Fast diagonalisation of nonlocal pseudopotential Hamiltonians

1989 J. Phys.: Condens. Matter 1 525

(http://iopscience.iop.org/0953-8984/1/3/004)

J. Phys.: Condens. Matter 1 (1989) 525–540. Printed in the UK

Fast diagonalisation of non-local pseudopotential Hamiltonians

X Gonze†, J P Vigneron‡ and J-P Michenaud†

† Unité de Physico-Chimie et de Physique des Matériaux, Université Catholique de Louvain, Place Croix du Sud 1, B-1348 Louvain-la-Neuve, Belgium

‡ Institute for Studies in Interface Sciences, Facultés Universitaires Notre-Dame de la Paix, Rue de Bruxelles 61, B-5000 Namur, Belgium

Received 5 April 1988, in final form 26 August 1988

Abstract. The calculation of band structure and total energy of solids involves the search for the few lowest (M) eigenvectors and eigenvalues of large matrices (size $N \times N$). Standard algorithms which diagonalise the matrix entirely scale as N^3, while procedures for extracting only a subset of M eigenvalues and eigenvectors scale as MN^2. Two methods are described which aim at finding the lowest eigenvalues and corresponding eigenvectors of a non-local pseudopotential Hamiltonian. Both approaches lead to an algorithm which scales roughly as $MN^{4/3}$.

1. Introduction

The task of solving the huge quantum-mechanical many-body problem involved in the consideration of perfect or imperfect solids has been under constant development for many years. The density-functional formalism (Hohenberg and Kohn 1964, Kohn and Sham 1965) in the local density approximation has been satisfactorily used for the description of the ground-state properties of the electron gas (see, for instance, Koelling 1981, Yin and Cohen 1980, Nielsen and Martin 1985, Cohen 1986, and references therein). Today, the *ab initio* description of thermal properties of matter becomes an important issue: *ab initio* molecular dynamics, liquid state, etc. (Car and Parrinello 1985, 1988) require accurate and fast solutions of many-body Hamiltonians. The development of new, faster algorithms designed specifically to carry out the basic Hamiltonian analysis is then of paramount importance.

For periodic solids, or for other systems treated using a supercell approach, the usual self-consistent methods require the generation of a Hamiltonian matrix for a few k-points in the Brillouin zone and the search for its M lower eigenvalues and associated eigenvectors (accounting for the spin degeneracy, M is half the number of electrons per unit cell). These are needed in order to generate the self-consistent electronic charge density. The Hamiltonian is expanded in a basis set of size N, which differs according to the specific approach at hand LMTO, plane-wave basis set, etc). The factors which actually limit the system complexity (the number of atoms per unit cell for a supercell calculation) lie in the computational effort needed to generate and to diagonalise the large-size Hamiltonian matrix. These limitations are more rapidly apparent when one deals with systems exhibiting strong hard-core potentials.

0953-8984/89/030525 + 16 $02.50

Diagonalisation programs for moderate-size matrices most often use the Choleski–Householder (CH) algorithm and scale as N^3 while performing the complete matrix decomposition. For large basis sets (like those usually required in the momentum-space formalism for high-accuracy treatments) this becomes rapidly prohibitive. By contrast, some modern algorithms (Davidson 1975; RMM direct inversion in the iterative subspace (DIIS) method described by Wood and Zunger 1985) target only the first M eigenvectors and require iterative applications of an $N \times N$ matrix to a vector of length N (N^2 operations). Since the number of iterations is most often quasi-independent of N we are led to an MN^2 scaling.

For local pseudopotentials used in conjunction with the density-functional formalism, Car and Parinello (1985) pointed out that the Hamiltonian matrix could multiply a vector with less than N^2 operations: a fast Fourier transform (FFT) permits the whole operation to be carried out with no more than $N \log N$ operations. This was a key-point for undertaking molecular dynamics using *ab initio* inter-atomic potentials (Payne *et al* 1986, 1987, Needels *et al* 1987, Allan and Teter 1987, Car and Parinello 1988). In the following, we consider applying similar ideas to the diagonalisation of local Hamiltonians within the Davidson or DIIS schemes.

Moreover, the use of norm-conserving pseudopotentials (involving non-local contributions) should be preferred to obtain quantitative results, but the step towards fast use of non-local pseudopotentials does not seem to be completely straightforward. A few procedures have been designed to deal with this question (Allan and Teter 1987), but to date these procedures have needed some modification in the form of the non-local pseudopotential. We propose here two different methods to apply the non-local pseudopotential Hamiltonian of the usual form (see the work of Bachelet *et al* (1982) for the table of 94-element pseudopotentials) with a rough scaling rate of $N^{4/3}$ (we indicate in § 5 a better way to measure the method's efficiency).

Section 2 discusses in detail the background problem. Sections 3 and 4 expose the two approaches, which we refer to as 'direct projection' and 'real-space projection'. After precise comparison of the CPU times and central memory scalings (§ 5), we test the convergence and CPU performance in § 6. In the following, $\boldsymbol{R}_\tau$ systematically denotes an atom position, $\boldsymbol{k}$ is a point of the Brillouin zone, Ω the unit-cell volume, and $\boldsymbol{G}$ a vector of the reciprocal lattice. Atomic units are used.

2. Background

We first discuss the explicit form of the Hamiltonian we consider in this paper. The pseudopotential Hamiltonian (Ihm *et al* 1979) is made up of four parts:

$$H = H_{\text{kin}} + H_{\text{ps}} + H_{\text{Hart}} + H_{\text{xc}} \tag{1}$$

where H_{kin} is the kinetic operator; H_{ps} is the pseudopotential operator, which can be divided into three parts (Hamann *et al* 1979, Kleinman 1980, Bachelet and Schlüter 1982), the local, the non-local and the spin–orbit pseudopotential operators (H_{loc}, H_{nl}, H_{so}); H_{Hart} is the electrostatic Hartree operator; and H_{xc} is the exchange–correlation potential operator (taken here for facility in the local density approximation; other approximations are also tractable (Perdew 1986, Perdew and Yue 1986)).

For our purpose, we can consider three groups of terms: (i) H_{kin}; (ii) H_{loc}, H_{Hart}, H_{xc}; and (iii) H_{nl}, H_{so}. The kinetic operator is diagonal in reciprocal space. H_{loc}, H_{Hart} and H_{xc} are diagonal in direct space, and together they form a total local potential, call it

$V(r)$. H_{nl} and H_{so} are non-diagonal in each of the previous spaces, but have the special form of a sum of projectors, and act only locally around each atom. One writes for H_{nl}

$$H_{nl} = \sum_{\tau} \sum_{l} |l> V_l(\boldsymbol{r} - \boldsymbol{R}_\tau) <l| \tag{2}$$

where $V_l(\boldsymbol{r})$ has spherical symmetry and is short-range. The notation $<l|$ denotes projector on the l angular moment. For most atoms, only $l = 0, 1, 2$ need be considered. Some heavy atoms also need to include $l = 3$. For H_{so} one writes

$$H_{so} = \sum_{\tau} \sum_{l} |l> V_{so,l}(\boldsymbol{r} - \boldsymbol{R}_\tau) L \cdot S <l|. \tag{3}$$

Forgetting the slight complication introduced by the electron spin, this term is similar to the H_{nl} term and can be treated along the same lines (Hybertsen and Louie 1986, Gonze *et al* 1988). On the other hand, spin–orbit terms can be neglected for most elements.

Traditionally, the self-consistent approach to the electronic relaxation follows three steps leading to the electron density: (i) define a set of k-points in the Brillouin zone and construct the associated Hamiltonians, using the appropriate basis set; (ii) diagonalise the Hamiltonians and obtain the first M eigenvectors for each k-point; (iii) sum up the partial charge densities associated with all eigenvectors to obtain the total charge density. Using an N-plane-wave basis set ($e^{i(k+G)\cdot r}$) (Ihm *et al* 1979), with N_{at} atoms in the unit cell, one obtains the $N \times N$ matrix between G and G' for one k-point: (i) $\frac{1}{2}(k + G)^2 \delta_{GG'}$ for H_{kin} in N operations; (ii) $V(G - G')$ in N^2 operations ($V(G'')$) is the Fourier transform of $V(r)$, evaluated for $|G''| < 2G_{max}$); (iii) $H^{nl}_{k,G,G'}$ for H_{nl} in $N_{at}N^2$ operations. The last part is the computation bottleneck. The storage of the Hermitian matrix requires N^2 real locations. The diagonalisation, using the Choleski–Householder algorithm, scales as N^3 and becomes rapidly untractable for large basis sets (see § 6). Amongst the modern algorithms for which the principal step scales as MN^2, we will consider the DIIS method, although the Davidson (1975) method is equally appropriate.

In order to obtain one eigenvector using the DIIS method, N_{it} iterations must be carried out, involving mainly a matrix-by-vector multiplication (N^2 operations). As the number of iterations is practically independent of N and M (the number of eigenvectors sought), we obtain an MN^2 scaling. This step is usually the most time-consuming. The times required for the matrix generation and subsequent DIIS diagonalisations are of the same order of magnitude and both scale together (because the number of atoms and electrons are closely correlated).

In the case of local pseudopotentials, the DIIS scheme can be considerably accelerated by the use of a fast Fourier transform (FFT) technique for the matrix–vector multiplication. The key idea is the merging of the generation and diagonalisation steps. H_{kin}, being diagonal in the momentum-space representation, can be applied very fast to a vector φ in reciprocal space. For the local potential, one can use a backward FFT to transform φ in the direct space, where the potential is diagonal, apply the local potential, and return to reciprocal space by a forward FFT. If we consider a mesh of N_{FFT} points in direct space, we can apply $V(r)$ to a discrete wavefunction in N_{FFT} operations. The mesh can be refined to increase accuracy. In this way, the application of H becomes an $N \log N$ operation (instead of N^2), which makes it as efficient as previous methods for basis sets of the order 200–300 plane waves, and much more efficient for larger sets. Another advantage is that no N^2 storage is needed, since storing the real-space potential only needs approximately $64N$ real locations.

The FFT approach to the Hamiltonian matrix–vector multiplication has been intro-

duced by Car and Parinello in the context of simulated annealing molecular dynamics (Car and Parinello 1985). As discussed above, the idea is also implementable for band-structure calculations requiring very large basis sets. However, the quantitative use of pseudopotentials (norm-conserving pseudopotentials, see Hamann *et al* 1979) implies the use of a non-local part H_{nl}. Then, in order to be useful in practical applications, a fast procedure must be found for non-local Hamiltonians. This is considered in the next sections, where we discuss two different approaches, the direct projection and the real-space projection methods, both relying on the fact that $|l\rangle\langle l|$ is a projector: it transforms $\varphi_k(\boldsymbol{G})$ (or $\varphi_k(\boldsymbol{r})$) into an expansion of the form $\Sigma_{lm} f_{lm}(r) Y_{lm}(\theta, \varphi)$, where only a few l, m values are needed (up to $l = 3$, which means at most 16 terms $f_{lm}(r)$).

3. Direct projection method

For simplicity, let us consider only one atom, centred at the origin. We will use spherical coordinates θ, φ, r with their usual meaning, and denote the vectors $\boldsymbol{k} + \boldsymbol{G}$ as $\boldsymbol{K}$ and $\boldsymbol{k} + \boldsymbol{G}'$ as $\boldsymbol{K}'$.

We start from the plane-wave expansion of the crytal wavefunction,

$$\varphi_k(\boldsymbol{r}) = \sum_{G'} \varphi_k(\boldsymbol{G}')\, \mathrm{e}^{\mathrm{i}\boldsymbol{K}'\cdot\boldsymbol{r}} \tag{4}$$

and expand it in spherical waves, using the identity

$$\mathrm{e}^{\mathrm{i}\boldsymbol{K}\cdot\boldsymbol{r}} = \sum_{l} (2l + 1)\, \mathrm{i}^l j_l(Kr) P_l(\cos\chi) \tag{5}$$

where χ is the angle between $\boldsymbol{K}$ and $\boldsymbol{r}$. In this representation, the non-local matrix element $V^l_{kGG'}$ associated with the angular momentum l, expressing transitions between states of wavevectors $\boldsymbol{k} + \boldsymbol{G}$ and $\boldsymbol{k} + \boldsymbol{G}'$, is given explicitly by Ihm *et al* (1979) as

$$V^l_{kGG'} = \frac{4\pi(2l + 1)}{\Omega} P_l(\cos\gamma) \int_0^\infty V_l(r) j_l(Kr) j_l(K'r) r^2\, \mathrm{d}r. \tag{6}$$

Here $\cos\gamma = \boldsymbol{K}\cdot\boldsymbol{K}'/|\boldsymbol{K}|\,|\boldsymbol{K}'|$. If we now apply this non-local part of the Hamiltonian to some vector $\varphi(\boldsymbol{G}')$ we get

$$V^l\varphi_k(\boldsymbol{G}) = \frac{4\pi}{\Omega} \int_0^\infty V_l(r) j_l(Kr) \left(\sum_{G'} (2l + 1) \varphi_k(\boldsymbol{G}') j_l(K'r) P_l(\cos\gamma) \right) r^2\, \mathrm{d}r. \tag{7}$$

In contrast with the case of a local potential, the FFT approach is of no use here essentially because the summation over G' does not take the form of a convolution product. However, if we expand the function $j_l(K'r)$ into a Taylor series (truncated for powers above n_{max})

$$j_l(K'r) = \sum_{n=0}^{n_{max}} j_{ln} K'^n r^n \tag{8}$$

the non-local pseudopotential becomes 'separable'—in the sense recently introduced by Allan and Teter (1987). Using the notation

$$C^n_l(K) = \frac{1}{K^1} \int_0^\infty V_l(r) j_l(Kr) r^{n+2}\, \mathrm{d}r \tag{9}$$

and expanding the Legendre polynomials according to the addition theorem of spherical harmonics we are led to the final result

$$V^l\varphi_k(\boldsymbol{G}) = \frac{(4\pi)^2}{\Omega}\sum_{n=0}^{n_{\max}} C_{nl}(\boldsymbol{K})K^l j_{nl}\sum_{m=-l}^{l} Y_{lm}(\theta,\varphi)\left(\sum_{\boldsymbol{G}'}(\boldsymbol{G}')K'^n Y^*_{lm}(\theta',\varphi')\right). \tag{10}$$

The advantage of this formulation lies in the fact that the $\boldsymbol{G}'$ summation enclosed in brackets, denoted B_{nlm}, need be performed only once, for a given wavefunction, independently of the vector $\boldsymbol{G}$ considered.

The potential $V_l(r)$ is usually given either in an analytical form, or tabulated for some set of r values. The last case gives no problem of evaluation, and for a potential of the form

$$V(r) = \sum_i (\beta_i + \delta_i r^2)\exp(-\alpha_i r^2) \tag{11}$$

the result of the calculation of $C_l^n(K)$ is given in Appendix 1. Considering the algorithm construction (with more than one atom located at the origin), the direct projection method can be decomposed in three steps: (i) For each $\boldsymbol{k}$ and each atom, before starting the diagonalisation procedure, evaluate the $C_l^n(K)$ and structure factors $\exp(i\boldsymbol{K}'\cdot\boldsymbol{R}_\tau)$. The $C_l^n(K)$, as a representation of the pseudopotential, are atom-dependent, but not 'location-dependent'. For the same atoms located in different places, it need not be evaluated twice and stored. (ii) At each application of the Hamiltonian and for each atom, evaluate the B_{nlm} by summation on the $\boldsymbol{G}'$. (iii) At each application of the Hamiltonian and for each atom, generate the $V^l\varphi_k(\boldsymbol{G})$, using the B_{nlm} and $C_l^n(K)$, by summation on the Taylor indices. The two last steps scale as $N_{\text{at}}Nn_{\max}$.

The number of Taylor coefficients need not be large, because after the projection, $f_{lm}(r)$ will be multiplied by $V_l(r)$ which is short-ranged; moreover, for even (odd) l, odd (even) power coefficients vanish. Using dimensional arguments, it can be seen that if the size of the basis set which describes $\varphi(\boldsymbol{r})$ is enlarged by a factor F, the set of Taylor coefficients which describes $f(r)$ must be increased by only a factor $F^{1/3}$, in order to maintain a constant accuracy (see also Appendix 2).

4. Real-space projection method

In this method, the set of B_{nlm} will be found using a polynomial interpolation of the wavefunction in real space (obtained by a fast Fourier transform, also needed for the local part of the Hamiltonian) and, similarly, the set of $V^l\varphi_k(\boldsymbol{r})$ is determined in real space and transformed using the reverse FFT. Care must be taken to keep control of the truncation error due to real-space discretisation and interpolation procedure. This will be discussed in § 6.

The coefficients B_{lmn} are extracted from the following identity:

$$\sum_{n=0}^{n_{\max}} r^n B_{nlm} = \int Y^*_{lm}(\theta,\varphi)\varphi_k(\boldsymbol{r})\,\mathrm{d}\Omega. \tag{12}$$

In order to evaluate the wavefunction polar projections, a uniform set of Q points (part or, if necessary, a periodic extension of the FFT grid) is chosen around each atom. A three-dimensional polynomial interpolation for $\varphi_k(\boldsymbol{r})$ is then fitted to the known

wavefunction values. In terms of the FFT basis vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$, the relevant mesh points are denoted $\boldsymbol{r}_p = a\boldsymbol{e}_1 + b\boldsymbol{e}_2 + c\boldsymbol{r}_3$, the origin being attached to the nearby atomic site.

We consider the three-dimensional polynomial interpolation which will be written as $\varphi_k(\boldsymbol{r}) = \Sigma_i\Sigma_j\Sigma_k \alpha_{ijk} a^i b^j c^k$. In order to obtain a $Q^{4/3}$ scaling of this calculation step, we can treat each dimension in a recursive way:

$$\varphi_k(\boldsymbol{r}) = \varphi_k(a, b, c) = \sum_i \beta_i(b, c) a^i = \sum_i \sum_j \delta_{ij}(c) a^i b^j = \sum_i \sum_j \sum_k \alpha_{ijk} a^i b^j c^k. \tag{13}$$

For each step, the one-dimensional interpolation can easily be performed using a standard least-square procedure, or a direct inversion of the (a^i), (b^i), (c^i) matrices. The convergence of the interpolating sequences on larger and larger sets of points is found to be appropriate for all practical purposes.

From these α_{ijk}, a simple projection of the expression $a^i b^j c^k$ on the l angular momentum space provides the coefficients B_{nlm}. This calculation proceeds in such a way that the transformation coefficients, for each angular function,

$$T_{lm}^{ijk} = \int Y_{lm}(\theta, \varphi)^* a^i b^j c^k \, \mathrm{d}\Omega \tag{14}$$

are first obtained in $Qn_{\max}$ operations using a recursive approach (see Appendix 3), and are then fed into the identity

$$\sum_n B_{nlm} r^n = \sum_i \sum_j \sum_k a_{ijk} T_{lm}^{ijk} r^{i+j+k}. \tag{15}$$

This last step scales linearly with Q, as the number of angular functions is kept constant.

Using the set of B_{nlm}, the application of the non-local potential operator to the wavefunction can be performed on a restricted zone around the atom, which by convenience, is chosen to be the Q points zone:

$$\sum_l V^l \varphi_k(\boldsymbol{r}) = \sum_l \sum_{m=-l}^{l} \sum_{n=0}^{n_{\max}} B_{nlm} Y_{lm}(\theta_r, \varphi_r) V_l(r) r^n. \tag{16}$$

The remarkable fact about this procedure is that it can be split into two distinct stages, the first one being independent of the explicit representation of the wavefunction, which leads to a considerable saving in computational effort. This first stage involves (i) the choice of an appropriate set of Q points for each atom, (ii) the evaluation of the non-local potential $V_l(\boldsymbol{r})$ and of the phase factor $\exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r})$ on this mesh, (iii) the construction of transformation matrices (equation (13)) and (iv) the generation of the T_{lm}^{ijk} set. Following this, the Hamiltonian can be applied to the real-space wavefunction with minimal computational effort: (v) evaluation of the α_{ijk} using the FFT values of the function in real space, (vi) evaluation of the B_{nlm} and (vii) generation of the $V^l\varphi_k(\boldsymbol{r})$, using the B_{nlm} and $V_l(\boldsymbol{r})$, by direct summation. Step (v) scales as $N_{\mathrm{at}}Q^{4/3}$ while steps (vi) and (vii) both scale as $N_{\mathrm{at}}Qn_{\max}^{1/3}$. All other steps are significantly less demanding.

5. Theoretical evaluation of each algorithm's performance

As can be noted, CPU time and memory scaling both grow with the number of plane waves N, the number of atoms N_{at}, the number of electrons M, the number of FFT points N_{FFT}, the maximum polynomial exponent $n_{\max}$, and the number of points Q around each

atom. The k-point sampling scaling is also an important parameter, but can be considered separately, as all methods are similarly affected. To compare the five methods (Choleski–Householder (CH), DIIS), direct projection, real-space projection, and local potentials using FFT), we need some normalisation. We express the scaling of all relevant parameters as a function of only two variables: the 'volume scaling V' and the 'precision scaling P'.

By 'precision scaling', we mean an increase of the number of plane waves by a factor P, keeping constant the number of electrons, the number of atoms and the volume of the system. In the CH and RMM methods, increasing the number of plane waves leads to an accuracy improvement for the total energy and the band structure. To ensure a comparable accuracy with other methods one needs to improve N_{FFT} by a factor P, and $n_{\max}$ by a factor $P^{1/3}$ (see Appendix 2). The number of mesh points Q kept to represent the potential at each atomic site varies with the analytical form given to the pseudopotential, a Gaussian representation leading to a $P^{3/2}$ dependence (a careful discussion of this point can be found in Appendix 4).

By 'volume scaling', we mean an increase of the unit-cell volume by a factor V, keeping constant the kinetic energy cut-off, the average electronic charge density and the average 'atomic density'. For example, defining a frozen-phonon supercell needs a V scaling of the original unit cell. In such an operation N, N_{at}, M and N_{FFT} are all multiplied by a factor V, while Q and $n_{\max}$ are unchanged, as they only depend on the range of the pseudopotentials.

Figure 1 shows the effect of typical $P = 4$ and $V = 4$ scalings of a square planar lattice on real and reciprocal spaces.

We now turn to the scaling evaluation of the algorithms with respect to CPU and memory scaling. The bottleneck of both the CH and RMM schemes is the diagonalisation step. The N^3 algorithm leads to a V^3P^3 scaling in the CH approach, and the MN^2 algorithm

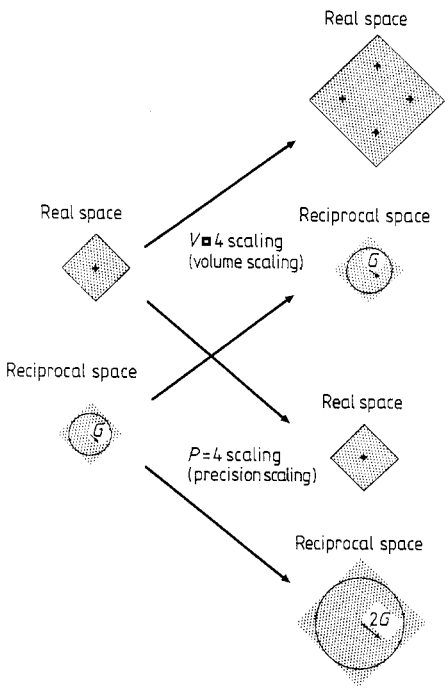

Figure 1. Representation of $V = 4$ and $P = 4$ scalings, on a square planar lattice, in both real space and reciprocal space. $V = 4$ scaling causes the increase of size of the unit cell, but without modifying the density of real-space FFT points, while in reciprocal space, the size of the zone is kept constant, and the density of G-points is increased. For $P = 4$ scaling, the effects on real and reciprocal spaces are interchanged.

leads to a V^3P^2 scaling characteristic of the RMM procedure. Both require that the full N^2 matrices be stored, which accounts for a V^2P^2 scaling. With local potentials, the combination of the DIIS approach with a fast Fourier treatment of the matrix multiplication leads to a CPU time proportional to $MN \log N$, which means a $V^2P \log(VP)$ scaling. Here, about $64N$ storage locations are needed, leading to an interesting VP scaling. The treatment of a non-local contribution to the pseudopotential involves an $MN_{at}Nn_{max}$ step, thus a $V^3P^{4/3}$ scaling for the direct projection method (respectively, an $MN_{at}Q^{4/3}$ step and a V^2P^2 scaling for the real-space projection method). The most stringent storage constraint for the direct projection method arises from the necessity of keeping available the whole set of $C_l^n(K)$ coefficients: the $Nn_{max}/2$ locations required for each atomic species and each angular momentum state gives a $V^2P^{4/3}$ scaling rule, to be compared to the $VP^{3/2}$ scaling of the memory requirement motivated by the discrete representation of real-space pseudopotential components $V_l(r)$ in the real-space projection approach. In the case of large unit cells containing the same atomic species many times (like in primitive cells associated with superstructures used for the description of superlattices, frozen phonons, or non-periodic systems approached by a supercell technique), the scaling rule in the direct projection method can be reduced to an even more favourable $VP^{4/3}$ memory scaling, as the set of $C_l^n(K)$ are identical for the same atoms. However, the volume scaling is then dominated by the V^2P memory usage needed to store the irreducible number of structure factors. Table 1 summarises the various scaling rules found through the present analysis.

The above discussion justifies the need for two distinct methods in order to account properly for the non-local character of the pseudopotentials in fast diagonalisation procedure. Basically, one can distinguish the case where a small unit cell must be treated with high accuracy (i.e. when rather hard-core pseudopotentials are present which require the use of an especially large basis set) from the case where the overall limitation comes from the use of very large unit cells. A low P-scaling is needed for the first class of problems, while a low V-scaling is beneficial to the second. From this argument, it is clear that the direct projection method permits better speed and lower memory requirements for the high-accuracy treatment of small-unit-cell crystal structures, while the real-space method turns out to be the more advantageous for large-volume unit cells. The gain in this case is due to the localised character of the non-local part of the pseudopotentials.

This correspondence between direct projection method, suitable for high-accuracy applications, and real-space projection method, suitable for high-volume applications, also exists for other forms of non-local pseudopotentials, and is then somewhat simpler (e.g. the Kleinmann–Bylander pseudopotential, used by Allan and Teter (1987)). Work is in progress in this direction, and will be the subject of a future article.

Table 1. Comparison of leading CPU and memory scaling, for the different methods considered in this paper. For the direct method, with only a few different atomic species per unit cell, the leading central memory scaling is $VP^{4/3} + V^2P$ (see text).

	CH	DIIS	Direct	Real space	Local potentials
CPU time	V^3P^3	V^3P^2	$V^3P^{4/3}$	V^2P^2	$V^2P \log VP$
Central memory	V^2P^2	V^2P^2	$V^2P^{4/3}$	$VP^{4/3}$	VP

6. Tests

Tests have been carried out for the well known silicon band structure, total energy and lattice geometry. Many studies, including those of Holzschuh (1983) have extensively discussed the influence of cut-off, pseudopotential and exchange–correlation potential on the calculated ground-state properties.

We used the momentum-space formalism of Ihm *et al* (1979), with Ceperley–Alder form of exchange–correlation functional (Ceperley and Alder 1980), as parametrised by Perdew and Zunger (1981), and an ionic pseudopotential taken from the table published by Bachelet *et al* (1982). Integration in the Brillouin zone was performed using a 10-special-point set (Monkhorst and Pack 1976). The self-consistency convergence was considered to be reached when the total energy improvement was found to be smaller than 10^{-6} Ryd/atom. The experimental lattice parameter $a_0 = 5.43$ Å was used for the total energy and band-width calculations.

The use of $l = 0, 1, 2$ non-local potentials involves the estimation of nine sets of coefficients B_{nlm} needed to apply both projection methods. Instead of using the local 'core potential' provided by the Bachelet tables, we found it more advantageous to use the $l = 2$ potential as the local part (we call it d potential), which leaves only four sets of B_{nlm} to be determined. Table 2 shows a comparison between the results obtained when using d and core potentials as local contributions. The differences have no significance, in view of the convergence reached.

The first set of tests concern the direct projection method. Table 3 describes a convergence test which involves an increasing number n_{max} of terms retained in the Taylor expansion of the wavefunction. The convergence shown when increasing n_{max} is found to be appropriate, and, as expected, the quasi-doubling of the plane-wave set size only requires a moderate change of n_{max}.

Figure 2 compares the typical CPU time needed for one diagonalisation. CH and RMM methods have N^3 and N^2 dependence, while the direct projection method (with both d or core potential) gives a stepped increase associated with the use of a variable-size fast Fourier transform (see Appendix 5).

The cross-over between RMM and direct projection methods happens at about 400

Table 2. Convergence of total energy, band width and equilibrium lattice constant of silicon, upon increase of the size N of the plane-wave basis set. The d and core potential results appear to agree within the uncertainty related to the convergence actually reached. Experimental values from Yin and Cohen (1980) and Hybertsen and Louie (1985), where other LDA values can also be found.

	Total energy (Ryd/atom)		Valence band width (eV)		Lattice constant (Å)	
N	d	core	d	core	d	core
59	−7.78387	−7.78397	11.19884	11.19277	5.1279	5.1292
169	−7.91260	−7.91249	11.86748	11.86743	5.2674	5.2672
229	−7.92908	−7.92895	11.91840	11.91823	5.3108	5.3108
307	−7.93827	−7.93816	11.93064	11.93061	5.3474	5.3476
411	−7.94111	−7.94101	11.92261	11.92253	5.3634	5.3637
531	−7.94212	−7.94202	11.91048	11.91045	5.3653	5.3656
Expt	−7.919		12.5 ± 0.6		5.43	

Table 3. Convergence of total energy and valance band width of silicon with respect to n_{max}, the number of terms retained in the Taylor expansion of the wavefunction. The cases of d and core potentials are considered, for 229 and 411 plane waves. Only differences with respect to the converged value are given, without sign.

		229 plane waves		411 plane waves	
	n_{max}	d	core	d	core
Energy	20	3×10^{-4}	1×10^{-4}	9×10^{-3}	1×10^{-3}
(Ryd/atom)	24	6×10^{-6}	2×10^{-5}	1×10^{-3}	5×10^{-4}
	28	2×10^{-6}	2×10^{-6}	1×10^{-4}	9×10^{-5}
	32	$<10^{-6}$	$<10^{-6}$	1×10^{-5}	1×10^{-5}
	36			1×10^{-6}	1×10^{-6}
	40			$<10^{-6}$	$<10^{-6}$
	Converged	−7.929081	−7.928951	−7.941115	−7.941015
Band width	20	2×10^{-3}	1×10^{-3}	1×10^{-2}	1×10^{-2}
(eV)	24	5×10^{-5}	2×10^{-4}	4×10^{-4}	8×10^{-4}
	28	$<2 \times 10^{-5}$	$<2 \times 10^{-5}$	8×10^{-5}	3×10^{-5}
	32			2×10^{-5}	$<2 \times 10^{-5}$
	Converged	11.91840	11.91823	11.92261	11.92253

plane waves. As already remarked by Nielsen and Martin (1985), about 500 plane waves are needed in order to produce well converged values of the lattice parameter and elastic constants of silicon. For other materials, in particular those involving Mendeleev first-row elements or for transition metals, a much larger number of plane waves could be necessary (Wentzcovitch and Cohen 1986).

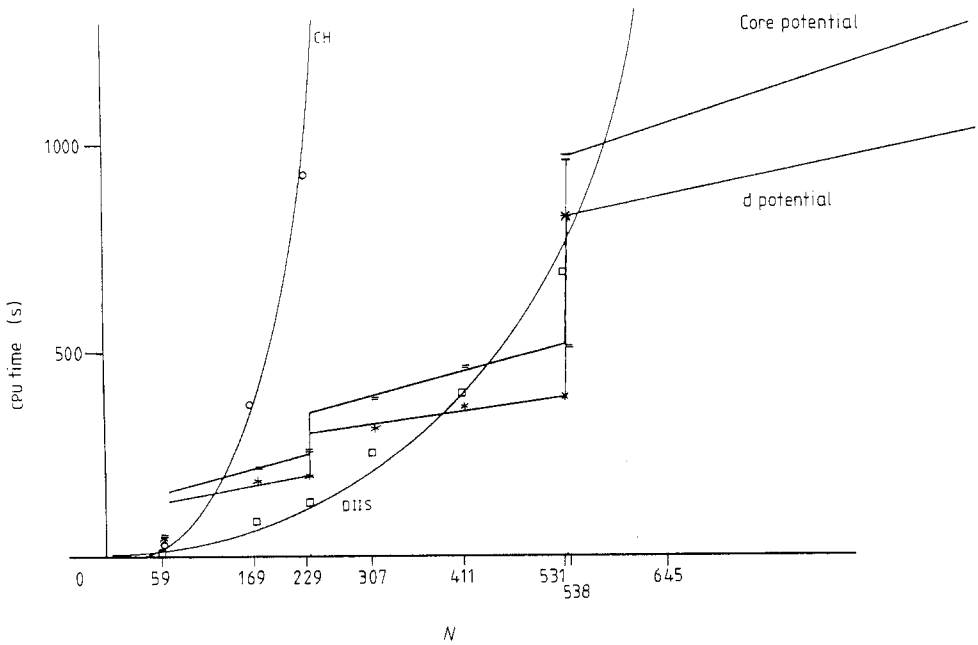

Figure 2. Representation of the CPU time (average) dependence on N, using the different methods CH, DIIS and direct space projection, where for the last, both d potential and core potential have been used. The test has been performed using an IBM 4381.

Table 4. Convergence of total energy and valence band width of silicon with respect to Q, the number of real-space points on which the d non-local potential acts, for some values of N and N_{FFT}. Only differences with respect to the converged value are given, without sign.

N	N_{FFT}	$Q^{1/3}$	ΔE (Ryd/atom)	Band width (eV)
229	3456	11	1×10^{-3}	9×10^{-4}
		13	7×10^{-5}	3×10^{-5}
		15	1×10^{-6}	$<2 \times 10^{-5}$
		17	$<10^{-6}$	
		Converged	−7.928950	11.91856
229	8192	15	1×10^{-3}	7×10^{-4}
		17	1×10^{-4}	2×10^{-5}
		19	1×10^{-5}	$<2 \times 10^{-5}$
		21	1×10^{-6}	
		23	$<10^{-6}$	
		Converged	−7.929079	11.91842
411	8192	15	1×10^{-3}	7×10^{-4}
		17	1×10^{-4}	3×10^{-5}
		19	9×10^{-6}	$<2 \times 10^{-5}$
		21	1×10^{-6}	
		23	$<10^{-6}$	
		Converged	−7.941115	11.92264

The application of the second approach (the real-space projection method) calls for some specific remarks. This method requires two more cut-off parameters: the step size in the FFT grid, and the extent of the non-local pseudopotential around each atom. Table 4 describes the typical convergence behaviour of the procedure with respect to the number Q of grid points used in the pseudopotential discretisation.

It is instructive to notice that, with the chosen pseudopotentials, the pseudopotential range must be extended beyond the first nearest-neighbour distance. Furthermore, reaching a total energy accuracy below the very stringent level of 10^{-6} Ryd/atom requires a rather fine FFT grid. The V- and P-scaling laws have in this case large proportionality coefficients, and this weakens the interest of the method for small systems. This could be improved by using another form for the pseudopotential (no more the Gaussian fitting of the Bachelet table). For a two-atom cell with 229 plane waves, the ratio of CPU time for the real projection method versus DIIS reaches about 5. Our approximate scaling arguments indicate that the real-space projection method should become competitive when more than about 10 atoms are involved.

7. Conclusions

We have proposed two different methods which aim at solving the problem of the fast diagonalisation of Hamiltonians involving a non-local electron–ion interaction term. Those provide an adequate basis for studying the ground-state properties of simple crystal structures where hard-core pseudopotentials are not easily avoided, or complex crystalline materials with large unit cells. The direct projection method is especially well suited to the treatment of non-local hard-core potentials, while the real-space projection method applies efficiently to large and complex unit cells.

We have not yet mentioned the possibility of combining those methods with other

theoretical tools sensitive to basis-set convergence. One may think of Hellmann–Feynman force calculations (Ihm *et al* 1979, Yin and Cohen 1982), the stress theorem (Nielsen and Martin 1985), or the linear perturbations treatment (Baroni *et al* 1987). The application of the present ideas to such theoretical treatments seems rather straightforward.

Acknowledgments

One of us (XG) has benefited from a 'Mandat d'aspirant' of the National Belgian Foundation for Scientific Research (FNRS) for this work. We also acknowledge the use of the Namur Scientific Computing Facility (Namur-SCF), a common project between the FNRS, IBM Belgium and the Facultés Universitaires Notre Dame de la Paix (FNDP).

Appendix 1

The integrals to be computed are

$$I_{l,n}(\alpha, K) = \int_0^\infty \exp(-\alpha r^2) r^{n+2} j_l(Kr)\, \mathrm{d}r$$

with n even if l even, n odd if l odd, and $n \geqslant l$. Using spherical Bessel function recurrence, we obtain

$$I_{l,n}(\alpha, K) = K^l \left(-\frac{1}{K}\frac{\mathrm{d}}{\mathrm{d}K}\right)^l I_{0,n-l}(\alpha, K).$$

We now compute $I_{0,m}(\alpha, K)$ (m even):

$$\begin{aligned} I_{0,m}(\alpha, K) &= \int_0^\infty \exp(-\alpha r^2) r^{m+2} [\sin(Kr)/Kr]\, \mathrm{d}r \\ &= -\frac{1}{K}\frac{\mathrm{d}}{\mathrm{d}K}\left(\frac{\mathrm{d}}{\mathrm{d}K}\right)^m (-1)^{m/2} \int_0^\infty \exp(-\alpha r^2)\cos(Kr)\, \mathrm{d}r \\ &= -\frac{1}{K}\frac{\mathrm{d}}{\mathrm{d}K}\left(\frac{\mathrm{d}}{\mathrm{d}K}\right)^m (-1)^{m/2} \frac{1}{2}\left(\frac{\pi}{\alpha}\right)^{1/2} \exp\left(\frac{-K^2}{4\alpha}\right) \\ &= \frac{1}{2}\left(\frac{\pi}{\alpha}\right)^{1/2} \frac{1}{K} \frac{(-1)^{m/2}}{(2\alpha^{1/2})^{m+1}} \exp\left(\frac{-K^2}{4\alpha}\right) H_{m+1}(x) \end{aligned}$$

where $x = K/2\alpha^{1/2}$, and H_m denotes the Hermite polynomial of order m. Using the recurrence properties of Hermite polynomials we finally obtain

$$I_{l,n}(\alpha, K) = \frac{1}{2}\left(\frac{\pi}{\alpha}\right)^{1/2} \frac{(-1)^{(n-1)/2}}{(2\alpha^{1/2})^{n+2}} \exp\left(\frac{-K^2}{4\alpha}\right) G_{l,n}(x)$$

with

$$G_{0,n}(x) = (1/x) H_{n+1}(x)$$

$$G_{1,n}(x) = (1/x)[H_{n+1}(x) + (1/x) H_n(x)]$$

$$G_{2,n}(x) = (1/x)[H_{n+1}(x) + (3/x)H_n(x) + (3/x^2)H_{n-1}(x)]$$
$$G_{3,n}(x) = (1/x)[H_{n+1}(x) + (6/x)H_n(x) + (15/x^2)H_{n-1}(x) + (15/x^3)H_{n-2}(x)].$$

Appendix 2

We will show in this appendix that a precision scaling of factor F (increase of the maximum norm for plane waves $G_{\max}$ by a factor $F^{1/3}$) needs only a $F^{1/3}$ scaling of $n_{\max}$. If we multiply G by a factor α in the expression $j_l(Gr)$, we affect the various terms in its r Taylor expansion differently. In order to correct for this change and keep a constant accuracy, we need to collect more terms in this expansion. We search the maximum coefficient α, by which we may multiply G in the Taylor development of $j_l(Gr)$, if we consider not only the $n_{\max}$ first terms, but the $m = F^{1/3}n_{\max}$ first terms.

We will restrict ourselves, without loss of generality, to the $l = 0$ angular momentum, and to a typical term of the form

$$T(n) = (Gr)^n/(n+1)!.$$

$n_{\max}$ is considerably larger than 1 and, as we assume that the development was well converged when restricted to the $n_{\max}$ first terms, we have $T(n_{\max}) = e^{-\beta} \ll 1$. Consider then the function

$$S(\alpha, n) = (\alpha Gr)^n/(n+1)!.$$

We must show that $S(\alpha, m) < S(1, n_{\max})$, with α less than $F^{1/3}$. We use the Stirling formula and obtain

$$\log S(\alpha, m) = m(\log \alpha Gr + 1) - m \log m$$

neglecting irrelevant supplementary terms. After some simple algebra, the inequality becomes

$$\log \alpha - \log(m/n_{\max}) - (1/n_{\max} - 1/m)\beta < 0.$$

As $\beta > 0$ and $m > n_{\max}$, we may certainly increase α as $F^{1/3}$, and $G_{\max}$ as $n_{\max}$, without loss of precision.

Appendix 3

We have defined the coordinates a, b, c with respect to a non-orthonormal basis $\boldsymbol{r}_p = a\boldsymbol{e}_1 + b\boldsymbol{e}_2 + c\boldsymbol{e}_3$. The relations with the orthonormal basis $\boldsymbol{e}_x$, $\boldsymbol{e}_y$, $\boldsymbol{e}_z$ for which $\boldsymbol{r}_p = x\boldsymbol{e}_x + y\boldsymbol{e}_y + z\boldsymbol{e}_z$ are the following:

$$\boldsymbol{e}_1 = e_{1x}\boldsymbol{e}_x + e_{1y}\boldsymbol{e}_y + e_{1z}\boldsymbol{e}_z \qquad \boldsymbol{e}_x = e_{x1}\boldsymbol{e}_1 + e_{x2}\boldsymbol{e}_2 + e_{x2}\boldsymbol{e}_3$$
$$\boldsymbol{e}_2 = e_{2x}\boldsymbol{e}_x + e_{2y}\boldsymbol{e}_y + e_{2z}\boldsymbol{e}_z \quad \text{or} \quad \boldsymbol{e}_y = e_{y1}\boldsymbol{e}_1 + e_{y2}\boldsymbol{e}_2 + e_{y3}\boldsymbol{e}_3$$
$$\boldsymbol{e}_3 = e_{3x}\boldsymbol{e}_x + e_{3y}\boldsymbol{e}_y + e_{3z}\boldsymbol{e}_z \qquad \boldsymbol{e}_z = e_{z1}\boldsymbol{e}_1 + e_{z2}\boldsymbol{e}_2 + e_{z3}\boldsymbol{e}_3.$$

We will now show the method to construct recurrence relations needed to evaluate T^{ijk}_{lm}. We only detail the special case $l = 0$. For the other angular momenta, the construction will be similar, after some differentiation with respect to a, b and/or c.

We define first

$$X^{ijk} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp[-(x^2 + y^2 + z^2)]a^i b^j c^k \, dx \, dy \, dz$$

which can be related to T_{00}^{ijk}:

$$X^{ijk} = \int_{4\pi} a^i b^j c^k \, d\Omega \int_0^{\infty} \exp(-r^2) r^{i+j+k+2} \, dr = T_{00\frac{1}{2}}^{ijk} \Gamma\left(\frac{i+j+k+3}{2}\right)$$

where the angular integration is performed on the radius-1 sphere. Γ is the usual gamma function.

X^{ijk} can be integrated by parts on a (or b or c):

$$\begin{aligned} X^{ijk} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp[-(x^2 + y^2 + z^2)]a^{i-1} b^j c^k \, (x e_{x1} + y e_{y1} + z e_{z1}) \, dx \, dy \, dz \\ &= e_{x1} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} -\tfrac{1}{2} \exp[-(x^2 + y^2 + z^2)] \frac{d}{dx}(a^{i-1} b^j c^k) \, dx \, dy \, dz \\ &\quad + e_{y1} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} -\tfrac{1}{2} \exp[-(x^2 + y^2 + z^2)] \frac{d}{dy}(a^i b^{j-1} c^k) \, dx \, dy \, dz \\ &\quad + e_{z1} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} -\tfrac{1}{2} \exp[-(x^2 + y^2 + z^2)] \frac{d}{dz}(a^i b^j c^{k-1}) \, dx \, dy \, dz \end{aligned}$$

giving

$$\begin{aligned} X^{ijk} &= e_{x1}[e_{x1}(i-1)X^{(i-2)jk} + e_{y1} j X^{(i-1)(j-1)k} + e_{z1} k X^{(i-1)j(k-1)}] \\ &\quad + e_{y1}[e_{x1} i X^{(i-1)(j-1)k} + e_{y1}(j-1)X^{i(j-2)k} + e_{z1} k X^{i(j-1)(k-1)}] \\ &\quad + e_{z1}[e_{x1} i X^{(i-1)j(k-1)} + e_{y1} j X^{i(j-1)(k-1)} + e_{z1}(k-1)X^{ij(k-2)}]. \end{aligned}$$

For cubic systems, some e-coefficients vanish.

Appendix 4

If the non-local pseudopotential is short-ranged in real space ($V_l(r) = (0)$ for $r > r_{max}$), the number of FFT points included in this zone scales with P, and we obtain the simple P-scaling of Q. This is not true for another pseudopotential form. It is easy to show that a Gaussian pseudopotential (or a combination of Gaussians), $V_l(r) = \alpha \exp(-\beta r^2)$, for example, leads to $P^{3/2}$ scaling. As n_{max} increases as $P^{1/3}$, the maximum of the function $r^n V_l(r) = r^n \alpha \exp(-\beta r^2)$, situated in $r = (n_{max}/2\beta)^{1/2}$, also moves, and enlarges the zone in which the pseudopotential action has some importance. The volume of this zone scales as $P^{1/2}$, leading to a $P^{3/2}$ scaling in conjunction with the FFT density of point scaling.

Appendix 5. Fourier transform

We shall refer to the book of Brigham (1974), where extensive discussion of Fourier transform as well as fast Fourier transform algorithms can be found. In this appendix,

only some technical points relevant to the application of the FFT algorithm to our problem will be mentioned.

The three-dimensional FFT can be reduced to a nested one-dimensional FFT, using the following recursive scheme:

$$Tf_{l_1l_2l_3} = \sum_{k_1=0}^{n-1} \exp\left(\mathrm{i}2\pi \frac{k_1 l_1}{n}\right)\left\{\sum_{k_2=0}^{n-1} \exp\left(\mathrm{i}2\pi \frac{k_2 l_2}{n}\right)\left[\sum_{k_3=0}^{n-1} \exp\left(\mathrm{i}2\pi \frac{k_3 l_3}{n}\right) f_{k_1k_2k_3}\right]\right\}.$$

The numer of points could be different for each dimension, but in the case of crystals with cubic symmetry, it is advisable to use a cubic discretisation mesh. The G vectors set, which is generally obtained from a spherical cut-off kinetic energy, must be contained in this cubic grid of FFT points. Furthermore, as a consequence of the Nyquist sampling theorem (Brigham 1974, p 83), the FFT has to be computed with a set of points of double density in each of the dimensions. Thus, the number of FFT points is at least about 16 times the number of G vectors.

The second point concerns the size of each one-dimensional FFT. If we consider only radix-2 FFT, we obtain three-dimensional grids with 8, 64, 512, 4096, 32768, . . . points. As the step-like behaviour of those grids leads to a very weak flexibility in the choice of the step size, it is recommended to allow for the use of radix of the form 3×2^n. This can be done by using the following decomposition:

$$Tf_l = \sum_{k'=0}^{n/3-2} \exp\left(\mathrm{i}2\pi \frac{k'l}{(n/3)}\right) f_{3k'} + \exp\left(\mathrm{i}2\pi \frac{l}{n}\right) \sum_{k''=0}^{n/3-1} \exp\left(\mathrm{i}2\pi \frac{k''l}{(n/3)}\right) f_{3k''+1}$$
$$+ \exp\left(\mathrm{i}2\pi \frac{2l}{n}\right) \sum_{k'''=0}^{n/3} \exp\left(\mathrm{i}2\pi \frac{k'''l}{(n/3)}\right) f_{3k'''+1}$$

which leads to three radix-2 FFT. To handle three-dimensional FFT on FCC lattices efficiently, we treat separately odd and even coordinate G points, and generate the real-space functions in the parallelepipedic volume represented in figure A1.

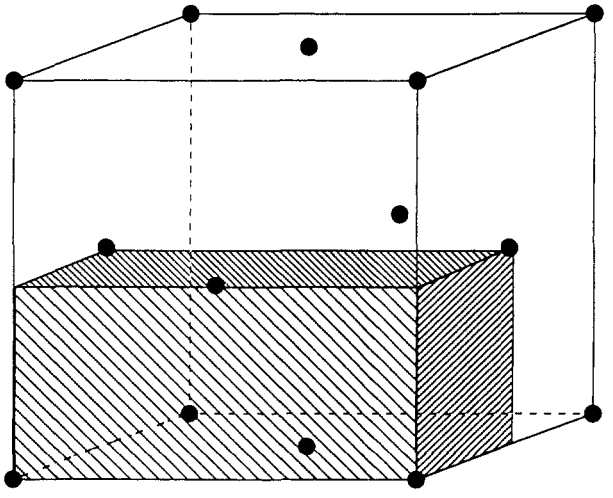

Figure A1. The real-space volume for FCC lattice on which the fast Fourier transform is performed.

Applying those considerations to the FFC diamond lattice, we obtain the natural grids:

$2 \times 8^3 = 1024$ points, for 59 (or less) G vectors sets
$2 \times 12^3 = 3456$ points, for 229 (or less) G vectors sets
$2 \times 16^3 = 8192$ points, for 531 (or less) G vectors sets
$2 \times 24^3 = 27648$ points, for about 1700 (or less) G vectors sets.

Note added in proof. Shortly after the submission of this paper, a complete discussion of the local potential FFT method (Martins and Cohen 1988) was published.

References

Allan D C and Teter M P 1987 *Phys. Rev. Lett.* **59** 1136
Bachelet G B, Hamann D R and Schlüter M 1982 *Phys. Rev.* B **26** 4199
Bachelet G B and Schlüter M 1982 *Phys. Rev.* B **25** 2103
Baroni S, Giannozzi P and Testa A 1987 *Phys. Rev. Lett.* **59** 2662
Brigham E O 1974 *The Fast Fourier Transform* (Englewood Cliffs, NJ: Prentice-Hall)
Car R and Parinello M 1985 *Phys. Rev. Lett.* **55** 2471
—— 1988 *Phys. Rev. Lett.* **60** 204
Ceperley D M and Alder B J 1980 *Phys. Rev. Lett.* **45** 566
Cohen M L 1986 *Science* **234** 549
Davidson E R 1975 *J. Comput. Phys.* **17** 87
Gonze X, Vigneron J P and Michenaud J P 1988 *Phys. Scr.* **37** 785
Hamann D R, Schlüter M and Chiang C 1979 *Phys. Rev. Lett.* **43** 1494
Honenberg P and Kohn W 1964 *Phys. Rev.* **136** B864
Holzschuh E 1983 *Phys. Rev.* B **28** 7346
Hybertsen M S and Louie S G 1985 *Phys. Rev. Lett.* **43** 1418
—— 1986 *Phys. Rev.* B **34** 2920
Ihm J, Zunger A and Cohen M L 1979 *J. Phys. C: Solid State Phys.* **12** 4409
Kleinman L 1980 *Phys. Rev.* B **21** 2630
Koelling D D 1981 *Rep. Prog. Phys.* **44** 139
Kohn W and Sham L J 1965 *Phys. Rev.* **140** A1133
Martins J L and Cohen M L 1988 *Phys. Rev.* B **37** 6134
Monkhorst H J and Pack J D 1976 *Phys. Rev.* B **13** 5188
Needels M, Payne M C and Joannopoulos J D 1987 *Phys. Rev. Lett.* **58** 1765
Nielsen O H and Martin R M 1985 *Phys. Rev.* B **32** 3792
Payne M C, Bristowe P D and Joannopoulos J D 1987 *Phys. Rev. Lett.* **58** 1348
Payne M C, Joannopoulos J D, Allan D C, Teter M P and Vanderbilt D H 1986 *Phys. Rev. Lett.* **56** 2656
Perdew J P 1986 *Phys. Rev.* B **33** 8822
Perdew J P and Yue W 1986 *Phys. Rev.* B **33** 8800
Perdew J P and Zunger A 1981 *Phys. Rev.* B **23** 5048
Wentzcovitch R M and Cohen M L 1986 *J. Phys. C: Solid State Phys.* **19** 6791
Wood D M and Zunger A 1985 *J. Phys. A: Math. Gen.* **18** 1343
Yin M T and Cohen M L 1980 *Phys. Rev. Lett.* **45** 1004
—— 1982 *Phys. Rev.* **26** 3259