Localized non-orthogonal orbitals in silicon

2001 J. Phys.: Condens. Matter 13 5731

(http://iopscience.iop.org/0953-8984/13/25/301)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **13** (2001) 5731–5741 www.iop.org/Journals/cm PII: S0953-8984(01)25134-0

Localized non-orthogonal orbitals in silicon

Jens Jørgen Mortensen[1] **and Michele Parrinello**

Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1, 70569 Stuttgart, Germany

E-mail: jensj@fysik.dtu.dk

Received 23 May 2001

Abstract

A comparison of orthogonal and non-orthogonal localized wavefunctions for Si in the diamond structure is carried out. We have used a real-space grid formulation of density functional theory in combination with the local density approximation for exchange and correlation to describe the energetics. Maximally localized wavefunctions, obtained from the extended Kohn–Sham states with and without an orthogonality constraint, are calculated and it is found that the wavefunctions calculated without any orthogonality constraint are the most localized. When solving directly for localized states, by applying a localization constraint to each electronic state, we find that there is a large difference between orthogonal and non-orthogonal states: when the localization region is a sphere with a radius of 3.0 Å, we get an error in the total energy due to the localization constraint of 0.2 and 2.7 eV/atom for non-orthogonal and orthogonal wavefunctions respectively.

1. Introduction

The study of large and complex systems with *ab initio* electronic structure methods is today done routinely for systems containing of the order of 10^2 atoms [1–3]. The desire to treat even larger systems has boosted interest in the development of so-called order-N methods, where the computational cost, in terms of time and memory, scales linearly with the number of atoms or electrons (for a recent review see [4]). A key ingredient in these methods is exploiting the locality of chemical binding [5]. In a wavefunction-based approach this means that the wavefunctions are localized in space and therefore only interact with a fixed number of other electronic states, this number being independent of system size. Localized wavefunctions are appealing also from a conceptual point of view: each state can be ascribed to a specific position in space, such as a bond between atoms or a lone pair. The fact that the wavefunctions depend only on the local chemistry makes them transferable from one system to a similar one, which could be useful in many situations. For example, localized orbitals could serve as a buffer at the interface between a classical model and a quantum mechanical model [6].

[1] Author to whom correspondence should be addressed. Present adress: Techinical University of Denmark, Building 307, 2800 Lyngby, Denmark

0953-8984/01/255731+11$30.00

Recently, non-orthogonal wavefunctions have attracted much attention due to their better transferability and more localized nature [7–12]. In this paper we compare localized wavefunctions for silicon in the diamond structure with and without an orthogonality constraint.

Solving directly for localized wavefunctions by forcing the wavefunctions to be zero outside a localization region introduces an error in the total energy. We address the question of how this error decays with localization region size, and show that it decays much faster for non-orthogonal wavefunctions than for orthogonal wavefunctions. This effect of non-orthogonality has been known for many years, but how large the effect is, is not well known.

The paper is organized as follows: we first introduce a generalized Kohn–Sham functional [10, 13–15], which depends on a set of non-orthogonal wavefunctions. This functional can be minimized without any orthogonality constraint, in contrast to the original Kohn–Sham functional [16, 17]. In section 3, we describe the details of our real-space grid calculation for Si. Section 4 compares the results of calculations with localized wavefunctions with and without an orthogonality constraint. In section 5, a number of physical quantities, such as the lattice constant and the bulk modulus, are calculated as a function of localization region size. Finally, the efficiency of the method developed is discussed, as well as the possibility of obtaining true order-N scaling in DFT calculations by using non-orthogonal wavefunctions. In the appendix, we show how localized non-orthogonal wavefunctions can be obtained indirectly from a given set of Kohn–Sham wavefunctions: we present an approach to transforming Kohn–Sham wavefunctions, extended over all atoms, into non-orthogonal wavefunctions maximally localized in space.

2. Energy functionals

Using density functional theory [16], the electronic energy of a spin unpolarized system can be expressed in terms of a set of doubly occupied wavefunctions $\{\psi_i\}$ that minimize the functional:

$$E[\{\psi_i\}] = 2\sum_i \int \mathrm{d}\vec{r}\,\psi_i\left(-\frac{1}{2}\nabla^2\right)\psi_i + \int \mathrm{d}\vec{r}\,v_{\mathrm{ext}}n + F[n] \tag{1}$$

under the constraint that the wavefunctions are orthonormal ($\int \mathrm{d}\vec{r}\,\psi_i\psi_j = \delta_{ij}$). In this equation we assume that the occupied states are separated from the unoccupied states by a bandgap, as is the case for bulk silicon, which is our test case in this paper. The first term in equation (1) is the kinetic energy of a non-interacting electron gas with electron density $n = 2\sum_i \psi_i^2$ [17]. The second term is the external energy, and $F[n]$ is a universal density functional that includes the Hartree energy, and the exchange and correlation energy:

$$F[n] = \frac{1}{2}\int \mathrm{d}\vec{r} \int \mathrm{d}\vec{r}\,'\frac{n(\vec{r})n(\vec{r}\,')}{|\vec{r}\,' - \vec{r}|} + E_{\mathrm{xc}}[n] \tag{2}$$

The external potential v_{ext} contains the Coulomb potential from the ions. The Kohn–Sham wavefunctions are the solutions to the Kohn–Sham equation:

$$-\frac{1}{2}\nabla^2\psi_i + v_{\mathrm{ext}}\psi_i + \frac{\partial F}{\partial n}\psi_i = \epsilon_i\psi_i \tag{3}$$

which must be solved self-consistently.

The ψ_i will in general extend over the whole system. However, any linear combination of the Kohn–Sham wavefunctions, such as:

$$\phi_i = \sum_j \mathbf{U}_{ij}\psi_j, \tag{4}$$

where $\mathbf{U}$ is a unitary matrix, will also minimize equation (1). A unitary matrix ensures that the ϕ_i form an orthonormal set of wavefunctions. The extra degrees of freedom available to us

by choosing **U** as any unitary matrix allows us to localize the wavefunctions [18–20]. As an example, we can choose **U** so that we get maximally localized Wannier functions (WF). The solid line in figure 1 shows what the wavefunctions will look like for bulk Si. This WF has been generated with the method described in the appendix. Each wavefunction has its WF centre (WFC) at the middle of a Si–Si bond, and the figure shows the weight of the wavefunction as a function of the distance from the bond-centre.

It can be seen that the Wannier function is localized and its weight is mainly on the two neighbouring Si atoms approximately 1 Å away. However, we still see some structure in the tail of the wavefunction. This structure is present because of the orthogonality constraint. The position of the minima in ϕ_i^2 correspond to the positions where ϕ_i must have nodes in order to be orthogonal to the neighbouring orbitals.

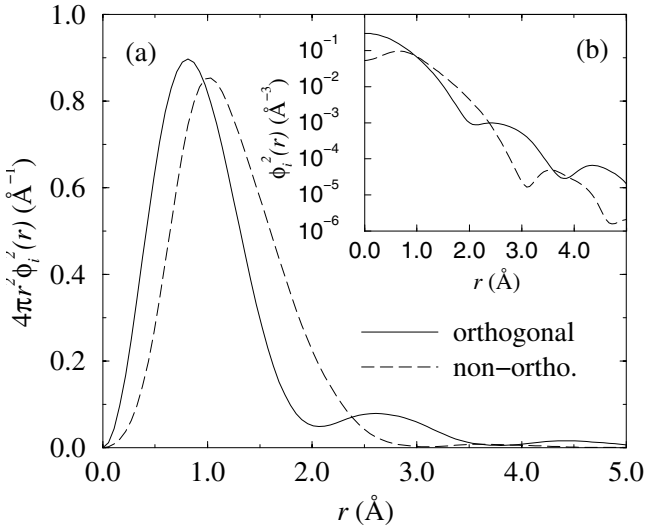

Figure 1. (a) Spherical average of $4\pi r^2\phi_i^2$ as a function of the distance from the corresponing WFC. Solid line: orthogonal wavefunction. Dashed line: non-orthogonal wavefunction. Both functions integrate to one electron. (b) Square of the wavefunctions on a logarithmic scale.

If we allow **U** to be any non-singular matrix, then we have more degrees of freedom, which, in principle, can lead to more localized wavefunctions. These wavefunctions will, in general, not be orthogonal to each other. One such maximally localized non-orthogonal wavefunction for Si is also shown in figure 1 (dashed line). Clearly this function decays faster than the solid line. The way we choose **U** to get maximally localized non-orthogonal wavefunctions is also described in the appendix.

2.1. A generalized Kohn–Sham functional

If we insert the transformation in equation (4) into equation (1) we obtain [10, 13–15]:

$$E[\{\phi_i\}] = 2\sum_{ij}\mathbf{S}_{ij}^{-1}\int \mathrm{d}\vec{r}\,\phi_i\left(-\frac{1}{2}\nabla^2\right)\phi_j + \int \mathrm{d}\vec{r}\,v_{\mathrm{ext}}n + F[n] \tag{5}$$

where $\mathbf{S}_{ij} = (\mathbf{U}\mathbf{U}^{\mathrm{T}})_{ij} = \int \mathrm{d}\vec{r}\phi_i\phi_j$ is the overlap matrix. The density is:

$$n = 2\sum_{ij}\mathbf{S}_{ij}^{-1}\phi_i\phi_j. \tag{6}$$

With this functional we can work with non-orthogonal wavefunctions. Notice that we no longer have any orthogonality constraint to fulfill.

The electronic gradient needed for the minimization of the energy functional in equation (5) is:

$$G_i = 1/2\frac{\delta E}{\delta \phi_i} = \sum_j \mathbf{S}_{ij}^{-1}\hat{H}\phi_j - \sum_j (\mathbf{S}^{-1}\mathbf{H}\mathbf{S}^{-1})_{ij}\phi_j \tag{7}$$

where $\mathbf{H}_{ij} = \int \mathrm{d}\vec{r}\phi_i \hat{H}\phi_j$ and $\hat{H} = -\frac{1}{2}\nabla^2 + v_{\mathrm{ext}} + \delta F/\delta n$. At the electronic groundstate we must have for all points in space that $G_i = 0$. By multiplying $G_{i'}$ by $\mathbf{S}_{ii'}$ and summing over i' we obtain:

$$G_i' = \hat{H}\phi_i - \sum_j (\mathbf{H}\mathbf{S}^{-1})_{ij}\phi_j. \tag{8}$$

We see that $G_i = 0$ is equivalent to $G_i' = 0$. Using the function G_i' as the electronic gradient in the minimization procedure was found by Fattebert *et al* [12] to give faster convergence.

We shall also be working with the functional proposed by Ordejón *et al* [21] and Mauri *et al* [22]. When minimized, this functional will automatically give orthogonal wavefunctions (OWF), without applying any explicit orthogonality constraints. The OWF functional looks like the functional in equation (5) with $\mathbf{S}_{ij}^{-1}$ replaced by $2\delta_{ij} - \mathbf{S}_{ij}$

$$E[\{\phi_i\}] = 2\sum_{ij}(2\delta_{ij} - \mathbf{S}_{ij}) \int \mathrm{d}\vec{r}\phi_i\left(-\frac{1}{2}\nabla^2\right)\phi_j + \int \mathrm{d}\vec{r} v_{\mathrm{ext}} n + F[n]. \tag{9}$$

The same replacement should be made in equation (6) for the density[2].

$$n = 2\sum_{ij}(2\delta_{ij} - \mathbf{S}_{ij})\phi_i\phi_j. \tag{10}$$

3. Bulk silicon

We have minimized the energy functional in equation (5) with respect to the ϕ_i for a system of 64 bulk Si atoms in a cubic box with periodic boundary conditions and Brillouin zone sampling limited to the $\vec{\Gamma}$-point. The local density approximation for the exchange and correlation energy [23,24] and a local pseudopotential for Si[3] [25] are used. All wavefunctions, potentials and electron densities are described on grids in real space with $32 \times 32 \times 32$ points. The grid spacing is $h \simeq 0.34$ Å, equivalent to a planewave cut-off of 24 Ry ($E_{\mathrm{cut}} = (\pi/h)^2/2$). A finite difference formula is used for the Laplacian operator, $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$. For the x-component, the formula looks as follows [26]:

$$\frac{\partial^2\psi(x)}{\partial x^2} = \frac{1}{h^2}\sum_{n=-2}^{2} c_n \psi(x + nh) \tag{11}$$

$$c_0 = -\frac{5}{2} \qquad c_{\pm 1} = \frac{4}{3} \qquad c_{\pm 2} = -\frac{1}{12} \tag{12}$$

with similar equations for $\partial^2/\partial y^2$ and $\partial^2/\partial z^2$. An iterative technique is used to find the groundstate: the electronic gradient is preconditioned [27] and then used to update the wavefunctions within the DIIS scheme [28] (direct inversion in iterative subspace). The density

[2] For the OWF functional, we have shifted the potential down by 3 Ry, so that all eigenvalues are negative. See [21] and [22] for details.

[3] The α parameter of the Appelbaum–Hamann pseudopotential has been multiplied by a factor of 1.17, in order to achieve better results for the lattice constant.

is updated by simply mixing the old density with the new density in a ratio of approximately 2:1. The Hartree potential and Hartree energy is found using Fourier transforms.

Preconditioning of the electronic gradient considerably improves the convergence of the total energy. Preconditioning is done as in [27], only slightly modified. We define the 'smearing' operator $\hat{P}$ as:

$$\hat{P} f(\vec{r}) = \alpha f(\vec{r}) + \frac{1-\alpha}{6} \sum_{\vec{\Delta}} f(\vec{r} + \vec{\Delta}) \tag{13}$$

where the sum runs over the six nearest neighbour grid points. The preconditioning operator that we apply to the electronic gradient is defined as $\hat{K} = \hat{P}^n$, where n is between 5 and 7. With $\alpha = 0.5$ this preconditioner is identical to that in [27]. However, for our system, we achieve better results with $\alpha = 0.8$.

4. Localized wavefunctions

The dashed line in figure 1 shows that it is possible to describe the electronic structure of bulk silicon in terms of highly localized non-orthogonal wavefunctions, by transforming the Kohn–Sham wavefunctions into Wannier-like functions. We now wish to calculate the ϕ_i directly, without having to calculate the extended Kohn–Sham wavefunctions first. We do this by minimizing the energy functional under the constraint that each wavefunction should be zero outside a given localization region. Since there are two bonds and two doubly occupied electronic states for each Si atom, it is natural to centre each wavefunction or localization region on the midpoint of the Si–Si bonds. We let each localization region be a sphere with radius r_{cut}. Any discontinuities in the wavefunctions at the border of the localization region will give a large contribution to the kinetic energy. Therefore, the wavefunctions will decay smoothly to zero at the border when the energy is minimized.

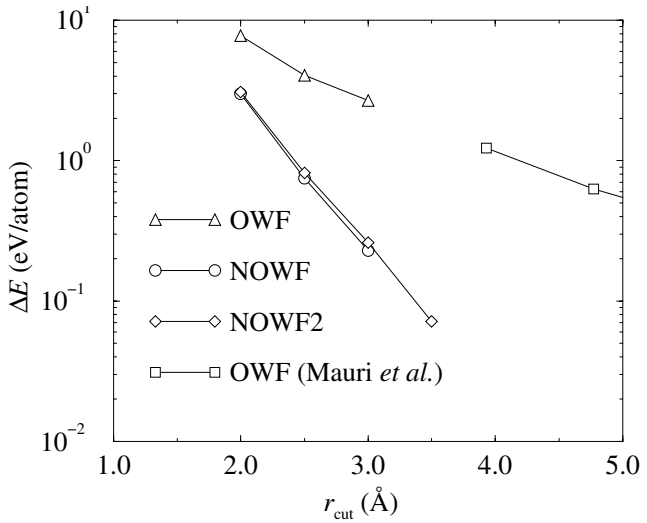

Figure 2. Error in the total energy as a function of localization region radius. Triangles: orthogonal wavefunctions (OWF functional). Circles: non-orthogonal wavefunctions and $G_i = 0$. Diamonds: non-orthogonal wavefunctions and $G'_i = 0$. Squares: results from [29] with the OWF functional. (In [29] the localization region is a cube and not a sphere—we have converted the cube length to a sphere radius so that the volumes are the same.)

Putting a constraint on the wavefunctions introduces an error, ΔE, into the calculation of the energy relative to the calculation without any constraints. Figure 1 suggests that the error should go to zero with increasing localization region. Since our energy functional in equation (5) is variational we will have $\Delta E > 0$.

We will now calculate the energy using two different functionals: the non-orthogonal wavefunction (NOWF) functional in equation (5) and the OWF functional of Ordejón *et al* [21] and Mauri *et al* [22] in equation (9).

4.1. Non-orthogonal wavefunctions

In figure 2 we have plotted the error for non-orthogonal wavefunctions as a function of r_{cut} (circles). As expected, the error goes to zero as the localization region increases in size. The energies are obtained by solving, self-consistently, the equations $G_i = 0$ (or $\delta E/\delta\phi_i = 0$) for each i. For each electronic state i, the equation must hold at each gridpoint inside the localization region of state i. This procedure gives us the minimum of the functional in equation (5) under the localization constraint.

When we have a localization constraint for each wavefunction, $G_i = 0$ is no longer equivalent to $G_i' = 0$. If we solve $G_i' = 0$, as was done in the work of Fattebert *et al* [12], instead of $\delta E/\delta\phi_i = 0$, we get an energy slightly higher than the minimum energy (we will denote this type of calculation by NOWF2). These results are also shown in figure 2 (diamonds). We shall return to this small energy difference later when we look at physical quantities such as lattice constants and phonon frequencies.

4.2. Orthogonal wavefunctions

Looking at figure 1, one would expect that the localization error would be bigger for orthogonal wavefunctions, since they decay more slowly. In order to find out if this is the case, we also calculate localized orthogonal (or nearly orthogonal) wavefunctions using the density functional proposed by Ordejón *et al* [21] and Mauri *et al* [22] (OWF, equation (9)). When minimized with a localization constraint, the OWF functional will give approximately orthogonal wavefunctions. The reason that we do not obtain exactly orthogonal wavefunctions is that strict localization is not compatible with orthogonality. Exactly how close to orthogonality the wavefunctions are can be illustrated by looking at the quantity $\Delta N = \int \mathrm{d}\vec{r} n - N = -2\mathrm{Tr}\{(\mathbf{S} - \mathbf{1})^2\}$ ($N = 256$ electrons), which will be zero without a localization constraint. We get $\Delta N/N = -4 \times 10^{-4}$ for $r_{\text{cut}} = 3.0$ Å, in good agreement with the work of Fernández *et al* [29].

Figure 2 shows the error for orthogonal wavefunctions (triangles) as well as the results of Fernández *et al* [29] for a similar calculation with orthogonal wavefunctions (squares). We see that our calculations with the OWF functional are in good agreement with those from [29] (see also [30] and [31]). We also see that the error goes much faster to zero when the wavefunctions are allowed to be non-orthogonal. If, for example, we accept an error in the total energy of 0.1 eV/atom (the error in energy differences and other physical quantities of interest are expected to be smaller due to error cancellations), then we need a localization region with a radius of approximately 3.5 and 7 Å in the case of non-orthogonal and orthogonal wavefunctions respectively. This amounts to a difference of a factor of 8 in the amount of real space volume needed to represent each wavefunction.

5. Physical quantities

Total energies are not interesting in themselves. We have therefore calculated several physical properties for silicon: the lattice constant (a), the bulk modulus (B), and the frequency of a zone-centre transverse optical phonon (ω_{TO}). We will now look at how these quantities converge to the values calculated without any localization constraint, as a function of r_{cut}.

It should be noted that for quantitatively accurate LDA results an all-electron calculation or an accurate non-local pseudopotential would be required. For this reason we do not compare our results with other LDA calculations, although the agreement is good.

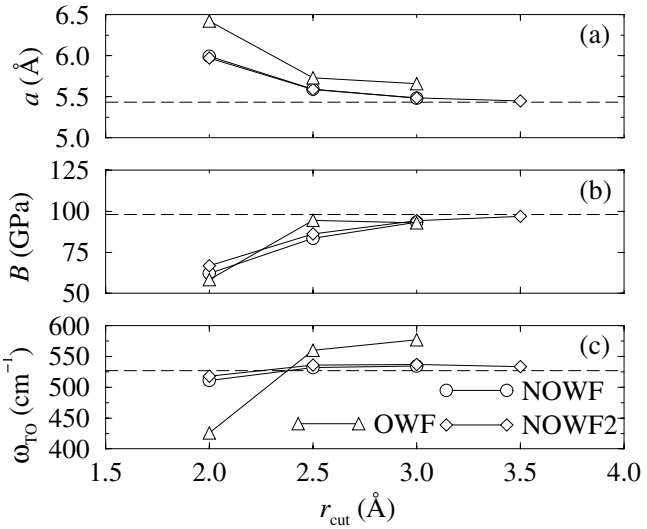

Figure 3. (a) Lattice constant. (b) Bulk modulus. (c) Frequency of a zone-centre transverse optical phonon. Triangles: orthogonal wavefunctions (OWF functional). Circles: non-orthogonal wavefunctions and $G_i = 0$. Diamonds: non-orthogonal wavefunctions and $G'_i = 0$. The dashed lines show the exact results with no localization constraints.

Figure 3 shows results for a, B, and ω_{TO} using: orthogonal wavefunctions (OWF functional, triangles), non-orthogonal wavefunctions and $G_i = 0$ (circles), and non-orthogonal wavefunctions and $G'_i = 0$ (diamonds). As observed for the convergence of the total energy with localization region size, we find also for the convergence of physical quantities that $G_i = 0$ and $G'_i = 0$ give almost identical result (NOWF and NOWF2). As mentioned above, for a NOWF2 calculation, we do not obtain the minimum energy, but this small error in the total energy is almost completely cancelled out when we calculate energy differences.

One can also observe that the convergence is much faster with non-orthogonal wavefunctions than with orthogonal wavefunctions. As an example, a, B, and ω_{TO} are all converged to within $\sim$1% of their exact values, with no localization constraint, at a radius of $r_{\mathrm{cut}} \simeq 3.0$–$3.5$ Å for non-orthogonal wavefunctions. For orthogonal wavefunctions, it seems that a radius of at least 5.0 Å would be needed for a calculation of the same quality.

6. Efficiency

With the functional in equation (5), we need to invert the overlap matrix at each iteration. This requires, in principle, of the order of N^3 operations. However, this part of the calculation is

Table 1. Lattice constant (a), bulk modulus (B), and frequency of a zone-centre transverse optical phonon (ω_{TO}). Calculations for different values of r_{cut} using the NOWF functional. The last row gives the results with no localization constraints.

r_{cut} (Å)	ΔE (eV/atom)	a (Å)	B (GPa)	ω_{TO} (cm^{-1})
2.0	3.0	5.99	62.1	511
2.5	0.75	5.59	83.6	532
3.0	0.23	5.48	93.3	535
3.5[a]	0.072	5.45	96.9	533
–	0.0	5.43	98.0	527

[a] Results for $r_{\mathrm{cut}} = 3.5$ Å are obtained from a NOWF2 calculation.

a small part of the total calculation, and we could reduce the scaling of the matrix inversion by taking advantage of the fact that the overlap matrix is sparse. In practice each iteration is therefore an order-N operation, as it is for the OWF functional, where $\mathbf{S}^{-1}$ is not needed. Unfortunately, both equation (5) and the OWF functional are ill conditioned when a localization constraint is applied, and the number of iterations needed to obtain the groundstate increases drastically with the size of the localization region. This is a well-known problem of many order-N schemes [4, 32, 33]. The reason is that the introduction of a localization constraint destroys the invariance of the energy with respect to a unitary transformation of the wavefunctions [4].

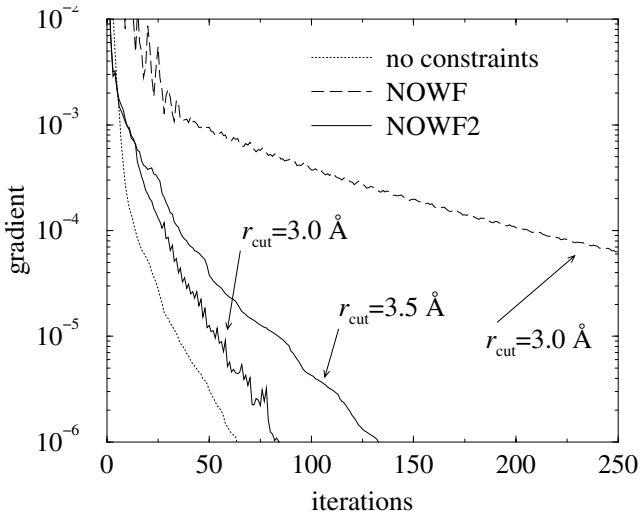

Figure 4. Convergence of the electronic gradient with iteration number. Dotted line: calculation with no constraints. Dashed line: NOWF calculation. Solid lines: NOWF2 calculations with $r_{\mathrm{cut}} = 3.0$ and 3.5 Å. Four old wavefunctions and gradients are used in the DIIS update.

In order to see how fast or how slow the energy converges, we make a small random displacement of all atoms (maximum displacement: 0.05 Å), and then start the calculation from the fully-converged wavefunctions for the perfect bulk structure. Figure 4 shows how the norm of the electronic gradient converges to zero as a function of the number of iterations. The norm is defined as:

$$\sqrt{\frac{2}{N}\sum_i \int \mathrm{d}\vec{r}\, G_i^2}. \tag{14}$$

The dotted line shows that without any localization constraints, the convergence is quite fast. In order to converge the energy to an accuracy of 10^{-5} eV/atom, 21 iterations are needed.

The dashed line shows the results of a NOWF calculation with a localization region radius of 3.0 Å. In order to reach the same accuracy 231 iterations were needed. When increasing r_{cut} to 3.5 Å we were not able to converge the energy. This clearly shows that this scheme is not practical for electronic structure calculations.

The two solid lines show the results for the case where G_i' is used for the electronic gradient (NOWF2). For this case 59 and 105 iterations are required for $r_{\mathrm{cut}} = 3.0$ and 3.5 Å respectively. If we compare the timings for one iteration for a NOWF2 calculation and a calculation with no constraints and extended wavefunctions, we find that for $r_{\mathrm{cut}} = 3.0$ the NOWF2 calculation is faster by a factor of three and for $r_{\mathrm{cut}} = 3.5$ the two methods are equally fast. Furthermore, a NOWF2 calculation with $r_{\mathrm{cut}} = 3.5$ uses only a third of the memory that is required to do a standard calculation with extended wavefunctions. It should be noted that these numbers are for only 64 Si atoms, where the linear scaling of time and memory usage has not yet been reached.

One problem with the NOWF2 scheme is that the Hellmann–Feynman forces are not exact: the force acting on atom number I at position $\vec{R}_I$ is:

$$\vec{F}_I = -\frac{\mathrm{d}E}{\mathrm{d}\vec{R}_I} = -\sum_i \int \mathrm{d}\vec{r} \frac{\delta E}{\delta \phi_i} \frac{\mathrm{d}\phi_i}{\mathrm{d}\vec{R}_I} - \frac{\partial E}{\partial \vec{R}_I}. \tag{15}$$

In this equation we cannot throw away the first term, as is usually done, because $\delta E/\delta\phi_i$ is not zero (in the NOWF2 scheme, $G_i' = 0$ is solved and not $\delta E/\delta\phi_i = 0$).

7. Conclusions

Our results clearly show the advantages of working with non-orthogonal wavefunctions rather than orthogonal wavefunctions. We have shown that non-orthogonal wavefunctions can be localized considerably more than orthogonal wavefunctions, without affecting the physical results. Also the low memory consumption allows one to use more memory-expensive methods such as DIIS and wavefunction extrapolations [13].

For an order-N method to be faster than a standard method with scaling between order-N^2 and order-N^3, the number of atoms must be larger than some critical value. This critical number of atoms is proportional to the volume of the localization region. It is therefore essential that an efficient order-N scheme uses non-orthogonal wavefunctions.

However, there are a number of problems and questions that need to be solved and answered before this method can be used routinely: for example, how to calculate accurate forces and how to choose the localization regions for a dynamically evolving system.

Appendix

This appendix describes how we have obtained the localized wavefunctions shown in figure 1. We start from the Kohn–Sham wavefunctions ψ_i, and then look for a transformation matrix $\mathbf{U}$ that gives the most localized ϕ_i.

We first look at the situation where $\mathbf{U}$ must be unitary. For a periodic system, the optimal $\mathbf{U}$ is found by maximizing the function [18–20]:

$$\Omega = \sum_i \left(\left| \int \mathrm{d}\vec{r} \phi_i^2 \mathrm{e}^{\mathrm{i}2\pi x/L} \right|^2 + \left| \int \mathrm{d}\vec{r} \phi_i^2 \mathrm{e}^{\mathrm{i}2\pi y/L} \right|^2 + \left| \int \mathrm{d}\vec{r} \phi_i^2 \mathrm{e}^{\mathrm{i}2\pi z/L} \right|^2 \right). \tag{.1}$$

If the the ϕ_i are spread out over the whole box (box length L), then the integrals with the oscillating exponential functions will be small. On the other hand, if all ϕ_i^2 are delta-functions, then Ω will have its maximum value $3N/2$, where $N/2$ is the number of electronic states.

The function Ω can be written in a more compact form:

$$\Omega = \sum_i |\vec{Z}_i|^2 \tag{.2}$$

where $\vec{Z}_i$ is a complex vector and $\vec{\mathbf{Q}}$ is a matrix of complex vectors:

$$\vec{Z}_i = \int \mathrm{d}\vec{r}\phi_i^2 \exp\left(\mathrm{i}\frac{2\pi}{L}\vec{r}\right) = (\mathbf{U}\vec{\mathbf{Q}}\mathbf{U}^{\mathrm{T}})_{ii} \tag{.3}$$

$$\vec{\mathbf{Q}}_{mn} = \int \mathrm{d}\vec{r}\psi_m\psi_n \exp\left(\mathrm{i}\frac{2\pi}{L}\vec{r}\right). \tag{.4}$$

We have used the notation: $\exp(\vec{v}) = \{\exp(v_x), \exp(v_y), \exp(v_z)\}$. Maximizing Ω with respect to the elements of $\mathbf{U}$, while keeping $\mathbf{U}$ unitary, is done using the scheme of Edmiston *et al* [34, 35], where $\mathbf{U}$ is built up from a series of 2×2 rotation matrices. This results in Wannier functions with Wannier function centres (WFC's):

$$\vec{R}_i = \frac{L}{2\pi}\mathrm{Im}\{\log\vec{Z}_i\} \tag{.5}$$

positioned on the bonds between the Si atoms.

In order to find maximally localized non-orthogonal Wannier functions, we use the WFC's determined above[5] and find the matrix $\mathbf{U}$ by minimizing the following function independently for each i:

$$\omega_i = \int \mathrm{d}\vec{r}\, p(\vec{r} - \vec{R}_i)\phi_i^2 \tag{.6}$$

under the constraint that ϕ_i should be normalized. The penalty function $p(\vec{r})$ must have a minimum at $r = 0$. This will localize each ϕ_i around its WFC, $\vec{R}_i$. Inserting equation (4) into equation (.6), we see that ω_i depends only on the ith row of $\mathbf{U}$, which we denote $\vec{u}_i$:

$$\omega_i = (\mathbf{U}\mathbf{P}^{(i)}\mathbf{U}^{\mathrm{T}})_{ii} = \vec{u}_i^{\mathrm{T}}\mathbf{P}^{(i)}\vec{u}_i \tag{.7}$$

where:

$$\mathbf{P}^{(i)}_{mn} = \int \mathrm{d}\vec{r}\psi_m\, p(\vec{r} - \vec{R}_i)\psi_n. \tag{.8}$$

The minimum of ω_i is obtained when $\vec{u}_i$ is equal to the normalized eigenvector corresponding to the smallest eigenvalue of $\mathbf{P}^{(i)}$. A natural choice, for a periodic system, for $p(\vec{r})$ is:

$$p(\vec{r}) = \sum_{\alpha=x,y,z}\left[1 - \cos\left(\frac{2\pi}{L}r_\alpha\right)\right] \tag{.9}$$

$$= \frac{2\pi^2}{L^2}r^2 \quad \text{for} \quad r \ll L. \tag{.10}$$

With this choice $\mathbf{P}^{(i)}_{mn}$ can easily be calculated from $\vec{\mathbf{Q}}_{mn}$:

$$\mathbf{P}^{(i)}_{mn} = 3\delta_{mn} - \mathrm{Re}\left\{\vec{\mathbf{Q}}_{mn}\exp\left(-\mathrm{i}\frac{2\pi}{L}\vec{R}_i\right)\right\}. \tag{.11}$$

[5] One might envisage a procedure in which the $\vec{R}_i$ and the non-orthogonal wavefunctions are simultaneously optimized. However, since this paper has an exploratory character, we have not attempted to do so. Furthermore, in bulk Si, symmetry dictates the WFC positions.

References

[1] Marx D and Hutter J 2000 1997 *Modern Methods and Algorithms of Quantum Chemistry* ed J Grotendorst Vol 1
[2] Parrinello M 1997 *Solid State Commun.* **102** 107
[3] Kresse G and Furthmüller J 1996 *Comput. Mat. Sci.* **6** 15
[4] Goedecker S 1999 *Rev. Mod. Phys.* **71** 1085
[5] Kohn W 1996 *Phys. Rev. Lett.* **76** 3168
[6] Assfeld X and Rivail J-L 1996 *Chem. Phys. Lett.* **263** 100
[7] Anderson P W 1968 *Phys. Rev. Lett.* **21** 13
[8] Yang W 1997 *Phys. Rev.* B **56** 9294
[9] Hierse W and Stechel E B 1994 *Phys. Rev.* B **50** 17811
[10] Stechel E B, Williams A R and Feibelman P J 1994 *Phys. Rev.* B **49** 10088
[11] Liu S, Péres-Jordá J M and Yang W 2000 *J. Chem. Phys.* **112** 1634
[12] Fattebert J-L and Bernholc J 2000 *Phys. Rev.* B **62** 1713
[13] Arias T A, Payne M C and Joannopoulos J D 1992 *Phys. Rev. Lett.* **69** 1077
[14] Galli G and Parrinello M 1992 *Phys. Rev. Lett.* **69** 3547
[15] Pandey K C, Williams A R and Janak J F 1995 *Phys. Rev.* B **52** 14415
[16] Hohenberg P and Kohn W 1964 *Phys. Rev.* **136** B864
[17] Kohn W and Sham L J 1965 *Phys. Rev.* A **140** 1133
[18] Marzari N and Vanderbilt D 1997 *Phys. Rev.* B **56** 12847
[19] Resta R 1998 *Phys. Rev. Lett.* **80** 1800
[20] Silvestrelli P L, Marzari N, Vanderbilt D and Parrinello M 1998 *Solid State Commun.* **107** 7
[21] Ordejón P, Drabold D A, Grumbach M P and Martin R M 1993 *Phys. Rev.* B **48** 14646
[22] Mauri F, Galli G and Car R 1993 *Phys. Rev.* B **47** 9973
[23] Ceperly D M and Alder B J 1980 *Phys. Rev. Lett.* **45** 566
[24] Perdew J P and Zunger A 1981 *Phys. Rev.* B **23** 5048
[25] Appelbaum J A and Hamann D R 1973 *Phys. Rev.* B **8** 1777
[26] Chelikowsky J R, Troullier N, Wu K and Saad Y 1994 *Phys. Rev.* B **50** 11355
[27] Hoshi T, Arai M and Fujiwara T 1995 *Phys. Rev.* B **52** R5459
[28] Pulay P 1980 *Chem. Phys. Lett.* **73** 393
[29] Fernández P, Corso A D, Baldereschi A and Mauri F 1997 *Phys. Rev.* B **55** R1909
[30] Stephan U and Drabold D A 1998 *Phys. Rev.* B **57** 6391
[31] Stephan U, Drabold D A and Martin R M 1998 *Phys. Rev.* B **58** 13472
[32] Ordejón P, Drabold D A, Martin R M and Grumbach M P 1995 *Phys. Rev.* B **51** 1456
[33] Bowker D R and Gillan M J 1998 *Comp. Phys. Commun.* **112** 103
[34] Edmiston C and Ruedenberg K 1963 *Rev. Mod. Phys.* **35** 457
[35] Berghold G 1999 *et al Phys. Rev.* B **61** 10040