alpha\beta)(V_{\text{Hlxc}})_B\big]\widehat{P}(\alpha\beta) | \phi_\beta \rangle, \tag{36}$$

and since all computations are done using the FFT box, the matrix elements can be calculated in $\mathrm{O}(N)$ operations.

Finally, the total energy, calculated using our FFT box method, may be written as

$$E^{\text{box}}[n] = 2K^{\alpha\beta}H^{\text{box}}_{\beta\alpha} - E^{\text{box}}_{\text{DC}}[n], \tag{37}$$

where $H^{\text{box}}_{\beta\alpha}$ is given by

$$H^{\text{box}}_{\beta\alpha} = T^{\text{box}}_{\beta\alpha} + V^{\text{box}}_{\text{nl},\beta\alpha} + V^{\text{box}}_{\text{Hlxc},\beta\alpha}\,, \tag{38}$$

and $E^{\text{box}}_{\text{DC}}[n]$ is as defined in Eq. (18), but with the charge density calculated according to Eq. (33). The FFT box method enables the sparse matrix represented by $H_{\beta\alpha}$ to be computed with an effort that scales linearly with system size. With just one extra variational approximation, namely truncation of the density kernel, the charge density of Eq. (33) and hence total energy of Eq. (37) may be calculated in $\mathrm{O}(N)$ operations.

5. Total energy optimization

The total energy is a functional of the charge density: $E = E[n(\mathbf{r})]$. The charge density itself is expanded in terms of the basis $\{B(\mathbf{r})\}$ and depends upon the density kernel elements, $\{K^{\alpha\beta}\}$, and the NGWF expansion coefficients, $\{C_{KLM,\alpha}\}$. Provided that it is N-representable, this dependence should be variational, i.e. the ground state energy, E_{min}, is given by

$$E_{\text{min}} = \min_{\{K^{\alpha\beta}\},\{C_{KLM,\alpha}\}} E\big(\{K^{\alpha\beta}\},\{C_{KLM,\alpha}\}\big). \tag{39}$$

In this work we are concerned principally with the optimization of the elements of the density kernel and we will consider the NGWF coefficients, $\{C_{KLM,\alpha}\}$, as being fixed. We use the pseudo-atomic orbitals (PAOs) of Sankey et al. [20] as our NGWFs.

The minimization must be performed under the constraints of constant electron number,

$$N_{\text{e}} = \int_V \mathrm{d}\mathbf{r}\, n(\mathbf{r}) = 2K^{\alpha\beta}S_{\beta\alpha} = 2\mathrm{Tr}[\mathbf{KS}], \tag{40}$$

and density-matrix idempotency,

$$\rho(\mathbf{r},\mathbf{r}') = \int_V \mathrm{d}\mathbf{r}''\,\rho(\mathbf{r},\mathbf{r}'')\rho(\mathbf{r}'',\mathbf{r}') \quad \Rightarrow \quad K^{\alpha\beta} = K^{\alpha\gamma}S_{\gamma\delta}K^{\delta\beta}, \tag{41}$$

where the overlap matrix, $S_{\alpha\beta}$ is given by

$$S_{\alpha\beta} = \int_V \mathrm{d}\mathbf{r}\,\phi_\alpha(\mathbf{r})\phi_\beta(\mathbf{r}). \tag{42}$$

In order to avoid explicitly imposing the idempotency constraint (41), we use the method suggested by Li, Nunes and Vanderbilt [21] and independently by Daw [22], and generalized to the case of non-orthogonal functions by Nunes and Vanderbilt [23]. Our implementation follows the simplified version of Millam and Scuseria [24]. We define the following function of the density kernel $\mathbf{K}$:

$$L(\mathbf{K}) = E(\widetilde{\mathbf{K}}) - \mu\big(2\mathrm{Tr}[\mathbf{KS}] - N_{\text{e}}\big), \tag{43}$$

where $\widetilde{\mathbf{K}}$ is the McWeeny purified density kernel [25],

$$\widetilde{\mathbf{K}} = 3\mathbf{KSK} - 2\mathbf{KSKSK}. \tag{44}$$

The contravariant, tensor-corrected gradient [15,16] that is used in the steepest descent or conjugate gradient iterative minimization is given by

$$\nabla L = \mathbf{S}^{-1}\frac{\partial L}{\partial \mathbf{K}}\mathbf{S}^{-1}$$
$$= 6\mathbf{KHS}^{-1} + 6\mathbf{S}^{-1}\mathbf{HK} - 4\mathbf{S}^{-1}\mathbf{HKSK} - 4\mathbf{KHK} - 4\mathbf{KSKHS}^{-1} - 2\mu\mathbf{S}^{-1}. \quad (45)$$

The value of μ is set at each step such that $\mathrm{Tr}[\mathbf{S}\nabla L] = 0$. This ensures that the total electron number remains unchanged, thus we simply require that our initial guess for the density kernel gives the correct number of electrons.

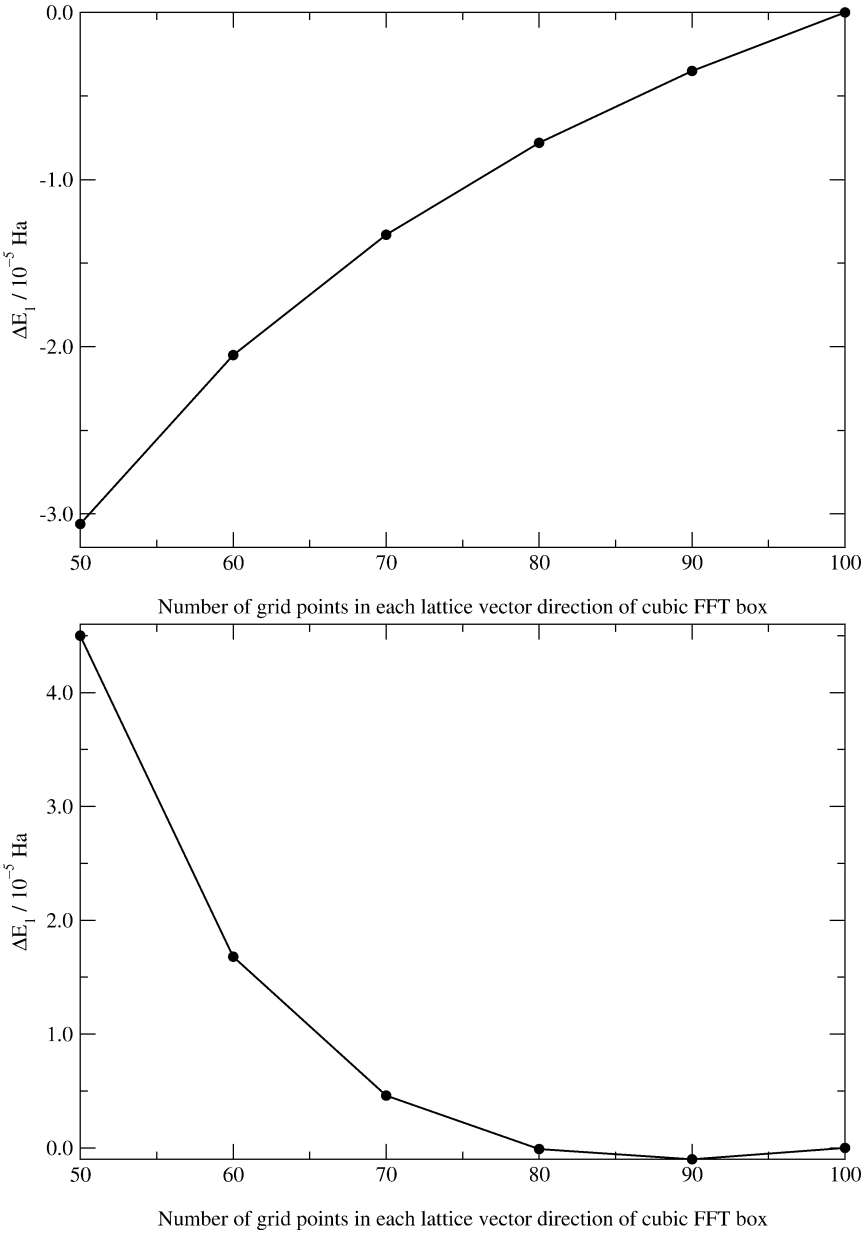

Fig. 4. Top panel: ΔE_1 plotted for a silane molecule as a function of FFT box size. Bottom panel: the same quantity plotted for cyclohexane. In both cases all PAOs were confined to atom centered localization regions with radius $6.0a_0$, and the grid spacing was $0.5a_0$.

6. Results and discussion

We now present some results of calculations we have performed using the methods described above. We first demonstrate the accuracy of the FFT box technique as compared to using the entire simulation cell as the FFT grid. Fig. 4 shows the quantity ΔE_1, defined as

$$\Delta E_1 \equiv E^{\text{box}}[n] - E[n], \tag{46}$$

for the molecules silane (SiH_4) and cyclohexane (C_6H_{12}), for different FFT box sizes. For these tests we used a cubic simulation cell of side length $50a_0$, although any other lattice symmetry could easily have been employed. The PAOs are confined within spherical regions of radius $6.0a_0$, and vanish exactly at the region boundary [20].

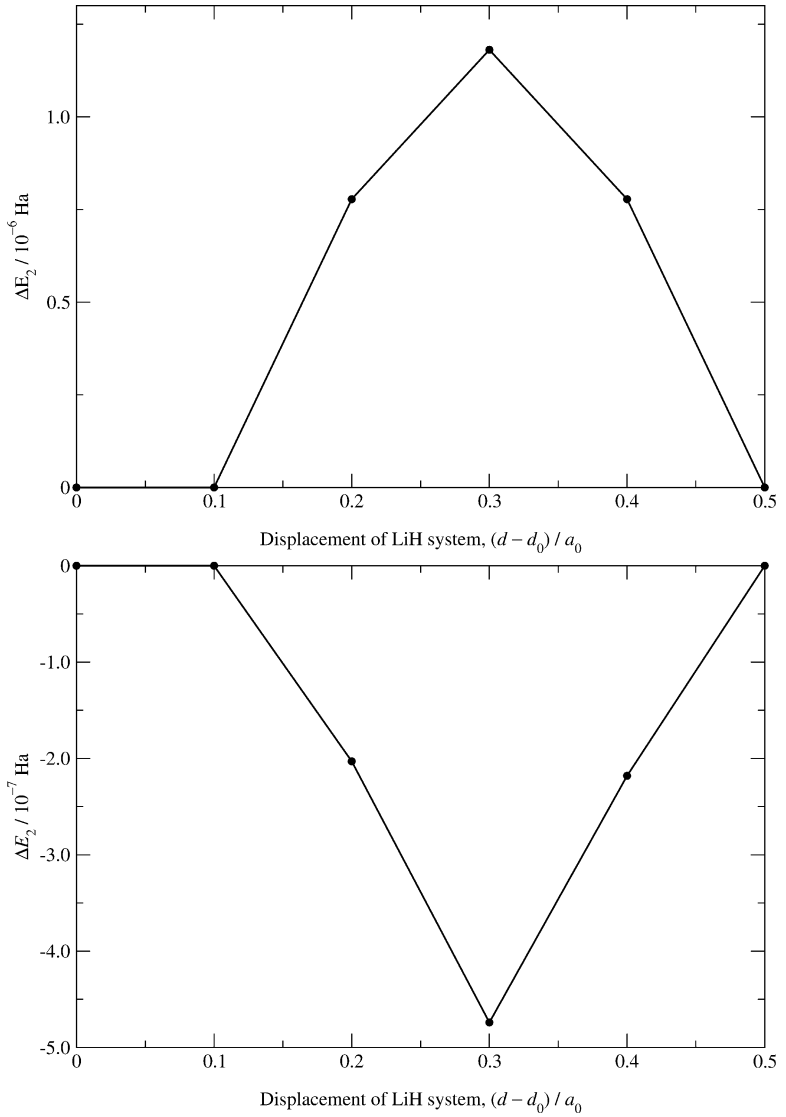

Fig. 5. Top panel: Variation in total energy, ΔE_2, as the LiH molecule is translated by fractions of a grid spacing. The PAOs on both the lithium and hydrogen atoms had a radius of $10.0a_0$. Bottom panel: The same quantity, but with optimization of the PAO coefficients [27].

They were generated using norm-conserving pseudopotentials [26], the local density approximation (LDA), and an energy cut-off of 538 eV (corresponding to our grid spacing of $0.5a_0$). In the case of silane, the silicon atom had one 3s and three 3p PAOs, and for cyclohexane, the carbon atoms had one 2s and three 2p PAOs. The hydrogen atoms all had a single 1s PAO.

We see that the error associated with the FFT box technique are of the order of a few μHartree per atom, which is comparable to the inherent errors present in plane-wave DFT. We note that the convergence of the total energy with FFT box size is not expected to be variational, and indeed it is not: the FFT box technique should be viewed as a good approximation to the 'correct' result that would be obtained by using the entire simulation cell to perform FFTs for calculating the total energy. For a given FFT box size, however, the kinetic energy cut-off of our basis functions (and hence the grid spacing) *is* a variational parameter, just as in traditional plane-wave DFT.

In addition, we look at a feature common to all real space grid methods, namely the variation of the energy as the system is translated by fractions of a grid spacing with respect to the simulation cell. We define the quantity ΔE_2 to be

$$\Delta E_2(d - d_0) \equiv E^{\text{box}}(d) - E^{\text{box}}(d_0), \tag{47}$$

where d_0 is a reference position of the system, and d is some rigid displacement with respect to the grid. The top panel in Fig. 5 shows this variation, in steps of $0.1a_0$, for a lithium hydride (LiH) molecule on a cubic grid with grid spacing $0.5a_0$. The PAOs of both lithium and hydrogen had localization radii of $10.0a_0$, and were generated in the same way as described above. ΔE_2 can be seen to be of the order of 10^{-6} Hartree. The bottom panel shows the same quantity, but this time with optimization of the NGWF coefficients, $\{C_{KLM,\alpha}\}$, in addition to the elements of the density kernel [27]. We see that the variation in total energy is reduced to the 10^{-7} Hartree level. We expect the optimization of the NGWF coefficients to be important for performing accurate geometry optimizations.

7. Conclusions

We have presented a novel formalism for performing DFT calculations on a real space grid. Features of our method include a systematic basis set, non-orthogonal functions localized in spherical regions, and the use of fast Fourier transforms over small regions of the simulation cell (the FFT box) for efficient and accurate calculation of matrix elements of the Hamiltonian. The total energy may be obtained in $\mathrm{O}(N)$ operations with only a single further variational approximation, namely the truncation of the density kernel.

Acknowledgements

A.A.M. would like to thank the EPSRC for a Ph.D. studentship. C.-K.S. would like to thank the EPSRC (grant number GR/M75525) for postdoctoral research funding. P.D.H. would like to thank Magdalene College, Cambridge for a research fellowship.

Appendix A. Basis set

A.1. Definition of basis functions

The basis functions $D_{KLM}(\mathrm{r}) \equiv D(\mathrm{r} - \mathrm{r}_{KLM})$ are defined by Eq. (1).

A.2. Localization and orthogonality

Each basis function is localized on the grid, i.e. its value is unity at the grid point at which it is centered and zero on all other grid points of the simulation cell:

$$D_{KLM}(\mathbf{r}_{FGH}) = \delta_{KF}\delta_{LG}\delta_{MH}. \tag{A.1}$$

Furthermore, basis functions centered on different sites are orthogonal:

$$\int_V \mathrm{d}\mathbf{r}\, D^*_{KLM}(\mathbf{r}) D_{FGH}(\mathbf{r}) = \Omega \delta_{FK}\delta_{GL}\delta_{HM}, \tag{A.2}$$

where Ω is the volume per grid point of the standard grid.

A.3. Analytic integrals

Consider a cell periodic function, $f(\mathbf{r})$, which may be written in terms of its discrete Fourier series

$$f(\mathbf{r}) = \frac{1}{V} \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \tilde{f}(l\mathbf{B}_1 + m\mathbf{B}_2 + n\mathbf{B}_3) \mathrm{e}^{\mathrm{i}(l\mathbf{B}_1 + m\mathbf{B}_2 + n\mathbf{B}_3)}. \tag{A.3}$$

Consider also the bandwidth limited version of this same function, $f_D(\mathbf{r})$, which has only the same frequency components as $D(\mathbf{r})$:

$$f_D(\mathbf{r}) = \frac{1}{V} \sum_{l=-J_1}^{J_1+1} \sum_{m=-J_2}^{J_2+1} \sum_{n=-J_3}^{J_3+1} \tilde{f}(l\mathbf{B}_1 + m\mathbf{B}_2 + n\mathbf{B}_3) \mathrm{e}^{\mathrm{i}(l\mathbf{B}_1 + m\mathbf{B}_2 + n\mathbf{B}_3)}. \tag{A.4}$$

It can be shown that the projection of $f(\mathbf{r})$ onto a particular basis function is exactly equal to that of $f_D(\mathbf{r})$, and that furthermore, replacing the integral by a discrete sum over grid points leads to *exactly the same answer*:

$$\begin{aligned} \int_V \mathrm{d}\mathbf{r}\, f^*(\mathbf{r}) D_{KLM}(\mathbf{r}) &= \int_V \mathrm{d}\mathbf{r}\, f^*_D(\mathbf{r}) D_{KLM}(\mathbf{r}) \\ &= \Omega \sum_{X=0}^{N_1-1} \sum_{Y=0}^{N_2-1} \sum_{Z=0}^{N_3-1} f^*_D(\mathbf{r}_{FGH}) D_{KLM}(\mathbf{r}_{FGH}) = \Omega f^*_{\mathrm{D}}(\mathbf{r}_{KLM}). \end{aligned} \tag{A.5}$$

This result is very useful for our purposes as it tells us that the overlap integral of *any cell periodic function* with a function that is represented by our basis set, can be evaluated *exactly* as a summation over grid points.

A.4. Basis for the fine grid

In addition, we define a set of fine grid basis functions with twice the cut-off frequency as the $D(\mathbf{r})$:

$$B_{XYZ}(\mathbf{r}) = \frac{1}{8N_1N_2N_3} \sum_{p=-N_1+1}^{N_1} \sum_{q=-N_2+1}^{N_2} \sum_{s=-N_3+1}^{N_3} \mathrm{e}^{\mathrm{i}(p\mathbf{B}_1 + q\mathbf{B}_2 + s\mathbf{B}_3)\cdot(\mathbf{r}-\mathbf{r}_{XYZ})}, \tag{A.6}$$

that correspond to a plane-wave representation that has twice the wavevector cutoff of the standard basis functions. $\{\mathbf{r}_{XYZ}\}$ are the points of the fine grid. This basis is localized on the grid points of the fine grid and is also orthogonal:

$$B_{XYZ}(\mathbf{r}_{ABC}) = \delta_{XA}\delta_{YB}\delta_{ZC}, \tag{A.7}$$

$$\int_V \mathrm{d}\mathbf{r}\, B^*_{XYZ}(\mathbf{r}) B_{ABC}(\mathbf{r}) = \frac{\Omega}{8}\delta_{AX}\delta_{BY}\delta_{CZ}, \tag{A.8}$$

and satisfies a property analogous to Eq. (A.5):

$$\int_V \mathrm{d}\mathbf{r}\, f^*(\mathbf{r}) B_{XYZ}(\mathbf{r}) = \int_V \mathrm{d}\mathbf{r}\, f^*_B(\mathbf{r}) B_{XYZ}(\mathbf{r}) = \frac{\Omega}{8} f_B(\mathbf{r}_{XYZ}), \tag{A.9}$$

where $f_B(\mathbf{r})$ is a version of the function $f(\mathbf{r})$ that is bandwidth limited to the same frequencies as the basis functions $B(\mathbf{r})$.

References

[1] M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias, J.D. Joannopolous, Rev. Mod. Phys. 64 (1992) 1045.
[2] W. Kohn, L.J. Sham, Phys. Rev. 140 (1965) 1133.
[3] S. Goedecker, Rev. Mod. Phys. 71 (1999) 1085.
[4] G. Galli, M. Parrinello, Phys. Rev. Lett. 69 (1992) 3547.
[5] G.E. Wannier, Phys. Rev. 52 (1937) 191.
[6] W. Kohn, Phys. Rev. 115 (1959) 809.
[7] J. des Cloizeaux, Phys. Rev. 135 (1964) 685.
[8] S. Ismail-Beigi, T.A. Arias, Phys. Rev. Lett. 82 (1999) 2127.
[9] L. He, D. Vanderbilt, Phys. Rev. Lett. 86 (2001) 5341.
[10] P.W. Anderson, Phys. Rev. Lett. 21 (1968) 13.
[11] E. Hernández, M.J. Gillan, C.M. Goringe, Phys. Rev. B 53 (1996) 7147.
[12] J.-L. Fattebert, J. Bernholc, Phys. Rev. B 62 (2000) 1713.
[13] E. Hernández, M.J. Gillan, C.M. Goringe, Phys. Rev. B 55 (1997) 13 485.
[14] J.E. Pask, B.M. Klein, C.Y. Fong, P.A. Sterne, Phys. Rev. B 59 (1999) 12 352.
[15] E. Artacho, L. Miláns del Bosch, Phys. Rev. A 43 (1991) 5770.
[16] C.A. White, P. Maslen, M.S. Lee, M. Head-Gordon, Chem. Phys. Lett. 276 (1997) 133.
[17] L. Kleinman, D.M. Bylander, Phys. Rev. Lett. 48 (1982) 1425.
[18] C.-K. Skylaris, A.A. Mostofi, P.D. Haynes, C.J. Pickard, M.C. Payne, Comput. Phys. Commun. 140 (2001) 135.
[19] R.D. King-Smith, M.C. Payne, J.S. Lin, Phys. Rev. B 44 (1991) 13 063.
[20] O.F. Sankey, D.J. Niklewski, Phys. Rev. B 40 (1989) 3979.
[21] X.-P. Li, R.W. Nunes, D. Vanderbilt, Phys. Rev. B 47 (1993) 10 891.
[22] M.S. Daw, Phys. Rev. B 47 (1993) 10 895.
[23] R.W. Nunes, D. Vanderbilt, Phys. Rev. B 50 (1994) 17 611.
[24] J.M. Millam, G.E. Scuseria, J. Chem. Phys. 106 (1997) 5569.
[25] R. McWeeny, Rev. Mod. Phys. 32 (1960) 335.
[26] N. Troullier, J.L. Martins, Phys. Rev. B 43 (1991) 1993.
[27] C.-K. Skylaris, A.A. Mostofi, P.D. Haynes, O. Diéguez, M.C. Payne, (2002), submitted.