Nonorthogonal generalized Wannier function pseudopotential plane-wave method

Chris-Kriton Skylaris, Arash A. Mostofi, Peter D. Haynes,* Oswaldo Diéguez, and Mike C. Payne
Theory of Condensed Matter, Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, United Kingdom
(Received 28 September 2001; revised manuscript received 12 March 2002; published 30 July 2002)

We present a reformulation of the plane-wave pseudopotential method for insulators. This new approach allows us to perform density-functional calculations by solving directly for ''nonorthogonal generalized Wannier functions'' rather than extended Bloch states. We outline the theory on which our method is based and present test calculations on a variety of systems. Comparison of our results with a standard plane-wave code shows that they are equivalent. Apart from the usual advantages of the plane-wave approach such as the applicability to any lattice symmetry and the high accuracy, our method also benefits from the localization properties of our functions in real space. The localization of all our functions greatly facilitates the future extension of our method to linear-scaling schemes or calculations of the electric polarization of crystalline insulators.

DOI: 10.1103/PhysRevB.66.035119 PACS number(s): 71.15.Ap, 31.15.Ew

I. INTRODUCTION

The pseudopotential plane-wave method for density-functional theory (DFT) calculations has been developed and perfected over many years into a reliable tool for predicting static and dynamic properties of molecules and solids.[1] Kohn–Sham DFT maps the interacting system of electrons to a fictitious system of noninteracting particles,[2] which can be fully described by the single-particle density matrix $\rho(\mathbf{r},\mathbf{r}')$, expressed as a sum of contributions from single-particle Bloch states $\psi_{n\mathbf{k}}(\mathbf{r})$,

$$\rho(\mathbf{r},\mathbf{r}')=\sum_n f_n \frac{V}{(2\pi)^3}\int_{1\mathrm{BZ}} \psi_{n\mathbf{k}}(\mathbf{r})\psi^*_{n\mathbf{k}}(\mathbf{r}')d\mathbf{k}. \qquad (1)$$

We have assumed that we are dealing with an insulator with completely filled (occupation number $f_n=1$) or empty ($f_n=0$) states and V is the volume of the simulation cell. The **k**-point integration is carried out in the first Brillouin zone (1BZ). The single-particle states are the eigenfunctions of the Kohn–Sham Hamiltonian at each **k**-point. They are required to be orthonormal and, in general, extend over the whole simulation cell. The consequence of this orthogonality requirement is that the cost of a DFT calculation involving the $\{\psi_{n\mathbf{k}}\}$ grows cubically with the system-size. The electronic charge density is equal to the diagonal part of the density matrix multipied by a factor of 2 to take into account the spin degeneracy and is commonly abbreviated as $n(\mathbf{r})=2\rho(\mathbf{r},\mathbf{r})$.

The most general representation of the density matrix, equivalent to (1), and first applied to linear-scaling DFT calculations by Hernández and Gillan,[3] is in terms of a set of localized nonorthogonal functions $\{\phi_{\alpha\mathbf{R}}\}$,

$$\rho(\mathbf{r},\mathbf{r}')=\sum_{\alpha\beta}\sum_{\mathbf{R}} \phi_{\alpha\mathbf{R}}(\mathbf{r})K^{\alpha\beta}\phi^*_{\beta\mathbf{R}}(\mathbf{r}'), \qquad (2)$$

where the sum over **R** runs over the lattice vectors of the crystal and the matrix $K^{\alpha\beta}$ is called the density kernel, a generalization of the occupation numbers $\{f_n\}$. We will call the $\{\phi_{\alpha\mathbf{R}}\}$ ''nonorthogonal generalized Wannier functions'' (NGWFs) as they can be derived from a subspace rotation **M** between a set of Bloch orbitals at each **k**-point and a unitary transformation of the results in **k**-space (Wannier transformation),

$$\phi_{\alpha\mathbf{R}}(\mathbf{r})=\frac{V}{(2\pi)^3}\int_{1\mathrm{BZ}} e^{-i\mathbf{k}\cdot\mathbf{R}}\left[\sum_n \psi_{n\mathbf{k}}(\mathbf{r})M_{n\alpha}\right]d\mathbf{k}, \qquad (3)$$

where the number of NGWFs at each lattice cell **R** can be greater than or equal to the number of occupied Bloch bands at each **k**-point. The density kernel is the result of applying the inverse of these transformations on each side on the diagonal occupation number matrix diag($\{f_n\}$),

$$K^{\alpha\beta}=\sum_n N_n^\alpha f_n (N^\dagger)_n^\beta, \qquad (4)$$

where $\mathbf{N}=\mathbf{M}^{-1}$.

In our presentation so far, we have considered the Bloch states as the natural representation and starting point from which to construct the charge density and NGWFs. This is also the usual order which has been followed in discussions and derivations of Wannier functions[4] in the literature. Generalized Wannier functions, orthonormal or not, are most commonly constructed in a postprocessing fashion[5–8] after the end of a plane-wave band structure calculation.

Since the energy is variational with respect to the charge density, directly varying the NGWFs and the density kernel of Eq. (2) is equivalent to varying the Bloch states of Eq. (1). The localization properties of the NGWFs are set *a priori*, i.e., each NGWF is nonzero only within a predefined localization region. As a result, the computational cost of a density-functional calculation scales only quadratically with system-size rather than cubically. Furthermore, it can be made to scale linearly with one extra variational approximation: the truncation of the density kernel in Eq. (2) when the centers of the functions $\phi_{\alpha\mathbf{R}}(\mathbf{r})$ and $\phi_{\beta\mathbf{R}}(\mathbf{r})$ lie beyond some cutoff distance.[32,33]

The NGWFs are expanded in terms of a basis of periodic, bandwidth limited delta functions (Appendix A). These are centered on the points of a regular real-space grid and are related to an equivalent plane-wave basis through a unitary

transformation (the Fourier transform). Hence, the method we present is directly equivalent to the Bloch state pseudopotential plane-wave density-functional approach. The delta function basis allows us to restrict our NGWF expansions to contain only the delta functions that are included in some spherical localization region. This should be an accurate assumption for insulators in which the NGWFs decay exponentially.[9] Our approach is closely related to that of Hernández *et al.*[10] who developed a method to do calculations by optimizing both the density kernel and the functions $\{\phi_{\alpha\mathbf{R}}\}$, which they call ''support functions.''

Real-space methods, as an alternative to pure reciprocal-space plane-wave methods, have been used by many other authors for DFT calculations[11–18] in the past in order to take advantage of the benefits of localization in real-space. In particular, approaches have been developed that use functions strictly localized in spherical regions on real-space grids.[19,20] These functions play the same role as the NGWFs we present here. A different approach is taken by the SIESTA program,[21–23] which uses a basis set of numerical atomic orbitals. These are generated as described by Artacho *et al.*[24,25] and are not optimized during the calculations.

Even though these real-space methods have led to some important methodological developments, it would be very desirable if they could be directly comparable with plane-wave pseudopotential DFT. In other words, we wish to have a method that rigorously adheres to a basis set we can improve systematically, such as the plane-wave basis where its convergence towards completeness is controlled by the kinetic energy cutoff parameter. Our method achieves this by working both in real- and reciprocal-space. We demonstrate that our approach can actually be viewed as an alternative way of performing plane-wave DFT calculations which is easy to turn into a linear-scaling method in the future with only trivial modifications. It is thus also directly applicable to any Bravais lattice symmetry, in contrast to common finite difference methods that are usually restricted to orthorhombic lattices.

We should note at this point that there exist other basis sets, apart from plane-waves, that can be improved systematically: the B-splines of Hernández *et al.*,[26] and the polynomial basis in the finite-element approach of Pask *et al.*,[27,28] who have done some pioneering work using a technique for solving differential equations common in engineering applications and adapting it for electronic structure calculations.

In what follows, we begin by describing the calculation of the total energy in our scheme, directly with NGWFs, in Sec. II. In Sec. III we describe our strategy for total energy optimization, i.e., minimization of the energy with respect to both the density kernel and the NGWFs. In Sec. IV we present the FFT box technique, an essential ingredient for lowering the cost of the calculations and for eventually achieving linear-scaling behavior. In Sec. V we present tests on a variety of systems showing the accuracy and efficiency of this method and finally we conclude and mention what we see as future developments.

II. CHARGE DENSITY AND TOTAL ELECTRONIC ENERGY WITH NONORTHOGONAL GENERALIZED WANNIER FUNCTIONS

Linear-scaling DFT calculations are aimed at large systems, and in particular, large unit cells. Thus in this work we will be concerned with calculations only at the Γ-point, i.e., $\mathbf{k}=0$. This means that the Bloch bands and therefore the NGWFs can be chosen to be real. We can also drop the dependence of the NGWFs on $\mathbf{R}$, so that $\phi_{\alpha\mathbf{R}}(\mathbf{r})=\phi_\alpha(\mathbf{r})$.

Our basis set is the set of periodic bandwidth limited delta functions that are centered on the points $\mathbf{r}_{KLM}$ of a regular real-space grid,

$$D_{KLM}(\mathbf{r})=\frac{1}{N_1N_2N_3}\times\sum_{P=-J_1}^{J_1}\sum_{Q=-J_2}^{J_2}\sum_{R=-J_3}^{J_3}e^{i(P\mathbf{B}_1+Q\mathbf{B}_2+R\mathbf{B}_3)\cdot(\mathbf{r}-\mathbf{r}_{KLM})}, \tag{5}$$

where $\mathbf{B}_1$ is one of the reciprocal lattice vectors of the simulation cell. N_1 is the number of grid points in the direction of direct lattice vector $\mathbf{A}_1$, and $N_1=2J_1+1$. The delta function basis is equivalent to the plane-waves that can be represented by the real-space grid since it is related to them via a unitary transformation. An important property of the basis set is that the projection of a function $f(\mathbf{r})$ on $D_{KLM}(\mathbf{r})$ is

$$\int_V D_{KLM}(\mathbf{r})f(\mathbf{r})d\mathbf{r}=Wf_D(\mathbf{r}_{KLM}), \tag{6}$$

where W is the volume per grid point and $f_D(\mathbf{r})$ is the result of bandwidth limiting the function $f(\mathbf{r})$ to the same plane-wave components as in (5).

We represent the NGWFs in the delta function basis by

$$\phi_\alpha(\mathbf{r})=\sum_{K=0}^{N_1-1}\sum_{L=0}^{N_2-1}\sum_{M=0}^{N_3-1}C_{KLM,\alpha}D_{KLM}(\mathbf{r}), \tag{7}$$

and in the plane-wave basis by

$$\phi_\alpha(\mathbf{r})=\frac{1}{V}\sum_{P=-J_1}^{J_1}\sum_{Q=-J_2}^{J_2}\sum_{R=-J_3}^{J_3}\times\tilde{\phi}_\alpha(P\mathbf{B}_1+Q\mathbf{B}_2+R\mathbf{B}_3)e^{i(P\mathbf{B}_1+Q\mathbf{B}_2+R\mathbf{B}_3)\cdot\mathbf{r}}. \tag{8}$$

where it is straightforward to show that the amplitudes $\tilde{\phi}_\alpha(P\mathbf{B}_1+Q\mathbf{B}_2+R\mathbf{B}_3)$ are the result of a discrete Fourier transform on the delta function expansion coefficients $C_{KLM,\alpha}$.

In (7) the sum over the K, L and M indices formally goes over the grid points of a regular grid that extends over the *whole* simulation cell. From now on however, we will restrict all NGWFs to have contributions *only* from delta functions centered inside a predefined spherical region. This spherical region is in general different for each NGWF. Thus we impose on (7) the condition

$$C_{KLM,\alpha}=0 \quad \text{if} \quad \mathbf{r}_{KLM} \quad \text{does not belong to the sphere of } \phi_\alpha. \tag{9}$$

This of course does not affect the form or the applicability of Eq. (8).

The charge density of Eq. (2) with our NGWFs becomes (from now on we will use the summation convention for repeated Greek indices)

$$\begin{aligned} n(\mathbf{r}) &= 2\rho(\mathbf{r},\mathbf{r}) = 2\phi_\alpha(\mathbf{r})K^{\alpha\beta}\phi_\beta(\mathbf{r}) = 2K^{\alpha\beta}\rho_{\alpha\beta}(\mathbf{r}) \\ &= 2\sum_{X=0}^{(2N_1-1)}\sum_{Y=0}^{(2N_2-1)}\sum_{Z=0}^{(2N_3-1)} K^{\alpha\beta} \\ &\quad \times \rho_{\alpha\beta}(\mathbf{r}_{XYZ})B_{XYZ}(\mathbf{r}), \end{aligned} \tag{10}$$

which involves the fine grid delta functions $B_{XYZ}(\mathbf{r})$ that are defined in a similar way to the $D_{KLM}(\mathbf{r})$ of Eq. (5) but include up to twice the maximum wave vector of $D_{KLM}(\mathbf{r})$ in every reciprocal lattice vector direction (see also Appendix A). This is necessary because a product of two $D_{KLM}(\mathbf{r})$ delta functions is a linear combination of fine grid delta functions $B_{XYZ}(\mathbf{r})$, a result reminiscent of the Gaussian function product rule.[29]

The expressions for the various contributions to the total electronic energy with the NGWFs are simple to derive from (10). The total energy is the sum of the kinetic energy E_K, the Hartree energy E_H, the local pseudopotential energy $E_{\rm loc}$, the nonlocal pseudopotential energy $E_{\rm nl}$, and the exchange and correlation energy $E_{\rm xc}$,

$$E[n]=E_K[n]+E_H[n]+E_{\rm loc}[n]+E_{\rm nl}[n]+E_{\rm xc}[n]. \tag{11}$$

The kinetic energy is written as a trace of the product of the density kernel and of the matrix elements of the kinetic energy operator $\hat{T}=-(1/2)\nabla^2$,

$$E_K[n]=2K^{\alpha\beta}\langle\phi_\beta|\hat{T}|\phi_\alpha\rangle. \tag{12}$$

To compute these matrix elements we can apply $\hat{T}$ to the plane-wave representation (8) of $\phi_\alpha(\mathbf{r})$ and then evaluate the integral in real-space where it is equal to a discrete sum over grid points where $\hat{T}\phi_\alpha(\mathbf{r})$ obviously plays the role of $f_D(\mathbf{r})$ of Eq. (6).

Calculation of the Hartree energy requires first the Hartree potential. From Eq. (10) we see that the charge density is a fine grid delta function expansion, thus the same should be true for the Hartree potential, which is a convolution of the charge density with the Coulomb potential. Therefore, $V_H(\mathbf{r})$ can be written as a linear combination of fine grid delta functions and extends over the whole simulation cell,

$$V_H(\mathbf{r})=\sum_{X=0}^{(2N_1-1)}\sum_{Y=0}^{(2N_2-1)}\sum_{Z=0}^{(2N_3-1)} V_H(\mathbf{r}_{XYZ})B_{XYZ}(\mathbf{r}). \tag{13}$$

The Hartree energy is

$$E_H[n]=\frac{1}{2}\int V_H(\mathbf{r})n(\mathbf{r})d\mathbf{r}=K^{\alpha\beta}\langle\phi_\beta|V_H|\phi_\alpha\rangle. \tag{14}$$

This quantity can be calculated as a discrete summation on the fine grid of the product of $V_H(\mathbf{r})$ with $n(\mathbf{r})$ or equivalently as a trace of the product of the density kernel and the potential matrix elements. The local potential matrix elements are integrals that are identically equal to discrete sums on the regular grid provided of course that $(V_H\phi_\alpha)_D(\mathbf{r})$ is first put on the regular grid.

The local pseudopotential energy is calculated in an entirely analogous manner to the Hartree energy and can be represented by Eq. (14) if we put $V_{\rm loc}$ in place of V_H and multiply it by a factor of 2 to take into account the lack of self-interaction in this case.

The nonlocal pseudopotential energy $E_{\rm nl}[n]$ is the expectation value of the nonlocal potential operator $\hat{V}_{\rm nl}$ in the Kleinman–Bylander form,[30]

$$\hat{V}_{\rm nl}=\sum_A\sum_{lm(A)}\frac{|\delta\hat{V}_l^{(A)}\Psi_{lm}^{(A)}\rangle\langle\Psi_{lm}^{(A)}\delta\hat{V}_l^{(A)}|}{\langle\Psi_{lm}^{(A)}|\delta\hat{V}_l^{(A)}|\Psi_{lm}^{(A)}\rangle}, \tag{15}$$

where the A-summation runs over the atoms in the system and the lm-summation runs over the pseudo-atomic orbitals of a particular atom. The $\delta\hat{V}_l^{(A)}$ is an angular momentum dependent component of the nonlocal potential of a pseudo-Hamiltonian for a particular atom and the $\Psi_{lm}^{(A)}$ are the atomic pseudo-orbitals associated with it. In the NGWF representation the nonlocal potential energy is again expressed as a matrix trace,

$$E_{\rm nl}[n]=2K^{\alpha\beta}\langle\phi_\beta|\hat{V}_{\rm nl}|\phi_\alpha\rangle. \tag{16}$$

The $\langle\phi_\beta|\hat{V}_{\rm nl}|\phi_\alpha\rangle$ matrix elements require the calculation of overlap integrals $\langle\phi_\alpha|\delta\hat{V}_l^{(A)}\Psi_{lm}^{(A)}\rangle$ between the NGWFs and the nonlocal projectors $|\delta\hat{V}_l^{(A)}\Psi_{lm}^{(A)}\rangle$. These are simple to compute as discrete summations on the regular grid, starting from the plane-wave representation of the nonlocal projectors which is analogous to the plane-wave representation of the NGWFs in Eq. (8). These integrals need only be calculated when the sphere of function $\phi_\alpha(\mathbf{r})$ overlaps the core of atom A.

The exchange-correlation energy is obtained by approximating the exchange-correlation functional expression as a direct summation on the fine grid, which first involves the evaluation of a function $F(n(\mathbf{r}))$ whose particular form depends on our choice of exchange-correlation functional,[2]

$$E_{\rm xc}[n]=\int_V F(n(\mathbf{r}))d\mathbf{r}\simeq\frac{V}{8N_1N_2N_3}\sum_{XYZ}F(n(\mathbf{r}_{XYZ})). \tag{17}$$

This is the only approximation in integral evaluation in our method as all direct summations described up to now were exactly equal to the analytic integrals. However, in the case of the exchange-correlation energy, the exchange-correlation functionals usually contain highly nonlinear expressions that can not be represented without any aliasing even when we use the delta functions of the fine grid. The resulting errors however will be of the same nature as in conventional plane-wave codes and therefore negligible.[1]

III. TOTAL ENERGY OPTIMIZATION

The total energy is a functional of the charge density $E[n]$. From Eq. (10) we see that the charge density is expanded in fine grid delta functions where $K^{\alpha\beta}\rho_{\beta\alpha}(\mathbf{r}_{XYZ})$ are the expansion coefficients. Therefore the energy will have a variational dependence on these coefficients provided they form an *N*-representable charge density. Consequently, the energy should also have a variational dependence on the density kernel $K^{\alpha\beta}$ and the NGWF expansion coefficients $C_{KLM,\alpha}$ since the $K^{\alpha\beta}\rho_{\beta\alpha}(\mathbf{r}_{XYZ})$ are constructed from them

$$E[n]=E(\{K^{\alpha\beta}\},\{C_{KLM,\alpha}\}). \tag{18}$$

It is thus sufficient to minimize the energy with respect to $\{K^{\alpha\beta}\}$ and $\{C_{KLM,\alpha}\}$. We must however do this under two constraints. The first is that the number of electrons corresponding to the charge density

$$N_e=\int_V n(\mathbf{r})d\mathbf{r}=2K^{\alpha\beta}S_{\beta\alpha} \tag{19}$$

should remain constant. The second is that the ground state density matrix should be idempotent, or in other words the eigenfunctions of the Kohn–Sham Hamiltonian have to be orthonormal,

$$\rho(\mathbf{r},\mathbf{r}')=\int_V \rho(\mathbf{r},\mathbf{r}'')\rho(\mathbf{r}'',\mathbf{r}')d\mathbf{r}''$$

or

$$K^{\alpha\beta}=K^{\alpha\gamma}S_{\gamma\delta}K^{\delta\beta}. \tag{20}$$

We choose to carry out the total energy minimization in two nested loops, in a fashion similar to the ensemble DFT method of Marzari *et al.*[31] The density kernel will play the role of the generalized occupation numbers and the NWGFs will play the role of the orbitals. So we can reach the minimum energy in two constrained-search stages,

$$E_{\min}=\min_{\{C_{KLM,\alpha}\}} L(\{C_{KLM,\alpha}\}), \tag{21}$$

with

$$L(\{C_{KLM,\alpha}\})=\min_{\{K^{\alpha\beta}\}} E(\{K^{\alpha\beta}\},\{C_{KLM,\alpha}\}), \tag{22}$$

where the minimization with respect to the density kernel in Eq. (22) ensures that L of Eq. (21) is a function of the NGWF coefficients only. In practice in Eq. (22) we do not just minimize the energy with respect to $K^{\alpha\beta}$ but we also impose the electron number and idempotency constraints (19) and (20). There are a variety of efficient methods for achieving this available in the literature, derived from the need to perform linear-scaling calculations with a localized basis.[32–36] Any of these methods would ensure that the density kernel in (22) adapts to the current NGWFs so that it minimizes the energy within the imposed constraints.

In the present work we have used the variant of the Li, Nunes, and Vanderbilt (LNV) (Ref. 32) method that was developed by Millam and Scuseria[37] in calculations with Gaussian basis sets. We emphasize again, though, that any of the other available methods could have been used as well. For simplicity of presentation, our analysis from now on will assume that the energy of Eq. (22) is minimized without any constraints. In order to take into account the constraints, the formulas we derive will have to be modified according to the density kernel minimization method one chooses to use. This is a straightforward but tedious exercise.[38]

The minimization of Eq. (22) can be performed iteratively with the conjugate gradients method.[39] As in the simpler steepest descents method, the essential ingredient is the gradient. It is easy to show[40] that this quantity is equal to twice the matrix elements of the Kohn–Sham Hamiltonian,

$$\frac{\partial E}{\partial K^{\alpha\beta}}=2\langle\phi_\beta|\hat{H}|\phi_\alpha\rangle. \tag{23}$$

The nonorthogonality of our NGWFs has to be taken into account when computing search directions with the above gradient by transforming it to a contravariant second order tensor.[41,42]

The minimization stage of Eq. (21) is also performed iteratively with the conjugate gradients method. In this case, one can show by using the properties of the delta function basis set that the gradient is

$$\frac{\partial L}{\partial C_{KLM,\alpha}}=4WK^{\alpha\beta}(\hat{H}\phi_\beta)_D(\mathbf{r}_{KLM}), \tag{24}$$

where W is the weight associated with each grid point. Here a contravariant-to-covariant tensor correction is needed when this gradient is used to calculate the search direction during a conjugate gradient step.[43] The $(\hat{H}\phi_\beta)_D(\mathbf{r})$ functions in general contain contributions from all delta functions of the simulation cell but we wish to keep $\phi_\alpha(\mathbf{r})$ restricted to its spherical region. For this reason in every minimization step of (21) we zero all the components of (24) that correspond to delta functions outside the sphere of $\phi_\alpha(\mathbf{r})$.

When the minimization with respect to the density kernel of Eq. (22) is carried out under the electron number and idempotency constaints, Eq. (24) contains extra terms as a result of the constraints imposed in (22). These terms ensure that the electron number and idempotency constraints are automatically obeyed in (21) and as a result, the optimization with respect to the support functions can be carried out in an unconstrained fashion.

IV. THE FFT BOX TECHNIQUE

Our discussion so far has demonstrated how to localize the NGWFs in real-space in a manner that ensures that they are composed by a number of delta functions that is constant with system-size. On the other hand, each delta function is expanded in the plane-waves that can be supported by the regular grid of the simulation cell. The number of these plane-waves is proportional to the size of the system. As a result, the cost of a calculation still scales cubically with system-size as in the traditional plane-wave case.

In order to reduce the computational cost we must restrict

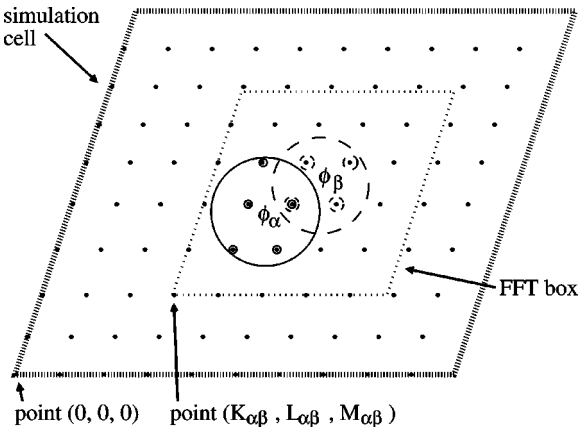

FIG. 1. The FFT box as defined for the pair of functions ϕ_α and ϕ_β.

the number of plane-waves that contribute to each delta function so that it is independent of system-size. This is not such a straightforward matter as the restriction of the NGWFs in real-space. Several factors have to be taken into account the most important of which is the Hermiticity of operators in integrals between NGWFs and also a representation of operators that is consistent when they are acting on different NGWFs. We have investigated these matters in detail in the context of the evaluation of kinetic energy integrals with NGWFs in a previous paper,[44] where we have proposed the ''FFT box'' technique as an accurate and efficient solution. It is shown there that all the imposed conditions are satisfied if the plane-waves that are used to expand the delta functions are restricted to belong to a miniature simulation cell which we call the ''FFT box.'' The FFT box must be large enough to contain any possible orientation of overlapping NGWFs, and must be a parallelepiped with a shape commensurate with the simulation cell. It should contain a regular grid which is a subset of the simulation cell grid and thus the origin of the grid of the FFT box should coincide with a particular grid point of the simulation cell. FFT techniques using smaller boxes have been used in the past to take advantage of localized functions: Pasquarello *et al.*[45,46] use them to efficiently deal with the augmentation charges that arise when using ultrasoft pseudopotentials; Hutter *et al.*[47] use them to FFT localized Gaussian orbitals. However, to the best of our knowledge, this is the first time that a small FFT box has been used to define a systematic basis set for the entire total energy calculation.

Figure 1 demonstrates in two dimensions the FFT box inside the simulation cell as defined for two overlapping NGWFs ϕ_α and ϕ_β. In the same figure the regular grid is also visible and we can observe that a portion of it is included in the FFT box. For the subset of grid points inside the FFT box we can define delta functions as we did for the simulation cell. We represent these FFT box delta functions by $d_{klm}(\mathbf{r})$ and in general we follow the convention of using *lowercase letters to denote quantities associated with the FFT box* and *uppercase letters for quantities associated with the simulation cell*.

Based on the knowledge gained through the use of the FFT box to compute kinetic energy integrals,[44] we extend here its use to the calculation of all the terms that constitute the total energy (Sec. II) and all the terms needed for its subsequent optimization with respect to the density kernel and NGWF coefficients (Sec. III). For this purpose we need to be able to express or ''project'' a function from the simulation cell to the FFT box and back. Assuming the function is expressed entirely by delta functions $D_{KLM}(\mathbf{r})$ on grid points common to the simulation cell and the FFT box, this task is straighforward and all we have to do is to rewrite the function as a linear combination of the FFT box delta functions that lie on the same centers as the simulation cell delta functions. This task is expressed formally by the $\hat{P}(\alpha\beta)$ projection operator. When this operator acts on a function in the simulation cell, it maps it onto a function in the FFT box. This is demonstrated for the function of Eq. (7) by the following expression which can also be considered as a definition of $\hat{P}(\alpha\beta)$ [a more explicit expression for $\hat{P}(\alpha\beta)$ is given in Appendix B]:

$$\begin{aligned}\hat{P}(\alpha\beta)\phi_\alpha(\mathbf{r}) &= \sum_{k=0}^{n_1-1}\sum_{l=0}^{n_2-1}\sum_{m=0}^{n_3-1} c_{klm,\alpha} d_{klm}(\mathbf{r}) \\ &= \sum_{k=0}^{n_1-1}\sum_{l=0}^{n_2-1}\sum_{m=0}^{n_3-1} C_{(k+K_{\alpha\beta})(l+L_{\alpha\beta})(m+M_{\alpha\beta})} d_{klm}(\mathbf{r})\end{aligned} \tag{25}$$

while the adjoint operator $\hat{P}^\dagger(\alpha\beta)$ can act on a function in the FFT box and turn it into a function in the simulation cell in an analogous way.

With this compact notation it is relatively straightforward to devise a way of calculating the total energy by using only the delta functions (and hence the plane-waves) periodic in the FFT box as the basis set for each NWGF. It is also equally straightforward to write formulas that represent this process in a concise way.

We start with the charge density of Eq. (2), which should of course extend over the whole simulation cell, however its contributions $\rho_{\alpha\beta}(\mathbf{r})$ from pairs of NGWFs need not, and do not, when they are calculated with the FFT box technique. Therefore the charge density is calculated by replacing these pair contributions in (10) by the following expression:

$$\rho_{\alpha\beta}^{\mathrm{box}}(\mathbf{r}) = \hat{Q}^\dagger(\alpha\beta)\{[\hat{P}(\alpha\beta)\phi_\alpha(\mathbf{r})][\hat{P}(\alpha\beta)\phi_\beta(\mathbf{r})]\}, \tag{26}$$

which involves multiplying $\phi_\alpha(\mathbf{r})$ with $\phi_\beta(\mathbf{r})$ *after* they have been transferred into the FFT box and as a consequence their product is limited to extend only over the volume of the FFT box. Here we have made use of the fine grid delta function projection operator $\hat{Q}(\alpha\beta)$ which defines a mapping between fine grid delta functions $B_{XYZ}(\mathbf{r})$ of the simulation cell and the fine grid delta functions $b_{xyz}(\mathbf{r})$ of the FFT box. It is defined in an analogous manner to $\hat{P}(\alpha\beta)$ of Eq. (25).

The matrix elements of one-electron operators, such as the kinetic energy and nonlocal potential can be easily calculated in the FFT box rather than in the simulation cell by

simply transferring the NGWFs in the FFT box before evaluating the integrals. The notation for this set of operations simply involves ''sandwiching'' the one-electon operator between the standard grid projector operator and its adjoint. The kinetic energy expression of Eq. (12) becomes

$$E_K[n]=2K^{\alpha\beta}\langle\phi_\beta|\hat{P}^\dagger(\alpha\beta)\hat{T}\hat{P}(\alpha\beta)|\phi_\beta\rangle \tag{27}$$

and a similar expression can be written for the nonlocal potential energy.

The matrix elements of the Hartree and local potentials can be calculated with the same ease provided however we have a way to transfer these potentials to the FFT box. This is indeed possible as these local potentials are expressed in terms of the fine grid delta functions and therefore we can transfer them if we use the fine grid delta function projector $\hat{Q}(\alpha\beta)$. Therefore the Hartree energy of Eq. (14) becomes

$$E_H[n]=K^{\alpha\beta}\langle\phi_\beta|\hat{P}^\dagger(\alpha\beta)[(\hat{Q}(\alpha\beta)V_H)(\mathbf{r})]\hat{P}(\alpha\beta)|\phi_\alpha\rangle. \tag{28}$$

Nothing changes in the evaluation of the exchange-correlation energy which is simply an integral over the fine grid of the whole simulation cell except of course for the fact that the charge density is now calculated by summing $\rho^{\mathrm{box}}_{\alpha\beta}(\mathbf{r})$ terms in Eq. (10) in place of the $\rho_{\alpha\beta}(\mathbf{r})$.

As far as the optimization of the energy with respect to the density kernel is concerned, the results of Sec. III are still valid provided the Hamiltonian matrix elements that constitute the density kernel gradient (23) are calculated with the FFT box method as we have described above.

In the same way, the total energy gradient with respect to the NGWF expansion coefficients of Eq. (24) can be calculated in the FFT box provided we use only the part of the Kohn-Sham Hamiltonian that exists in the FFT box. Therefore all we have to do is to substitute $\hat{H}$ of Eq. (24) by

$$\hat{H}(\alpha\beta)=\hat{P}^\dagger(\alpha\beta)[\hat{T}+\hat{V}_{\mathrm{nl}}+(\hat{Q}(\alpha\beta)V_{\mathrm{Hlxc}})(\mathbf{r})]\hat{P}(\alpha\beta), \tag{29}$$

where $V_{\mathrm{Hlxc}}(\mathbf{r})$ is the sum of the Hartree potential, local pseudopotential and exchange-correlation potential and of course now this Kohn–Sham operator is in general different for each pair of functions since it contains only the parts of the local potentials that fall inside the FFT box and therefore changes with the location of the FFT box relative to the simulation cell.

As far as implementation is concerned, there are many more issues and algorithmic details about the use of the FFT box that are beyond the scope of this paper. We describe these in another paper.[48]

V. THE NGWF PSEUDOPOTENTIAL PLANE-WAVE METHOD IN PRACTICE

We have implemented our method in a new code and we have performed extensive tests on a variety of systems. We have also performed comparisons with CASTEP,[1] an established pseudopotential plane-wave code that we use as our point of reference. We expect our approach to have efficiency comparable to the traditional cubic-scaling plane-wave pseudopotential method and thus it would be possible to use it in its place for systems with a band gap. In that case, a calculation with our method would afford a set of optimal localized functions which could be used directly in applications such as the calculation of polarization changes in crystalline solids.[49,50] However, the most important application that we envisage is the extension of the present formalism to linear-scaling calculations on very large systems. Such an extension requires the truncation of the density kernel, an issue which has been already investigated in detail.[34,36,51] The resulting linear-scaling method would be directly comparable to and have the same advantages as the plane-wave approach.

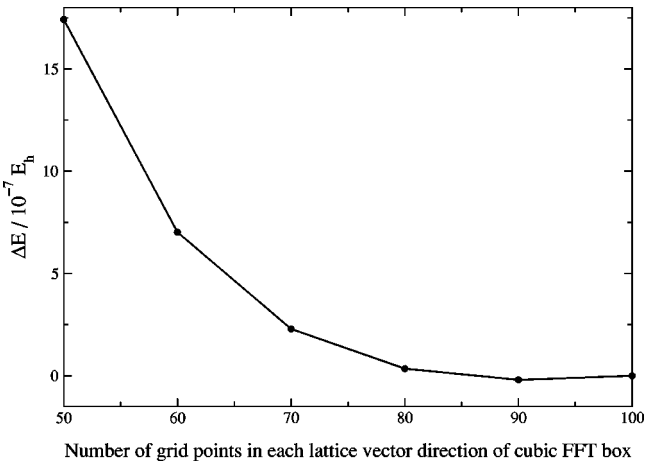

FIG. 2. ΔE plotted for a butane molecule as a function of FFT box size. All PAOs were confined to atom centered localization regions of radius $6.0a_0$, and the grid spacing was $0.5a_0$.

Since we optimize the NGWFs iteratively, some initial guesses are required for them. In this work we use pseudo-atomic orbitals (PAOs) that vanish outside a spherical region.[52] These orbitals are generated for the isolated atoms with the same radii, norm-conserving pseudopotentials and kinetic energy cutoff as in our calculations. Even though these NGWF guesses are optimal for the isolated atoms, they undergo large changes during our calculations so that in practice any guess that resembles an atomic orbital could be used, such as Slater or Gaussian functions.

We first demonstrate the accuracy of the FFT box technique as compared to using the entire simulation cell as the FFT grid. We define the quantity

$$\Delta E\equiv E^{\mathrm{box}}[n]-E[n], \tag{30}$$

where E^{box} is the total energy per atom calculated using the FFT box technique and E is that calculated using the entire simulation cell. Figure 2 shows ΔE for the butane molecule (C_4H_{10}) for different FFT box sizes. For this test we used a cubic simulation cell of side length $50a_0$ and grid spacing $0.5a_0$. The PAOs on all the atoms were confined within spherical regions of radius $6.0a_0$. The carbon atoms had one $2s$ and three $2p$ orbitals and the hydrogen atoms had a single $1s$ orbital. In this case the PAOs were not optimized during

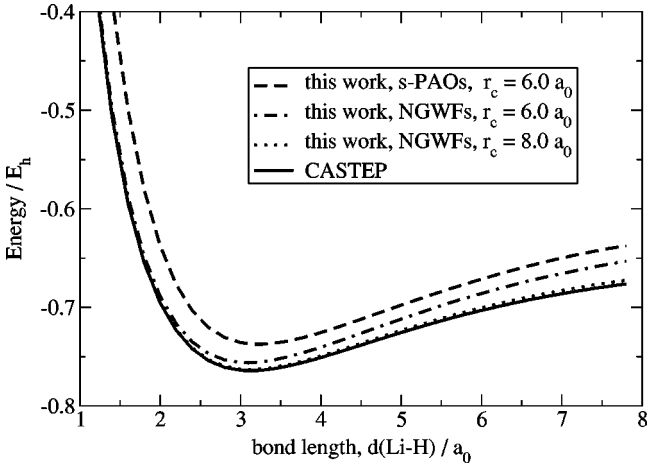

FIG. 3. Potential energy curves for LiH generated with the CASTEP plane-wave pseudopotential code and with our method for NGWF radii of $6.0a_0$ and $8.0a_0$ and for *s*-type PAOs with a radius of $6.0a_0$.

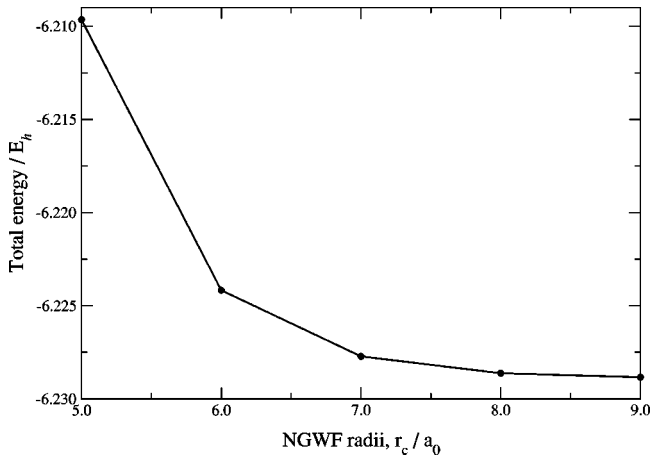

FIG. 4. Total energy of a silane molecule calculated with our method for various NGWF radii (r_c).

the calculation. It is seen that the error associated with using the FFT box rather than the entire simulation cell is only of the order of 10^{-7} E_{h} per atom, which is insignificant in the context of DFT calculations. We also note that the convergence of the total energy with FFT box size is not strictly variational, as is expected; as the FFT box size is increased, it is true that the basis set expands, but the smaller basis is not necessarily a subset of the larger one. For a given FFT box size, however, the kinetic energy cutoff of our basis functions (and hence the grid-spacing) *is* a variational parameter, just as in traditional plane-wave DFT. Further tests and discussion of the FFT box technique are published elsewhere.[48]

Our next example involves the potential energy curve of the LiH molecule inside a large cubic simulation cell of side length $40a_0$. In Fig. 3 the potential energy curve is shown as calculated by CASTEP and by our method with the same kinetic energy cutoff of 538 eV. As we have used norm-conserving Troullier–Martins[53] pseudopotentials, this is a two electron system which we describe by one NGWF on each atom. It can be seen that when we use NGWFs with radii of $8.0a_0$, we have mE_{h} agreement in total energies with the CASTEP results. Furthermore, the equilibrium bond length and vibrational frequency for this case differ from the CASTEP results by only -0.19% and 0.74%, respectively. For the smaller radius of $6.0a_0$ the curve diverges from the CASTEP curve at large bond lengths. This is because the NGWF sphere overlap, and therefore the number of delta functions between the atoms, decreases more rapidly for the small radii as the atoms are pulled apart. Also shown in the same figure is a curve that has been generated with our method but without optimization of the NGWFs, which were kept constant and equal to the initial PAO guesses. This is equivalent to a tight-binding calculation with a minimal PAO basis. As can be seen, the total energies deviate significantly from the CASTEP result, as one would expect. The equilibrium bond length for this case differs by 3.34% from CASTEP, as compared to -1.24% for the $6.0a_0$ NGWF calculation. Thus, optimizing the NGWFs improves the estimate of the bond length. The vibrational frequency obtained from the PAO case, however, differs by 3.51% from CASTEP, as compared to 5.33% for the $6.0a_0$ NGWF calculation. This, we believe, is an artifact of the localization constraint imposed on the NGWFs and suggests that in fact localization radii greater than $6.0a_0$ should be used in practice.

We now show convergence of the total energy with NGWF radius. For our tests we have used a silane (SiH_4) molecule with the same simulation cell and grid spacing as described above. A local Troullier–Martins[53] norm-conserving pseudopotential was used on the hydrogen atoms and a nonlocal one on the silicon atom. The number of NGWFs on each atom was as many as in the valence shells of the isolated atoms, i.e., one on hydrogen and four on silicon. Figure 4 shows total energy results calculated for this system as a function of NGWF sphere radii. Convergence is uniform and to mE_{h} accuracy by the time we get to a radius of $7.0a_0$. Such a NGWF radius should be adequate for practical calculations.

Here we also show that large qualitative changes occur to the shapes of the NGWFs during optimization. In Fig. 5 we show plots of isosurfaces of the NGWFs for an ethene molecule in a large simulation cell, before and after optimization. The NGWF radius was $8.0a_0$ for all atoms. In particular, the carbon $2p_x$ orbital, which is collinear with the C–C σ bond, focuses more around this bond and gains two more lobes and nodes at the positions of the hydrogen atoms farthest from its carbon center. The hydrogen functions, starting from $1s$, obtain after optimization a complicated shape that extends over the whole molecule and has nodal surfaces betwen the carbons and the rest of the hydrogens. The deep qualitative changes to the shapes of the NGWFs that occur during their optimization with our method are obviously necessary for obtaining a plane-wave equivalent result. Our optimized NGWFs in general look nothing like the atomic orbitals they started from and are adjusted to their particular molecular environment. We therefore conclude that using the delta function basis set and performing all operations consistently with the plane-wave formalism is important for obtaining the systematic convergence that plane-waves have.

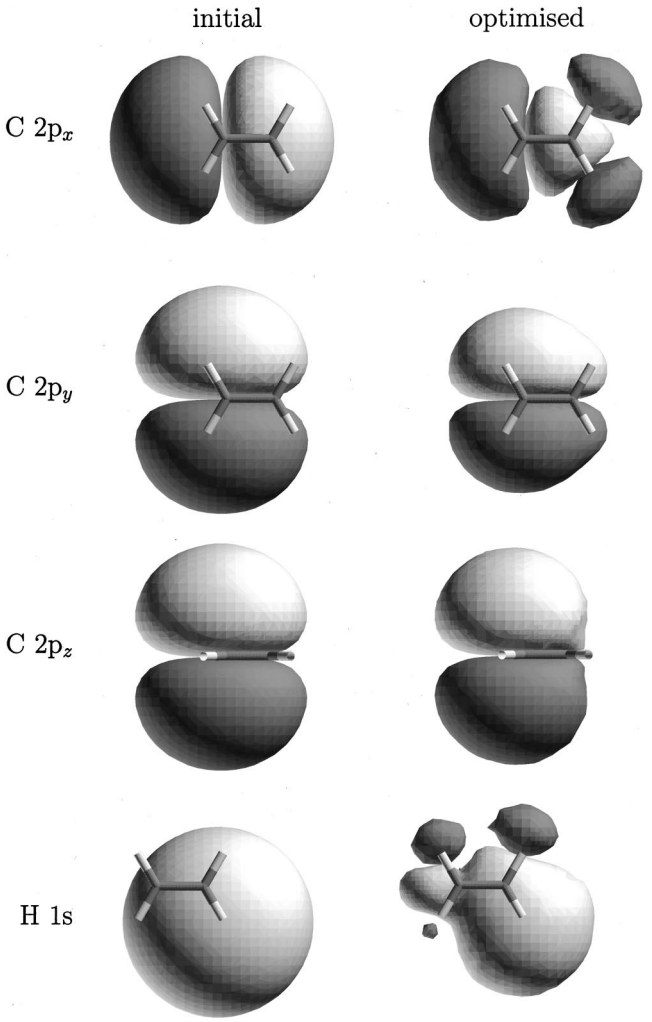

FIG. 5. Isosurfaces of NGWFs for the ethene molecule, before and after optimization. The light gray surfaces are positive and the dark gray are negative. A drawing of the ethene molecule is superimposed on each NGWF in order to show its location with respect to the atoms.

Our final example demonstrates the direct applicability of our method to any lattice symmetry without any modification. This is a consequence of being consistent throughout with the plane-wave formalism. As we have shown for the calculation of the kinetic energy in this way,[44] we also achieve better accuracy at no additional cost compared with a finite difference approach. In Fig. 6 we show a portion (Si_8H_{16}) of an infinite linear silane chain inside a hexagonal simulation cell on which we have performed a total energy calculation at a kinetic energy cutoff of 183 eV. The radii of the NGWFs were $6.0a_0$ on silicon and $5.0a_0$ on hydrogen. A total energy of -39.097 E_h was obtained when we optimized the density kernel only (with the NGWFs kept constant and equal to PAOs). When both the density kernel and the NGWFs were optimized, the energy lowered to -52.216 E_h, which is another manifestation of the fact that both the density kernel and the NGWFs should be optimized in calculations with our method.

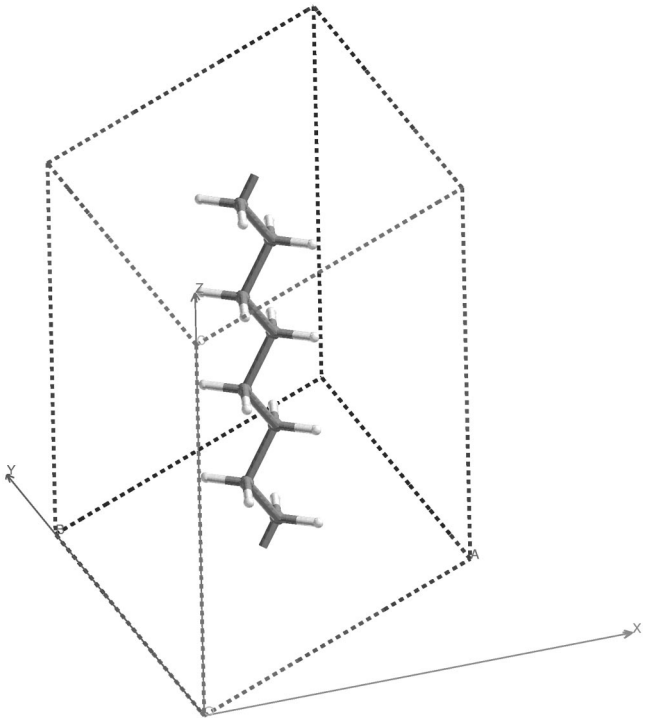

FIG. 6. A portion (Si_8H_{16}) of an infinite linear silane chain in a hexagonal simulation cell.

VI. CONCLUSIONS

We have developed a new formalism where we have recast the plane-wave pseudopotential method in terms of nonorthogonal localized functions instead of Kohn–Sham bands. A key ingredient for computationally efficient calculations with our approach is the restriction of the local functions both in real and in reciprocal space. We have written a new code to implement and test this approach. Even though it is equivalent to plane-waves, our method performs calculations directly with localized functions without ever resorting to Kohn–Sham states. As a consequence it could be more suitable for application to fields such as the theory of electric polarization of insulators. However we anticipate that the main use of this approach will be in density-functional calculations on insulators whose cost scales linearly with the size of the system. Its extension to linear-scaling calculations requires the reduction of the elements of the density kernel by truncation. Our test calculations on a variety of systems confirm that such a linear-scaling method should be directly comparable to traditional plane-waves. Advantages of our approach include high accuracy, applicability to any lattice symmetry, and systematic basis set improvement controlled by the kinetic energy cutoff.

ACKNOWLEDGMENTS

C.-K.S. would like to thank the EPSRC (Grant No. GR/M75525) for postdoctoral research funding. A.A.M. would like to thank the EPSRC for a Ph.D. studentship. P.D.H. would like to thank Magdalene College, Cambridge for a Research Fellowship. O.D. would like to thank the European

Commission for a Research Training Network Fellowship (Grant No. HPRN-CT-2000-00154).

APPENDIX A: DELTA FUNCTIONS

In this paper, whenever we refer to ''delta functions'' we will assume a periodic and bandwidth limited version of the Dirac delta functions. These delta functions are three-dimensional versions of the ''impulse functions'' that are common in signal processing applications of FFTs.[54] In an electronic structure, similar functions have been used as ''mesh delta functions'' in the ''exact finite difference method'' of Hoshi *et al.*[55] and in recent studies of their possible application when we consider the limit of an infinite simulation cell.[56]

In our derivations we will assume that we have a simulation cell of any symmetry, which in general is a parallelepiped defined by its primitive lattice vectors $\mathbf{A}_1$, $\mathbf{A}_2$, and $\mathbf{A}_3$. In this simulation cell we define a *regular grid* with an odd number of points $N_1=2J_1+1$, $N_2=2J_2+1$, and $N_3=2J_3+1$ in every direction (the adaptation of our results to the case of even numbers of points is straightforward). Therefore point $\mathbf{r}_{KLM}$ of this regular grid is defined as

$$\mathbf{r}_{KLM}=\frac{K}{N_1}\mathbf{A}_1+\frac{L}{N_2}\mathbf{A}_2+\frac{M}{N_3}\mathbf{A}_3 \tag{A1}$$

with $K=0,1,\ldots(N_1-1)$, etc.

Bandwidth limited delta functions centered at points of the regular grid are defined as

$$\begin{aligned} D_{KLM}(\mathbf{r}) &= D_{000}(\mathbf{r}-\mathbf{r}_{KLM}) \\ &= \frac{1}{N_1N_2N_3} \\ &\quad\times\sum_{P=-J_1}^{J_1}\sum_{Q=-J_2}^{J_2}\sum_{R=-J_3}^{J_3} e^{i(P\mathbf{B}_1+Q\mathbf{B}_2+R\mathbf{B}_3)\cdot(\mathbf{r}-\mathbf{r}_{KLM})}, \end{aligned} \tag{A2}$$

where $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{B}_3$ are the primitive reciprocal lattice vectors of the simulation cell. Plane-waves whose wave vector is a linear combination of these reciprocal lattice vectors have periodicity compatible with the simulation cell and therefore so do our delta functions, or any other function expanded in terms of these plane-waves. These *periodic bandwidth limited* delta functions are our basis set. A plot of a two-dimensional version of one of these delta functions is shown in Fig. 7. It is obvious from (A2) that the delta functions are real-valued everywhere in space. They are not normalized to unity but they are normalized to the grid point volume (V is the volume of the simulation cell)

$$W=\frac{V}{N_1N_2N_3}. \tag{A3}$$

Their value at grid points is equal to one when the grid point coincides with the center of the function and zero for all other grid points,

$$D_{KLM}(\mathbf{r}_{FGH})=\delta_{KF}\delta_{LG}\delta_{MH}. \tag{A4}$$

The delta functions act as Dirac delta functions with the added effect of filtering out any plane-wave components that are not part of them. For example, if $f(\mathbf{r})$ is a function periodic with the periodicity of the simulation cell but not bandwidth limited, it can be expressed in terms of its discrete Fourier transform (plane-wave) expansion,

$$\begin{aligned} f(\mathbf{r}) &= \frac{1}{V}\sum_{S=-\infty}^{\infty}\sum_{T=-\infty}^{\infty}\sum_{U=-\infty}^{\infty}\tilde{f}(S\mathbf{B}_1+T\mathbf{B}_2+U\mathbf{B}_3) \\ &\quad\times e^{i(S\mathbf{B}_1+T\mathbf{B}_2+U\mathbf{B}_3)\cdot\mathbf{r}}, \end{aligned} \tag{A5}$$

where V is the volume of the simulation cell.

It is straightforward to show that the projection of $f(\mathbf{r})$ onto $D_{KLM}(\mathbf{r})$ is

$$\begin{aligned} &\int_V D_{KLM}(\mathbf{r})f(\mathbf{r})d\mathbf{r} \\ &\quad= \frac{1}{N_1N_2N_3}\sum_{S=-J_1}^{J_1}\sum_{T=-J_2}^{J_2}\sum_{U=-J_3}^{J_3}\tilde{f}(S\mathbf{B}_1+T\mathbf{B}_2+U\mathbf{B}_3) \\ &\qquad\times e^{-i(S\mathbf{B}_1+T\mathbf{B}_2+U\mathbf{B}_3)\cdot\mathbf{r}_{KLM}} \\ &\quad= Wf_D(\mathbf{r}_{KLM}). \end{aligned}$$

We define here $f_D(\mathbf{r})$ to be the bandwidth limited version of the function $f(\mathbf{r})$, limited to the same frequency components as $D_{KLM}(\mathbf{r})$.

As the NGWFs are linear combinations of the delta functions according to (7), the result of Eq. (A6) is very important since it leads to the following relation:

$$\int_V \phi_\alpha(\mathbf{r})f(\mathbf{r})d\mathbf{r}=W\sum_{K=0}^{N_1-1}\sum_{L=0}^{N_2-1}\sum_{M=0}^{N_3-1} C_{KLM,\alpha}\, f_D(\mathbf{r}_{KLM}) \tag{A6}$$

which means that the integral in the left-hand side of the above equation is *exactly equal* to a discrete summation of values on the grid, provided we use the bandwidth limited version of $f(\mathbf{r})$.

As a corollary we observe that the delta functions are an orthogonal set since

$$\begin{aligned} \int_V D_{FGH}(\mathbf{r})D_{KLM}(\mathbf{r})d\mathbf{r} &= WD_{FGH}(\mathbf{r}_{KLM}) \\ &= W\delta_{FK}\delta_{GL}\delta_{HM}. \end{aligned} \tag{A7}$$

We also need to define the *fine grid* delta functions $B_{XYZ}(\mathbf{r})$ (here the XYZ are just grid point indices for the fine grid, they are *not related* to any Cartesian coordinates). These functions are the analogs of the delta functions we have just described that would be obtained if we doubled the minimum and maximum values that their wave vectors can take. Consequently, they have the same periodicity but they correspond to a grid with twice the number of points in every direction, i.e., $2N_1$, $2N_2$, and $2N_3$ points. They are defined by

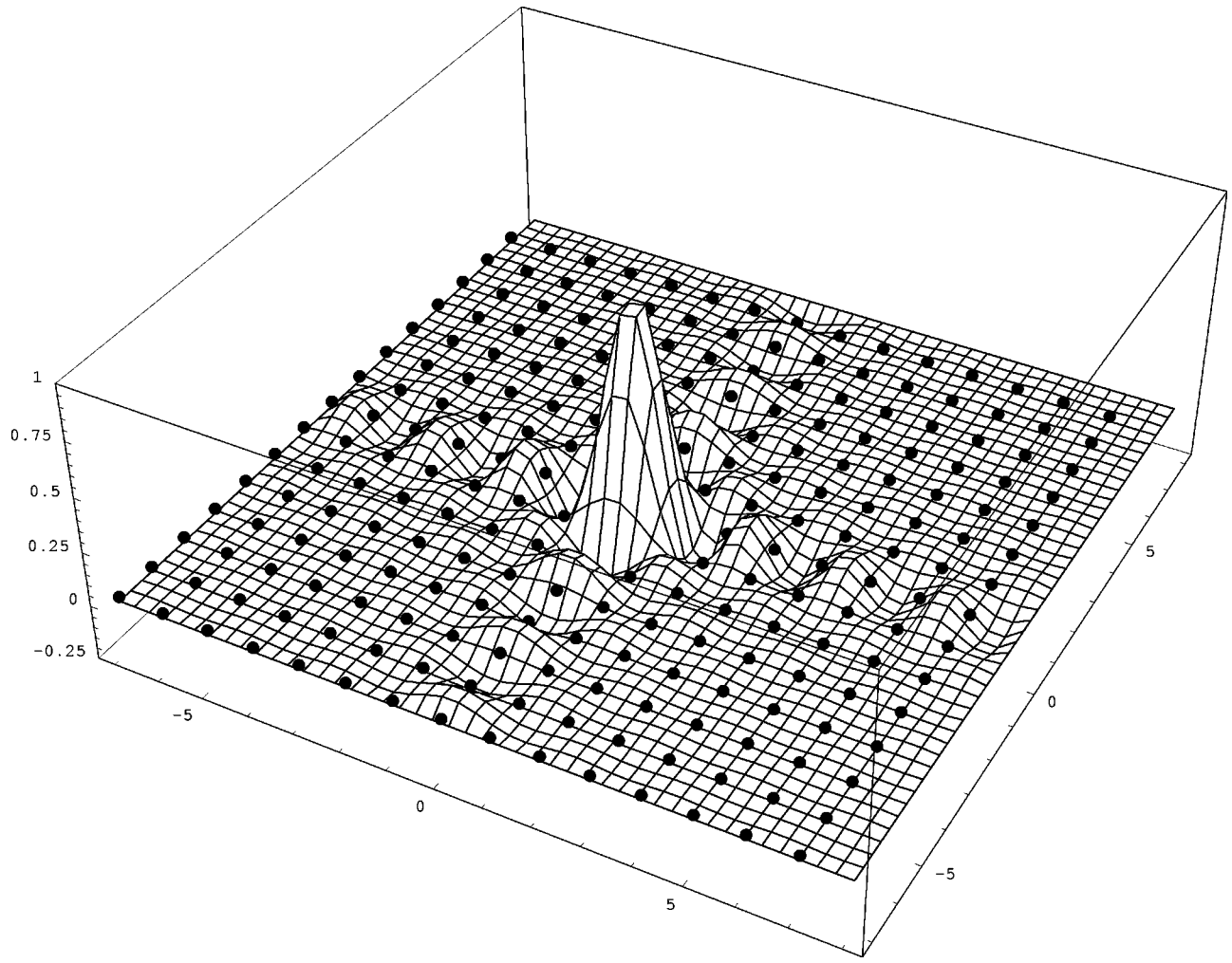

FIG. 7. A two-dimensional version of one of the functions that constitute our basis set. Here function $D_{00}(\mathbf{r})$ is shown which is identically equal to 1 at its center (point $\mathbf{r}_{00}$) and equal to zero at the centers (shown as black dots in the picture) of all other functions in the basis set.

$$B_{XYZ}(\mathbf{r}) = B_{000}(\mathbf{r}-\mathbf{r}_{XYZ}) = \frac{1}{8N_1N_2N_3} \sum_{P=(-N_1+1)}^{N_1} \sum_{Q=(-N_2+1)}^{N_2} \sum_{R=(-N_3+1)}^{N_3} \times e^{i(P\mathbf{B}_1+Q\mathbf{B}_2+R\mathbf{B}_3)\cdot(\mathbf{r}-\mathbf{r}_{XYZ})}. \quad \text{(A8)}$$

As expected, the fine grid delta functions also satisfy an equation similar to (A6),

$$\int_V B_{XYZ}(\mathbf{r}) f(\mathbf{r}) d\mathbf{r} = \frac{W}{8} f_B(\mathbf{r}_{XYZ}), \quad \text{(A9)}$$

where $f_B(\mathbf{r})$ is again a bandwidth limited version of $f(\mathbf{r})$ but this time it is limited to contain any of the plane-waves that constitute $B_{XYZ}(\mathbf{r})$ rather than $D_{KLM}(\mathbf{r})$. It is easy to verify that any function that can be written as a sum of products of pairs of delta functions can also be written as a fine grid delta function expansion. We define and use the fine grid delta functions because of this ''product rule'' property.

APPENDIX B: FROM THE SIMULATION CELL TO THE FFT BOX AND BACK

Even though the FFT box is universal in shape and size for a given system, its position with respect to the grid of the simulation cell is determined by the pair of overlapping NGWFs, say $\phi_\alpha(\mathbf{r})$ and $\phi_\beta(\mathbf{r})$, we are dealing with at any given time. An operator therefore that would map $\phi_\alpha(\mathbf{r})$ from one representation to another would depend also on the position of $\phi_\beta(\mathbf{r})$. We therefore define such an operator for the pair of functions $\phi_\alpha(\mathbf{r})$ and $\phi_\beta(\mathbf{r})$ by

$$\hat{P}(\alpha\beta) = \frac{1}{W} \sum_{k=0}^{(n_1-1)} \sum_{l=0}^{(n_2-1)} \sum_{m=0}^{(n_3-1)} |d_{klm}\rangle \times \langle D_{(k+K_{\alpha\beta})(l+L_{\alpha\beta})(m+M_{\alpha\beta})}|, \quad \text{(B1)}$$

where the numbers $K_{\alpha\beta}$, $L_{\alpha\beta}$, and $M_{\alpha\beta}$ denote the grid point of the simulation cell on which the origin of the FFT box is located. Here lowercase letters are used to represent quantities related to the FFT box, so n_1, n_2, and n_3 are the numbers of grid points in the FFT box in each lattice vector direction. Because of the periodic boundary conditions it

should also be understood that if the indices of a delta function of the simulation cell exceed the grid point indices, then this function coincides with its periodic image that falls within the simulation cell. As an example, assume $N_1=N_2=N_3=20$. Then,

$$D_{(5)(21)(23)}(\mathbf{r})=D_{(5)(1)(3)}(\mathbf{r}). \tag{B2}$$

We also need to define an operator that projects a function from the portion of the fine grid associated with functions $\phi_\alpha(\mathbf{r})$ and $\phi_\beta(\mathbf{r})$ to the FFT box. Such an operator is defined in a similar fashion to $\hat{P}(\alpha\beta)$ by

$$\hat{Q}(\alpha\beta)=\frac{8}{W}\sum_{x=0}^{(2n_1-1)}\sum_{y=0}^{(2n_2-1)}\sum_{z=0}^{(2n_3-1)}|b_{xyz}\rangle \times\langle B_{(x+2K_{\alpha\beta})(y+2L_{\alpha\beta})(z+2M_{\alpha\beta})}|. \tag{B3}$$

Operators $\hat{P}^\dagger(\alpha\beta)$ and $\hat{Q}^\dagger(\alpha\beta)$ map a function from the FFT box to the simulation cell in the standard and fine grids, respectively.

*Author to whom correspondence should be addressed. Fax: +44 (0)1223 337356; Electronic address: pdh1001@cam.ac.uk; URL: http://www.tcm.cam.ac.uk/~pdh1001/

[1] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).
[2] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, Oxford, 1989).
[3] E. Hernández and M. J. Gillan, Phys. Rev. B **51**, 10 157 (1995).
[4] E. I. Blount, Solid State Phys. **13**, 305 (1962).
[5] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12 847 (1997).
[6] G. Berghold, C. J. Mundy, A. H. Romero, J. Hutter, and M. Parrinello, Phys. Rev. B **61**, 10 040 (2000).
[7] I. Souza, R. M. Martin, N. Marzari, X. Zhao, and D. Vanderbilt, Phys. Rev. B **62**, 15 505 (2000).
[8] I. Souza, N. Marzari, and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2002).
[9] W. Kohn, Phys. Rev. **115**, 809 (1959).
[10] E. Hernández, M. J. Gillan, and C. M. Goringe, Phys. Rev. B **53**, 7147 (1996).
[11] T. L. Beck, Rev. Mod. Phys. **72**, 1041 (2000).
[12] E. L. Briggs, D. J. Sullivan, and J. Bernholc, Phys. Rev. B **52**, R5471 (1995).
[13] E. L. Briggs, D. J. Sullivan, and J. Bernholc, Phys. Rev. B **54**, 14 362 (1996).
[14] J. Bernholc, E. L. Briggs, D. J. Sullivan, C. J. Brabec, M. B. Nardelli, K. Rapcewicz, C. Roland, and M. Wensell, Int. J. Quantum Chem. **65**, 531 (1997).
[15] N. A. Modine, G. Zumbach, and E. Kaxiras, Phys. Rev. B **55**, 10 289 (1997).
[16] I.-H. Lee, Y.-H. Kim, and R. M. Martin, Phys. Rev. B **61**, 4397 (2000).
[17] J. R. Chelikowsky, N. Troullier, and Y. Saad, Phys. Rev. Lett. **72**, 1240 (1994).
[18] J. R. Chelikowsky, Y. Saad, S.Öğüt , I. Vasiliev, and A. Stathopoulos, Phys. Status Solidi B **217**, 173 (2000).
[19] J. Bernholc, E. L. Briggs, C. Bungaro, M. Buongiorno Nardelli, J.-L. Fattebert, K. Rapcewicz, C. Roland, W. G. Schmidt, and Q. Zhao, Phys. Status Solidi B **217**, 685 (2000).
[20] J.-L. Fattebert and J. Bernholc, Phys. Rev. B **62**, 1713 (2000).
[21] D. Sánchez-Portal, P. Ordejón, E. Artacho, and J. M. Soler, Int. J. Quantum Chem. **65**, 453 (1997).
[22] P. Ordejón, Phys. Status Solidi B **217**, 335 (2000).
[23] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J. Phys. Condens. Matter **14**, 2745 (2002).
[24] E. Artacho, D. Sánchez-Portal, P. Ordejón, A. García, and J. M. Soler, Phys. Status Solidi B **215**, 809 (1999).
[25] J. Junquera, O. Paz, D. Sánchez-Portal, and E. Artacho, Phys. Rev. B **64**, 235111 (2001).
[26] E. Hernández, M. J. Gillan, and C. M. Goringe, Phys. Rev. B **55**, 13 485 (1997).
[27] J. E. Pask, B. M. Klein, C. Y. Fong, and P. A. Sterne, Phys. Rev. B **59**, 12 352 (1999).
[28] J. E. Pask, B. M. Klein, P. A. Sterne, and C. Y. Fong, Comput. Phys. Commun. **135**, 1 (2001).
[29] S. F. Boys, Proc. R. Soc. London, Ser. A **200**, 542 (1950).
[30] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).
[31] N. Marzari, D. Vanderbilt, and M. C. Payne, Phys. Rev. Lett. **79**, 1337 (1997).
[32] X. P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10 891 (1993).
[33] M. S. Daw, Phys. Rev. B **47**, 10 895 (1993).
[34] P. D. Haynes and M. C. Payne, Phys. Rev. B **59**, 12 173 (1999).
[35] D. R. Bowler and M. J. Gillan, Comput. Phys. Commun. **120**, 95 (1999).
[36] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).
[37] J. M. Millam and G. E. Scuseria, J. Chem. Phys. **106**, 5569 (1997).
[38] C.-K. Skylaris, "Linear-Scaling Density Functional Calculations With Plane-Waves and Pseudopotentials" (unpublished results).
[39] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes*, 2nd ed. (Cambridge University Press, Cambridge, 1994).
[40] R. McWeeny, Rev. Mod. Phys. **32**, 335 (1960).
[41] E. Artacho and L. Miláns del Bosch, Phys. Rev. A **43**, 5770 (1991).
[42] C. A. White, P. Maslen, M. S. Lee, and M. Head-Gordon, Chem. Phys. Lett. **276**, 133 (1997).
[43] C. K. Gan, P. D. Haynes, and M. C. Payne, Comput. Phys. Commun. **134**, 33 (2001).
[44] C.-K. Skylaris, A. A. Mostofi, P. D. Haynes, C. J. Pickard, and M. C. Payne, Comput. Phys. Commun. **140**, 315 (2001).
[45] A. Pasquarello, K. Laasonen, R. Car, C. Lee, and D. Vanderbilt, Phys. Rev. Lett. **69**, 1982 (1992).
[46] K. Laasonen, A. Pasquarello, R. Car, C. Lee, and D. Vanderbilt, Phys. Rev. B **47**, 10 142 (1993).
[47] J. Hutter (private communication).
[48] A. A. Mostofi, C.-K. Skylaris, P. D. Haynes, and M. C. Payne, Comput. Phys. Commun. (to be published).
[49] R. D. King-Smith and D. Vanderbilt, Phys. Rev. B **47**, 1651 (1993).

[50] R. Resta, Rev. Mod. Phys. **66**, 899 (1994).
[51] G. Galli, Phys. Status Solidi B **217**, 231 (2000).
[52] O. F. Sankey and D. J. Niklewski, Phys. Rev. B **40**, 3979 (1989).
[53] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[54] E. O. Brigham, *The Fast Fourier Transform and its Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1988).
[55] T. Hoshi, M. Arai, and T. Fujiwara, Phys. Rev. B **52**, R5459 (1995).
[56] A. A. Mostofi, ''Analytic Kinetic Energy'' (unpublished results).