First Year Report
12 June 1996

Linear Scaling Total Energy Pseudopotential Calculations for Large Systems

Peter Haynes
Christ's College

Supervisor: Dr. M. C. Payne

Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge, CB3 0HE, U.K.

Much interest has recently been shown in developing methods to perform *ab initio* total energy pseudopotential calculations within density functional theory (DFT) in which both the computational effort and memory required scale linearly with the system size (so-called $O(N)$ methods.) One approach involves minimisation of an energy functional with respect to the single-particle density matrix, and here this is developed from the conventional formulation of DFT. This method is implemented by discretising the wave functions and Hamiltonian on a real-space grid. A method is also presented for obtaining an initial estimate of the density matrix from conventional $O(N^3)$ calculations and the results compared with the best estimate obtained from an LCAO approach. Finally, a new implementation of the method which is currently being developed is described, which has several advantages over the real-space grid method and is much closer to traditional plane-wave methods.

1 Introduction

During the last decade, *ab initio* total energy pseudopotential methods based on density functional theory have become established as a major tool in condensed matter theory, and more recently the range of applicability has been extended to many other fields. Advances in computing resources have always been matched by the growing demand to model larger and more complex systems. Traditional minimisation techniques, based on the diagonalisation of the Hamiltonian, require a computational effort which scales with the cube of the system size (i.e. the number of atoms N) and require an amount of memory which scales as the square of the system size. Such scaling means that even a ten-fold increase in computing power could only be converted into roughly a two-fold increase in system size. There has therefore been a great deal of interest in methods which scale linearly with system size, both in terms of computational effort and memory requirements. The search for these methods is motivated by the fact that the complexity of the problem scales linearly with system size, and this suggests that the poorer scaling results from the way in which the problem is being tackled. In this report, I review one approach to this problem, based on the single-particle density matrix, and present schemes to implement this method, as well as a method for linking the new calculations to the results of traditional calculations.

2 Density Functional Theory

There are many references on the aspects and applications of density functional theory (DFT) [1, 2]. This section contains a brief review of the essential aspects and approximations involved in the conventional formulation.

2.1 Born-Oppenheimer Approximation

In dealing with systems of large numbers of electrons and ions, we wish to make the number of degrees of freedom of the system which have to be dealt with at any one time as small as possible. The forces on both the electrons and ions are of the same order of magnitude, so that their momenta are also comparable. However since the ions are so much more massive than the electrons, the kinetic energy of the ions is much smaller than that of the electrons. This idea forms the basis of the Born-Oppenheimer approximation [3] in which we assume that the electrons respond instantaneously to the motion of the ions. For any ionic configuration, we assume that the electrons are in the instantaneous ground-state and calculate the total energy of the system. Varying the ionic positions defines a multi-dimensional ground-state potential energy surface, and the motion of the ions can then be treated as classical particles moving in this potential. Ionic relaxation and molecular dynamics [4, 5] can be performed within the scheme to be described but (as this is not relevant to this work) from now on we shall assume that we are treating a system with a fixed ionic configuration.

2.2 The Hohenberg-Kohn Theorem and Density Functional Theory

The total energy of a system of interacting ions and electrons can be written as

$$E_{\text{tot}} = E_{\text{e-e}} + E_{\text{e-ion}} + E_{\text{ion-ion}}. \tag{1}$$

The first term describes the kinetic energy of the electrons and the potential energy of their interaction with each other. The second term describes the potential energy of the interaction between the electrons and ions, and the final term describes the interaction energy of the ions with themselves.

The result of the Hohenberg-Kohn theorem [6] which is the foundation of DFT is that the first term $E_{\text{e-e}}$ can be expressed as a universal functional of the ground-state electronic density, $F[n(\mathbf{r})]$. The second term can also be written in terms of the electronic density, $n(\mathbf{r})$, as

$$E_{\text{e-ion}} = \int d^3r \; V_{\text{ion}}(\mathbf{r})n(\mathbf{r}) \tag{2}$$

where $V_{\text{ion}}(\mathbf{r})$ is simply the electrostatic potential due to the ions. Including the ion-ion term (calculated from classical electrostatics for point charges) we can therefore write the total energy of the interacting system as a functional of the electronic density:

$$E_{\text{tot}} = F[n(\mathbf{r})] + \int d^3r \; V_{\text{ion}}(\mathbf{r})n(\mathbf{r}) + E_{\text{ion-ion}}. \tag{3}$$

By minimising this expression for the total energy with respect to the electronic density, it should be possible to find the ground-state energy and density. Since the complexity of this problem scales linearly with system size, this should yield an $O(N)$ method. Unfortunately, the form of the universal functional $F[n(\mathbf{r})]$ is unknown, and this straightforward approach is not possible.

2.3 The Kohn-Sham Equations

Kohn and Sham [7] showed that it is possible to map the many-body problem of an interacting electron gas (in the presence of an external potential) to an exactly equivalent set of self-consistent single-electron equations. The functional is now written in terms of a set of doubly-occupied single-particle electronic states $\{\psi_i\}$,

$$E_{\text{e-e}} = F[\{\psi_i\}] = 2\sum_i \int d^3r \; \psi_i^*(\mathbf{r})(-\tfrac{1}{2})\nabla^2\psi_i(\mathbf{r}) + \tfrac{1}{2}\int d^3r \int d^3r' \; \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\text{XC}}[n(\mathbf{r})], \tag{4}$$

and the electronic density is given by

$$n(\mathbf{r}) = 2\sum_i |\psi_i(\mathbf{r})|^2. \tag{5}$$

The first term gives the kinetic energy of a non-interacting electron gas, the second term is the classical electrostatic potential (or Hartree) energy, and the third term is known as the exchange-correlation energy.

The many-body wave function has to be antisymmetric under the exchange of any two electrons because they are fermions, and this results in a spatial separation for electrons of identical spin which reduces the Coulomb energy of the system. Spatially separating electrons with opposite spin also reduces the Coulomb energy, albeit at the cost of increasing the kinetic energy, an effect known as correlation. These effects make up the exchange-correlation energy $E_{\mathrm{XC}}[n(\mathbf{r})]$ which has to be included explicitly in the total energy expression because the electrons are now treated as independent particles, replacing the full quantum mechanical treatment involving the many-body wave function.

Minimising the Kohn-Sham energy functional with respect to the set of wave functions $\{\psi_i\}$, subject to the constraint that these wave functions be orthonormal, yields the Kohn-Sham equations,

$$\Big[-\tfrac{1}{2}\nabla^2 + V_{\mathrm{ion}}(\mathbf{r}) + V_{\mathrm{H}}(\mathbf{r}) + V_{\mathrm{XC}}(\mathbf{r})\Big]\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}), \tag{6}$$

which must be solved self-consistently. V_{H} is the Hartree potential of the electrons given by

$$V_{\mathrm{H}}(\mathbf{r}) = \int d^3r' \, \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{7}$$

and V_{XC} is the exchange-correlation potential defined formally as the functional derivative

$$V_{\mathrm{XC}}(\mathbf{r}) = \frac{\delta E_{\mathrm{XC}}[n(\mathbf{r})]}{\delta n(\mathbf{r})}. \tag{8}$$

From equation 4 we see that the original problem has now been mapped onto the problem of non-interacting electrons moving in an effective potential. When the Kohn-Sham equations are solved self-consistently, such that the occupied states generate the same electronic density used to construct the potential, then the ground-state density and energy have been found. The total energy is then given by

$$E_{\mathrm{tot}} = 2\sum_i \varepsilon_i - \tfrac{1}{2}\int d^3r \int d^3r' \, \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} - \int d^3r \; V_{\mathrm{XC}}(\mathbf{r})n(\mathbf{r}) + E_{\mathrm{XC}} + E_{\mathrm{ion-ion}}. \tag{9}$$

2.4 Local Density Approximation

The results obtained so far are exact, if the exact form of the exchange-correlation functional is known. Unfortunately, it is extremely difficult to calculate the correlation energy for a complex system. Instead, we make the local density approximation (LDA) which assumes that the exchange-correlation energy per electron at a point $\mathbf{r}$ in the electron gas is simply what it would be in a homogeneous electron gas of the same density as is found at $\mathbf{r}$:

$$E_{\mathrm{XC}}[n(\mathbf{r})] = \int d^3r \; \varepsilon_{\mathrm{XC}}(\mathbf{r})n(\mathbf{r}) \tag{10}$$

with

$$\varepsilon_{\mathrm{XC}}(\mathbf{r}) = \varepsilon_{\mathrm{XC}}^{\mathrm{hom}}(n(\mathbf{r})), \tag{11}$$

and hence

$$V_{\mathrm{XC}}(\mathbf{r}) = \frac{\partial}{\partial n}\Big\{n(\mathbf{r})\varepsilon_{\mathrm{XC}}[n(\mathbf{r})]\Big\}. \tag{12}$$

This is a vast simplification of the problem, which although exact in the limit of slowly varying densities, is difficult to justify given that the electron density in systems of interest is usually rapidly varying. The justification comes in part from the fact that the LDA obeys the correct sum rule [8] for the exchange-correlation hole, but ultimately from its success in predicting physical properties. Here we use the LDA results obtained from quantum Monte Carlo calculations by Ceperley and Alder [9] and parametrised by Perdew and Zunger [10].

2.5 The Pseudopotential Approximation

We are generally only interested in the valence electrons of atoms which are responsible for the formation of chemical bonds and govern most physical properties, and not in the core electrons of inner closed shells. The valence wave functions have high kinetic energy near the nucleus due to the oscillations required to maintain orthogonality with lower-lying core states, but this kinetic energy is largely cancelled by the high electrostatic potential energy near the nucleus. By exploiting this effective cancellation of part of the nuclear charge and the core electrons, we find that it is possible to replace the nucleus and core electrons by a pseudo-ion core with a much weaker potential. We also remove the oscillations of the valence wave function in the core region and replace the wave function by a smoother pseudo-wave function [11]. Outside the core region, the pseudo-wave function and pseudopotential are identical to the original ones, so that they reproduce the same electronic density there. The benefit of this approximation comes from the fact that the resulting potential is weaker so that the absolute value of the total energy is much smaller and can be computed much more accurately. We can also expand the wave functions in terms of a smaller basis set of plane-waves, and this is a central feature common to all conventional methods.

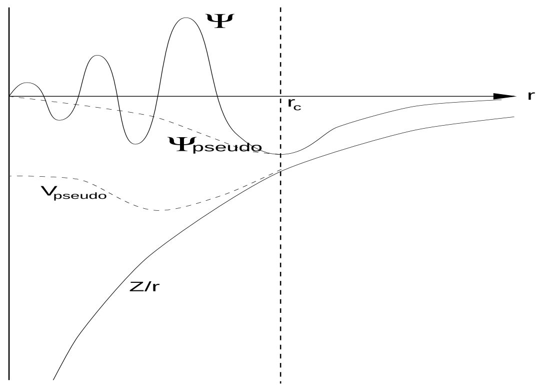

Figure 1: A sketch illustrating the introduction of a pseudopotential

In general, scattering from a potential produces different phase-shifts for each angular momentum component, so that the pseudopotential operator is of the form

$$\hat{V}_{\rm ps} = \sum_{l,m} |lm\rangle \, V_{{\rm ps},l} \, \langle lm| \tag{13}$$

where $|lm\rangle$ denotes the spherical harmonic Y_{lm}.

2.6 Periodic Boundary Conditions and Bloch's Theorem

One final point which should be mentioned is the role of Bloch's theorem in these calculations [12]. As the situation stands, we still have to calculate wave functions for an infinite number of electrons, which must be expressed in terms of an infinite basis because of the infinite extent of the system. As is well known, these problems can be overcome by performing calculations on periodic systems (which may involve introducing some fictitious long-range periodicity into aperiodic systems, as is demonstrated for the case of a single impurity atom in a crystal in figure 2) and applying Bloch's theorem.

Bloch's theorem states that the wave function in a periodic solid can be written as the product of a wave-like part and a cell-periodic part:

$$\psi_i(\mathbf{r}) = \exp[i\mathbf{k}\cdot\mathbf{r}]u_{\mathrm{i}}(\mathbf{r}). \tag{14}$$

The cell-periodic part can be written as a discrete sum of plane-waves with wave vectors which are reciprocal lattice vectors

$$u_{\mathrm{i}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathrm{i},\mathbf{G}} \exp[i\mathbf{G}\cdot\mathbf{r}] \tag{15}$$

so that each wave function can be written as

$$\psi_i(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathrm{i},\mathbf{k}+\mathbf{G}} \exp[i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}]. \tag{16}$$

Application of periodic boundary conditions means that electronic states are only allowed at discrete **k**-points and that these **k**-points can be restricted to the first Brillouin zone of the crystal. The density of **k**-points is proportional to the system volume, and so we now deal with the infinite number of electrons using a countably infinite set of **k**-points, with only a finite number of states occupied at each **k**-point. By choosing to truncate the plane-wave basis so that it only includes basis states with kinetic energy below some cut-off, we can also expand the states in terms of a finite basis. The calculations are usually repeated with increasing values of this cut-off until the quantity required has converged.

Since the form of the wave functions does not vary rapidly over the first Brillouin zone, the states can be accurately sampled by considering a much smaller set of **k**-points, and much work has been done on selecting efficient sets [13]. As with the kinetic energy cut-off, the calculations should be repeated with increasingly dense **k**-point sets until the quantity being calculated has converged. We note here however, that as the volume of the periodic unit (supercell) increases, so the volume of the first Brillouin zone decreases, so that if the number of **k**-points remains constant, the **k**-point density increases. Thus the larger the system modelled, the fewer the number of **k**-points required. Since this work is concerned with large systems, we only calculate states at a single **k**-point (usually the Γ-point) here.

Finally we note that the Kohn-Sham equations assume a particularly simple form in momentum space [14, 15] when plane-waves are used as the basis set, as the kinetic energy is diagonal in this representation, and the potentials are described by their Fourier transforms.

$$\sum_{\mathbf{G}'} \left[\tfrac{1}{2}|\mathbf{k}+\mathbf{G}|^2\delta_{\mathbf{G}\mathbf{G}'} + \tilde{V}_{\mathrm{ion}}(\mathbf{G}-\mathbf{G}') + \tilde{V}_{\mathrm{H}}(\mathbf{G}-\mathbf{G}') + \tilde{V}_{\mathrm{XC}}(\mathbf{G}-\mathbf{G}')\right] c_{\mathrm{i},\mathbf{k}+\mathbf{G}'} = \varepsilon_i c_{\mathrm{i},\mathbf{k}+\mathbf{G}} \tag{17}$$

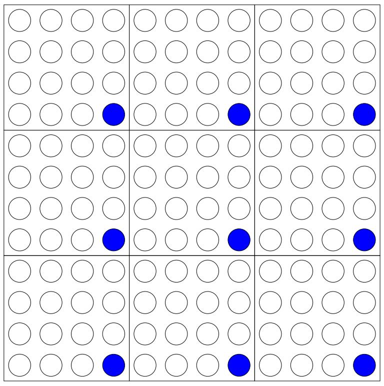

Figure 2: The introduction of a supercell to allow periodic boundary conditions to be applied to an aperiodic system

3 Density Matrix Formulation of DFT

In this section we develop the density matrix formulation of DFT from the Kohn-Sham formulation above, to obtain a method in which the total energy is minimised with respect to the single-particle density matrix, rather than the Kohn-Sham orbitals. This is very similar to the semi-empirical tight-binding method of Li, Nunes and Vanderbilt [16].

3.1 Partial Occupation of the Kohn-Sham Orbitals

We first generalise the theory outlined above so that the states $\{\psi_i\}$, rather than being doubly occupied, may now contain $2f_i$ electrons, where the occupation numbers $\{f_i\}$ lie in the range $0 \leq f_i \leq 1$. The electronic density $n(\mathbf{r})$ now becomes

$$n(\mathbf{r}) = 2 \sum_i f_i |\psi_i(\mathbf{r})|^2, \tag{18}$$

the kinetic energy becomes

$$E_{\mathrm{K}} = 2 \sum_i \int d^3r \ f_i \ \psi_i^*(\mathbf{r}) \left(-\tfrac{1}{2} \nabla^2 \right) \psi_i(\mathbf{r}), \tag{19}$$

and the pseudopotential energy (formerly $E_{\mathrm{e-ion}}$) is

$$E_{\mathrm{ps}} = 2 \sum_i \int d^3r \ f_i \ \psi_i^*(\mathbf{r}) \hat{V}_{\mathrm{ps}} \psi_i(\mathbf{r}). \tag{20}$$

The definitions for the remaining contributions to the total energy, E_{H} and E_{XC}, in terms of the electronic density remain unchanged. Allowing partial occupation cannot reduce the energy below that of the normal ground-state, and only increases the variational freedom, so that if the total energy is minimised both with respect to the $\{\psi_i\}$ (with the constraint of orthonormality) and with respect to the $\{f_i\}$ we arrive at the same ground-state that would be found by minimising with respect to fully occupied states $\{\psi_i\}$.

3.2 The Single-Particle Density Matrix

The single-particle density operator is defined by

$$\hat{\rho} = \sum_i f_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \tag{21}$$

which is therefore an Hermitian operator with eigenvalues in the interval [0,1] and in the coordinate representation this gives the density matrix $\rho(\mathbf{r}, \mathbf{r}')$ as

$$\rho(\mathbf{r}, \mathbf{r}') = \left\langle\mathbf{r}\right| \hat{\rho} \left|\mathbf{r}'\right\rangle = \sum_i f_i \ \psi_i(\mathbf{r})\psi_i^*(\mathbf{r}'). \tag{22}$$

Comparing this definition with the results given in the previous section, we define the electronic density, kinetic energy and pseudopotential energy as

$$\begin{aligned} n(\mathbf{r}) &= 2\rho(\mathbf{r}, \mathbf{r}) && (23) \\ E_{\mathrm{K}} &= -\int d^3r' \left[\nabla^2_{\mathbf{r}}\rho(\mathbf{r}, \mathbf{r}')\right]_{\mathbf{r}=\mathbf{r}'} && (24) \\ E_{\mathrm{ps}} &= 2\int d^3r \int d^3r' \ V_{\mathrm{ps}}(\mathbf{r}', \mathbf{r})\rho(\mathbf{r}, \mathbf{r}') && (25) \end{aligned}$$

with the Hartree energy E_{H} and exchange-correlation energy E_{XC} begin given in terms of the electronic density $n(\mathbf{r})$ as usual. $V_{\mathrm{ps}}(\mathbf{r}', \mathbf{r})$ is simply the pseudopotential operator in the coordinate representation:

$$V_{\mathrm{ps}}(\mathbf{r}', \mathbf{r}) = \left\langle\mathbf{r}'\right| \hat{V}_{\mathrm{ps}} \left|\mathbf{r}\right\rangle. \tag{26}$$

3.3 Constraints on the Density Matrix

3.3.1 Trace

In the coordinate representation, the trace of the density matrix is defined to be

$$\mathrm{Tr}(\rho(\mathbf{r}, \mathbf{r})) = \int d^3r \ \rho(\mathbf{r}, \mathbf{r}) = \tfrac{1}{2}\int d^3r \ n(\mathbf{r}) = \tfrac{1}{2}N_{\mathrm{e}} \tag{27}$$

where N_{e} is the number of electrons in the simulation cell. The physical significance of this constraint is obvious. However, we may prefer to make the Legendre transform to the grand canonical ensemble (at zero temperature), working at fixed chemical potential and variable electron number, and to minimise the grand potential $\Omega = E_{\mathrm{tot}} - \mu N_{\mathrm{e}}$ rather than the total energy. For insulators, it is sufficient that the chemical potential μ lie in the range $\varepsilon_{N_b} \leq \mu < \varepsilon_{N_b+1}$, where $N_b = \frac{1}{2}N_{\mathrm{e}}$ is the number of occupied bands. During the minimisation, various methods are used to keep the electron number close to its correct value, and these are described in section 4.5.

3.3.2 Idempotency

We have already noted that the eigenvalues of the density matrix should lie in the interval [0,1]. At self-consistency, all the eigenvalues should be either zero or unity, and this is a consequence of the property of *idempotency*, in which the density matrix obeys the equation $\rho^2 = \rho$. In this work, matrix products are defined as the continuum limit of matrix multiplication i.e.

$$(\mathrm{AB})(\mathbf{r},\mathbf{r}') = \int d^3r''\ \mathrm{A}(\mathbf{r},\mathbf{r}'')\mathrm{B}(\mathbf{r}'',\mathbf{r}')$$

. This constraint is hard to enforce, and two methods for doing this are described below.

3.3.3 McWeeny Purifying Transformation

Consider a matrix σ which is nearly idempotent (in a sense defined below) and define ρ in terms of this matrix as

$$\rho = 3\sigma^2 - 2\sigma^3. \tag{28}$$

Then

$$\rho^2 - \rho = 4(\sigma^2 - \sigma)^3 - 3(\sigma^2 - \sigma)^2 \tag{29}$$

so that ρ is therefore a more nearly idempotent matrix [17]. Equation 28 is known as a *purifying transformation*. From the definition of ρ it is clear that the diagonal frame of ρ is the same as the diagonal frame of σ and so the relationship between ρ and σ can be expressed as a relationship between their eigenvalues:

$$\lambda_\rho = 3\lambda_\sigma^2 - 2\lambda_\sigma^3 \tag{30}$$

The eigenvalues λ_ρ of ρ will satisfy $0 \le \lambda_\rho \le 1$ provided the eigenvalues λ_σ of σ are in the range $-\frac{1}{2} \le \lambda_\sigma \le \frac{3}{2}$. The function $\lambda_\rho(\lambda_\sigma)$ also has turning points at $\lambda_\rho = 0, 1$ and since the ground-state is obtained when $\lambda_\rho = 0, 1$ variation of σ will drive ρ towards idempotency.

We note however, that if any of the eigenvalues λ_σ fall out of the interval $[-\frac{1}{2}, \frac{3}{2}]$ at any stage during the minimisation process, then this transformation will in fact drive ρ away rather than toward idempotency. Since this formulation allows unphysical "run-away" solutions (in which the occupancies of states below the chemical potential $f_i \to \infty$ and occupancies of states above the chemical potential $f_i \to -\infty$) the ground-state being sought is in fact a metastable rather than a global minimum of the new functional so that this method of imposing the idempotency condition proves in practice to be rather unreliable.

3.3.4 Penalty Functional

Kohn [18] has suggested the use of the functional

$$P[\rho(\mathbf{r},\mathbf{r}')] = \left[\int d^3r\ \rho^2(1-\rho)^2 \Big|_{\mathbf{r}'=\mathbf{r}} \right]^{\frac{1}{2}} \tag{31}$$

as being an appropriate *penalty functional* for violating the idempotency condition. This can be seen by expressing the functional in terms of the eigenvalues λ_ρ of ρ

$$P[\rho(\mathbf{r},\mathbf{r}')] = \left[\sum_{\lambda_\rho} \lambda_\rho^2 (1-\lambda_\rho)^2 \right]^{\frac{1}{2}} \tag{32}$$

which shows that the functional vanishes only if all of the eigenvalues are zero or unity, and is otherwise positive.

Kohn has shown that for $\alpha > \alpha_c$, minimisation of the quantity

$$E_{\text{tot}} - \mu N_{\text{e}} + \alpha P[\rho(\mathbf{r}, \mathbf{r}')]$$

will converge to a value with $P[\rho(\mathbf{r}, \mathbf{r}')] = 0$ which therefore corresponds to the true ground-state energy, as long as $\text{Tr}(\rho) = \frac{1}{2} N_{\text{e}}$. He has also shown that

$$\alpha_c > 2 \left[\sum_{i, \varepsilon_i \leq \mu} (\varepsilon_i - \mu)^2 \right]^{\frac{1}{2}} \tag{33}$$

but it is not yet known whether this is a true lower bound. Naturally, this quantity cannot be calculated in advance, so it is necessary to increase α until the minimum is obtained with $P[\rho(\mathbf{r}, \mathbf{r}')] = 0$. In this case however, the ground-state is the global minimum of the functional, and so the problem with the purifying transformation mentioned above is eliminated.

3.4 Obtaining a Linear Scaling Method

So far we have shown that it is possible to reformulate DFT in terms of the single-particle density matrix, and that the electronic ground-state can be obtained by minimising a functional with respect to the elements of the density matrix (or, in the case of the McWeeny transformation, with respect to the elements of some auxiliary matrix). However, in order to obtain a tractable $O(N)$ method, it is necessary to impose some further restrictions on the form of the density matrix.

3.4.1 Separability

In practical calculations we cannot work with a six-dimensional function $\rho(\mathbf{r}, \mathbf{r}')$, even if it is localised as described below, and so it is essential that the density matrix is separable i.e.

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{\alpha\beta} \phi_\alpha(\mathbf{r}) K_{\alpha\beta} \phi_\beta(\mathbf{r}'). \tag{34}$$

We restrict ourselves to a finite number of functions ϕ_α which are referred to as *support functions.*

3.4.2 Spatial Localisation

In general, the density matrix in the ground-state $\rho(\mathbf{r}, \mathbf{r}')$ vanishes as the separation of its arguments $|\mathbf{r} - \mathbf{r}'|$ increases. The manner in which the density matrix vanishes depends upon the nature of the system being studied. It has been proved that in one-dimensional systems with a gap, $\rho(\mathbf{r}, \mathbf{r}')$ decays exponentially with separation, whereas in gapless systems it decays only algebraically. This suggests that we impose the restriction

$$\rho(\mathbf{r}, \mathbf{r}') = 0, \qquad |\mathbf{r} - \mathbf{r}'| > R_{\text{c}}. \tag{35}$$

The variational nature of DFT means that the minimum energy obtained $E_{\min}(R_c)$ will tend to the true ground-state energy E_0 from above as $R_c \to \infty$.

We impose the localisation of $\rho(\mathbf{r}, \mathbf{r}')$ by restricting the support functions to be non-zero only within finite regions called *support regions*, and by requiring the coefficients $K_{\alpha\beta}$ to vanish if the separation of the support regions of ϕ_α and ϕ_β exceeds a certain cut-off.

When using the McWeeny transformation, we write the auxiliary matrix as

$$\sigma(\mathbf{r}, \mathbf{r}') = \sum_{\alpha\beta} \phi_\alpha(\mathbf{r}) L_{\alpha\beta} \phi_\beta(\mathbf{r}') \tag{36}$$

where the coefficients $L_{\alpha\beta}$ vanish in the same way as $K_{\alpha\beta}$. This then implies the matrix relation

$$K = 3LSL - 2LSLSL \tag{37}$$

where $S_{\alpha\beta}$ is the overlap matrix for the support functions:

$$S_{\alpha\beta} = \int d^3r \ \phi_\alpha(\mathbf{r}) \phi_\beta(\mathbf{r}). \tag{38}$$

The localisation requirement means that all the matrices are sparse, and so matrix products can be calculated in $O(N)$ operations. The electronic density, total energy and Hamiltonian matrix elements can all be calculated in $O(N)$ operations so that a linear scaling method results. The memory requirement for support functions and matrices also scales linearly with system size.

4 Implementation on a Uniform Real-Space Grid

Hernández, Gillan and Goringe [19] have developed an implementation of the density matrix formulation using the McWeeny purifying transformation by discretising the support functions and Hamiltonian on a uniform grid in real space. The description below closely follows their work, and any differences are indicated.

4.1 Calculating the Total Energy and Hamiltonian

The support regions are chosen to be spherical and the support functions are represented by their values on the grid points lying within those support regions. In this work, the support regions were centred on the ionic positions, although the code allowed for arbitrary placement. Real-space integration becomes a summation over grid points, so that the overlap matrix elements are calculated as:

$$S_{\alpha\beta} \simeq \delta V \sum_{\mathbf{r}_g \in (\mathcal{R}_\alpha \cap \mathcal{R}_\beta)} \phi_\alpha(\mathbf{r}_g) \phi_\beta(\mathbf{r}_g) \tag{39}$$

in which $\mathcal{R}_\alpha$ denotes the support region for support function ϕ_α so that the sum runs over all grid points $\mathbf{r}_g$ common to the support regions of both ϕ_α and ϕ_β, and δV is the volume per grid point.

4.1.1 Kinetic energy

The action of the Laplacian operator on the support functions can be calculated using a finite difference technique. For a function $f(x)$ defined on a grid of spacing h, the second derivative to second order in the grid spacing is

$$\frac{d^2 f}{dx^2}(n_x) \simeq \frac{-f(n_x-2)+16f(n_x-1)-30f(n_x)+16f(n_x+1)-f(n_x+2)}{12h^2} \tag{40}$$

where n_x labels the grid point. This expression can be used to calculate $\partial^2\phi_\alpha/\partial x^2$, $\partial^2\phi_\alpha/\partial y^2$, $\partial^2\phi_\alpha/\partial z^2$ and therefore $\nabla^2\phi_\alpha$ approximately at each point. The kinetic energy is then given by

$$E_{\mathrm{K}} = 2\sum_{\alpha\beta} K_{\alpha\beta} T_{\beta\alpha} \tag{41}$$

where

$$T_{\beta\alpha} = -\tfrac{1}{2}\int d^3r\ \phi_\beta(\mathbf{r})\nabla^2_{\mathbf{r}}\phi_\alpha(\mathbf{r}) \tag{42}$$

and the integral is calculated by summing over grid points as for $S_{\alpha\beta}$. We have also calculated the kinetic energy by Fourier transforming each support function and then operating with the Laplacian, which is diagonal in reciprocal-space:

$$\nabla^2_{\mathbf{r}}\phi_\alpha(\mathbf{r}) = \mathcal{F}^{-1}\Big(-q^2\tilde{\phi}_\alpha(\mathbf{q})\Big) \tag{43}$$

where $\mathcal{F}^{-1}$ denotes taking the inverse Fourier transform. The matrix elements $T_{\alpha\beta}$ are then calculated by summing over grid points as before. The discrete Fourier transform is calculated using a fast Fourier transform (FFT) applied to the smallest rectangular grid which encloses the spherical support region.

4.1.2 Exchange-correlation energy and potential

The electronic density at grid point $\mathbf{r}_{\mathrm{g}}$ is:

$$n(\mathbf{r}_{\mathrm{g}}) = 2\sum_{\alpha\beta}\phi_\alpha(\mathbf{r}_{\mathrm{g}})K_{\alpha\beta}\phi_\beta(\mathbf{r}_{\mathrm{g}}). \tag{44}$$

The exchange-correlation energy E_{XC} can be calculated by summing $n(\mathbf{r}_{\mathrm{g}})\varepsilon_{\mathrm{XC}}[n(\mathbf{r}_{\mathrm{g}})]$ over grid points, and the exchange-correlation potential can be calculated at each grid point as

$$V_{\mathrm{XC}}(\mathbf{r}_{\mathrm{g}}) = \frac{\partial}{\partial n}\Big\{n(\mathbf{r}_{\mathrm{g}})\varepsilon_{\mathrm{XC}}[n(\mathbf{r}_{\mathrm{g}})]\Big\}. \tag{45}$$

4.1.3 Hartree energy and potential

The electronic density can be transformed into reciprocal-space using the FFT method to obtain its Fourier components $\tilde{n}(\mathbf{G})$.[1] The Hartree energy is then given by

$$E_{\rm H} = \frac{2\pi}{V} \sum_{\mathbf{G}\neq 0} \frac{|\tilde{n}(\mathbf{G})|^2}{G^2} \tag{46}$$

where V is the volume of the supercell and the $\mathbf{G} = 0$ term is omitted because the system is charge neutral overall and so this (infinite) term is cancelled by similar terms in $E_{\rm ion-ion}$ and $E_{\rm ps}$. The Hartree potential in reciprocal-space is

$$\tilde{V}_{\rm H}(\mathbf{G}) = \frac{4\pi}{V} \frac{\tilde{n}(\mathbf{G})}{G^2}. \tag{47}$$

4.1.4 Local pseudopotential

The local pseudopotential is also calculated in reciprocal-space as

$$\tilde{V}_{\rm ps,loc}(\mathbf{G}) = \sum_a \tilde{v}^a_{\rm ps,loc}(\mathbf{G}) S^a(\mathbf{G}) \tag{48}$$

where the summation is over ionic species a, $\tilde{v}^a_{\rm ps,loc}$ is the local pseudopotential for an isolated ion of species a in reciprocal-space and $S^a(\mathbf{G})$ is the structure factor for species a defined by

$$S^a(\mathbf{G}) = \sum_i \exp[-i\mathbf{G}\cdot\mathbf{R}^a_i] \tag{49}$$

where the sum is over all ions i of species a with positions $\mathbf{R}^a_i$. The pseudopotential energy can also be calculated in reciprocal-space as

$$E_{\rm ps} = \sum_{\mathbf{G}\neq 0} \sum_a \tilde{v}^a_{\rm ps,loc}(\mathbf{G}) S^a(\mathbf{G}) \tilde{n}^*(\mathbf{G}) + \sum_a N_a E^a_{\rm ps,core} \tag{50}$$

where $E^a_{\rm ps,core}$ is the pseudopotential core energy, and N_a the number of ions, of species a. The Hartree potential and pseudopotential can then be transformed back into real-space and added to the exchange-correlation potential to obtain the local part of the Kohn-Sham potential in real-space.

The FFT is not strictly an $O(N)$ operation but an $O(N \log_m N)$ operation (where m is some small number which depends upon the prime factors of the number of grid points) but since this is not likely to be the rate-limiting step in most calculations this is a minor detail.

4.1.5 Non-local pseudopotential

Non-local pseudopotentials were not used in this work, but since they are short-ranged in real-space [20], they could be included in an $O(N)$ scheme such as this by performing the

[1] The conventions used here for discrete Fourier transforms are

$$\tilde{n}(\mathbf{G}) = \sum_{\mathbf{r}} n(\mathbf{r}) \exp[-i\mathbf{G}\cdot\mathbf{r}] \quad ; \quad n(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{G}} \tilde{n}(\mathbf{G}) \exp[i\mathbf{G}\cdot\mathbf{r}]$$

projection onto spherical harmonics of only those support functions whose support regions overlap the pseudopotential core as summations over grid points in real-space, multiplying by the real-space non-local potential and then the appropriate spherical harmonic to obtain the non-local potential operating on the support functions. The matrix elements of the non-local potential can then be calculated, again by summations over grid points, and the non-local pseudopotential energy is given by twice the trace of the product of this matrix and the matrix K, the density matrix in the support function representation.

4.1.6 Electron Number N_e

Finally we note that

$$N_\mathrm{e} = \int d^3r\; n(\mathbf{r}) = 2\int d^3r\; \rho(\mathbf{r},\mathbf{r}) = 2\int d^3r\; \phi_\alpha(\mathbf{r}) K_{\alpha\beta} \phi_\beta(\mathbf{r}) = 2\mathrm{Tr}(KS). \tag{51}$$

4.2 Minimisation

Having calculated the total energy, both $L_{\alpha\beta}$ and ϕ_α need to be varied in order to minimise it. Since these are independent variables, the problem can be separated into two minimisations carried out in an alternating manner: varying ϕ_α at fixed $L_{\alpha\beta}$ and then varying $L_{\alpha\beta}$ at fixed ϕ_α. Since the object functions can be different for each minimisation, we choose here to minimise the total energy with respect to ϕ_α, and the grand potential with respect to $L_{\alpha\beta}$. Practical details concerning these minimisations follow in section 4.5. In the following sections, the summation convention is assumed.

4.3 Derivatives with respect to $L_{\alpha\beta}$

The total energy depends upon $L_{\alpha\beta}$ both explicitly and also through the electronic density $n(\mathbf{r})$. We use the result

$$\frac{\partial L_{ij}}{\partial L_{\alpha\beta}} = \delta_{i\alpha}\delta_{j\beta}. \tag{52}$$

4.3.1 Kinetic energy

From equations 41 and 37:

$$E_\mathrm{K} = 2T_{ji}(3L_{ik}S_{kl}L_{lj} - 2L_{ik}S_{kl}L_{lm}S_{mn}L_{nj}). \tag{53}$$

Explicitly:

$$\begin{aligned}\frac{\partial E_\mathrm{K}}{\partial L_{\alpha\beta}} &= 6[T_{j\alpha}S_{\beta l}L_{lj} + T_{\beta i}L_{ik}S_{k\alpha}] \\ &\qquad - 4[T_{j\alpha}S_{\beta l}L_{lm}S_{mn}L_{nj} + T_{ji}L_{ik}S_{k\alpha}S_{\beta n}L_{nj} + T_{\beta i}L_{ik}S_{kl}L_{lm}S_{m\alpha}] \\ &= [6(SLT+TLS) - 4(SLSLT + SLTLS + TLSLS)]_{\beta\alpha} \end{aligned} \tag{54}$$

4.3.2 Hartree energy

The Hartree energy depends upon $L_{\alpha\beta}$ through the electronic density $n(\mathbf{r})$:

$$E_{\mathrm{H}} = \tfrac{1}{2} \int d^3 r \int d^3 r' \, \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{55}$$

so

$$\frac{\partial E_{\mathrm{H}}}{\partial L_{\alpha\beta}} = \int d^3 r \, \frac{\delta E_{\mathrm{H}}}{\delta n(\mathbf{r})} \frac{\partial n(\mathbf{r})}{\partial L_{\alpha\beta}} \tag{56}$$

and the functional derivative of the Hartree energy with respect to the electronic density at point $\mathbf{r}$ is simply the Hartree potential at that point

$$\frac{\delta E_{\mathrm{H}}}{\delta n(\mathbf{r})} = V_{\mathrm{H}}(\mathbf{r}). \tag{57}$$

Since $n(\mathbf{r}) = 2\phi_\alpha(\mathbf{r}) K_{\alpha\beta} \phi_\beta(\mathbf{r})$, using equation 37 for $K_{\alpha\beta}$ again we obtain

$$\begin{aligned} \frac{\partial n(\mathbf{r})}{\partial L_{\alpha\beta}} &= 6[\phi_\alpha(\mathbf{r}) S_{\beta l} L_{lj} \phi_j(\mathbf{r}) + \phi_i(\mathbf{r}) L_{ik} S_{k\alpha} \phi_\beta(\mathbf{r})] \\ &- 4[\phi_\alpha(\mathbf{r}) S_{\beta l} L_{lm} S_{mn} L_{nj} \phi_j(\mathbf{r}) + \phi_i(\mathbf{r}) L_{ik} S_{k\alpha} S_{\beta n} L_{nj} \phi_j(\mathbf{r}) + \phi_i(\mathbf{r}) L_{ik} S_{kl} L_{lm} S_{m\alpha} \phi_\beta(\mathbf{r})]. \end{aligned} \tag{58}$$

We introduce the matrix element of the Hartree potential in the support function representation

$$V_{\mathrm{H},\alpha\beta} = \int d^3 r \, \phi_\alpha(\mathbf{r}) V_{\mathrm{H}}(\mathbf{r}) \phi_\beta(\mathbf{r}) \tag{59}$$

so that we can write

$$\begin{aligned} \frac{\partial E_{\mathrm{H}}}{\partial L_{\alpha\beta}} &= 6[V_{\mathrm{H},j\alpha} S_{\beta l} L_{lj} + V_{\mathrm{H},\beta i} L_{ik} S_{k\alpha}] \\ &\quad - 4[V_{\mathrm{H},j\alpha} S_{\beta l} L_{lm} S_{mn} L_{nj} + V_{\mathrm{H},\beta i} L_{ik} S_{k\alpha} S_{\beta n} L_{nj} + V_{\mathrm{H},\beta i} L_{ik} S_{kl} L_{lm} S_{m\alpha}] \\ &= [6(SLV_{\mathrm{H}} + V_{\mathrm{H}} LS) - 4(SLSLV_{\mathrm{H}} + SLV_{\mathrm{H}} LS + V_{\mathrm{H}} LSLS)]_{\beta\alpha}. \end{aligned} \tag{60}$$

4.3.3 Pseudopotential energy

The general non-local pseudopotential energy can be written in terms of the trace of the matrix K and the matrix elements of the pseudopotential operator in the support function representation:

$$V_{\mathrm{ps},\alpha\beta} = \int d^3 r \int d^3 r' \, \phi_\alpha(\mathbf{r}') \hat{V}_{\mathrm{ps}}(\mathbf{r}', \mathbf{r}) \phi_\beta(\mathbf{r}). \tag{61}$$

Thus $E_{\mathrm{ps}} = 2\mathrm{Tr}(KV_{\mathrm{ps}})$ as in the case of the kinetic energy, and so the derivative of the pseudopotential energy with respect to $L_{\alpha\beta}$ can be obtained from equation 54 simply by substitution:

$$\frac{\partial E_{\mathrm{ps}}}{\partial L_{\alpha\beta}} = [6(SLV_{\mathrm{ps}} + V_{\mathrm{ps}} LS) - 4(SLSLV_{\mathrm{ps}} + SLV_{\mathrm{ps}} LS + V_{\mathrm{ps}} LSLS)]_{\beta\alpha}. \tag{62}$$

4.3.4 Exchange-correlation energy

The derivative of the exchange-correlation energy is derived completely analogously to the derivative of the Hartree energy, and again the result can be obtained from equation 60 by substitution of the matrix elements of the exchange-correlation potential V_{XC}:

$$\frac{\partial E_{\mathrm{XC}}}{\partial L_{\alpha\beta}} = [6(SLV_{\mathrm{XC}} + V_{\mathrm{XC}}LS) - 4(SLSLV_{\mathrm{XC}} + SLV_{\mathrm{XC}}LS + V_{\mathrm{XC}}LSLS)]_{\beta\alpha}. \tag{63}$$

4.3.5 Total energy

Thus defining the matrix elements of the Kohn-Sham Hamiltonian in the representation of the support functions:

$$H_{\alpha\beta} = T_{\alpha\beta} + V_{\mathrm{H},\alpha\beta} + V_{\mathrm{ps},\alpha\beta} + V_{\mathrm{XC},\alpha\beta}. \tag{64}$$

we obtain the derivative of the total energy with respect to $L_{\alpha\beta}$ as:

$$\frac{\partial E_{\mathrm{tot}}}{\partial L_{\alpha\beta}} = [6(SLH + HLS) - 4(SLSLH + SLHLS + HLSLS)]_{\beta\alpha}. \tag{65}$$

4.3.6 Electron number

Since from equation 51 the electron number $N_{\mathrm{e}} = 2\mathrm{Tr}(KS)$, the derivative of the electron number with respect to $L_{\alpha\beta}$ can also be obtained from equation 54, substituting $S_{\alpha\beta}$ for $T_{\alpha\beta}$:

$$\frac{\partial N_{\mathrm{e}}}{\partial L_{\alpha\beta}} = 12(SLS - SLSLS)_{\beta\alpha}. \tag{66}$$

4.3.7 Grand potential

The grand potential $\Omega = E_{\mathrm{tot}} - \mu N_{\mathrm{e}}$ has derivatives $\frac{\partial\Omega}{\partial L_{\alpha\beta}} = \frac{\partial E_{\mathrm{tot}}}{\partial L_{\alpha\beta}} - \mu\frac{\partial N_{\mathrm{e}}}{\partial L_{\alpha\beta}}$, so that defining:

$$H'_{\alpha\beta} = H_{\alpha\beta} - \mu S_{\alpha\beta} \tag{67}$$

we obtain

$$\frac{\partial\Omega}{\partial L_{\alpha\beta}} = [6(SLH' + H'LS) - 4(SLSLH' + SLH'LS + H'LSLS)]_{\beta\alpha}. \tag{68}$$

4.4 Derivatives with respect to $\phi_\alpha(\mathbf{r}_{\mathrm{g}})$

In general, the quantities whose derivatives are required depend upon the support functions both explicitly through the electronic density, but also through the overlap matrix. We use the result

$$\frac{\partial\phi_i(\mathbf{r})}{\partial\phi_\alpha(\mathbf{r}_{\mathrm{g}})} = \delta_{i\alpha}\delta(\mathbf{r} - \mathbf{r}_{\mathrm{g}}). \tag{69}$$

Using this result in equation 38 we obtain

$$\frac{\partial S_{ij}}{\partial\phi_\alpha(\mathbf{r}_{\mathrm{g}})} = \delta_{i\alpha}\phi_j(\mathbf{r}_{\mathrm{g}}) + \phi_i(\mathbf{r}_{\mathrm{g}})\delta_{j\alpha}. \tag{70}$$

Then using this together with equation 37 for $K_{\alpha\beta}$:

$$\begin{aligned}\frac{\partial K_{ij}}{\partial \phi_\alpha(\mathbf{r}_\mathrm{g})} &= 3(L_{i\alpha}\phi_l(\mathbf{r}_\mathrm{g})L_{lj} + L_{ik}\phi_k(\mathbf{r}_\mathrm{g})L_{\alpha j}) - 2(L_{i\alpha}\phi_l(\mathbf{r}_\mathrm{g})L_{lm}S_{mn}L_{nj} \\ &+ L_{ik}\phi_k(\mathbf{r}_\mathrm{g})L_{\alpha m}S_{mn}L_{nj} + L_{ik}S_{kl}L_{l\alpha}\phi_n(\mathbf{r}_\mathrm{g})L_{nj} + L_{ik}S_{kl}L_{lm}\phi_m(\mathbf{r}_\mathrm{g})L_{\alpha j}).\end{aligned} \tag{71}$$

4.4.1 Kinetic energy

We define the Hermitian operator $\hat{T}$ as

$$\hat{T} = -\tfrac{1}{2}\nabla^2_{\mathbf{r}} \tag{72}$$

and in terms of this operator, the matrix elements $T_{\beta\alpha}$ are:

$$T_{\beta\alpha} = \int d^3r\ \phi_\beta(\mathbf{r})\hat{T}\phi_\alpha(\mathbf{r}). \tag{73}$$

Using the Hermiticity of $\hat{T}$ and equation 69:

$$\frac{\partial T_{ij}}{\partial \phi_\alpha(\mathbf{r}_\mathrm{g})} = \delta_{i\alpha}\hat{T}\phi_j(\mathbf{r}_\mathrm{g}) + \delta_{j\alpha}\hat{T}\phi_i(\mathbf{r}_\mathrm{g}). \tag{74}$$

Applying these results to equation 41:

$$\begin{aligned}\frac{\partial E_\mathrm{K}}{\partial \phi_\alpha(\mathbf{r}_\mathrm{g})} &= 2\frac{\partial}{\partial \phi_\alpha(\mathbf{r}_\mathrm{g})}(K_{ij}T_{ji}) = 2K_{ij}\frac{\partial T_{ji}}{\partial \phi_\alpha(\mathbf{r}_\mathrm{g})} + 2\frac{\partial K_{ij}}{\partial \phi_\alpha(\mathbf{r}_\mathrm{g})}T_{ji} \\ &= 4[K_{\alpha\beta}\hat{T} + (3LTL - 2LSLTL - 2LTLSL)_{\alpha\beta}]\phi_\beta(\mathbf{r}_\mathrm{g}).\end{aligned} \tag{75}$$

4.4.2 Hartree energy

The electronic density depends upon $\phi_\alpha(\mathbf{r})$ explicitly and through $K_{\alpha\beta}$ which depends upon the overlap matrix $S_{\alpha\beta}$:

$$\frac{\partial n(\mathbf{r})}{\partial \phi_\alpha(\mathbf{r}_\mathrm{g})} = 2\Big[K_{\alpha j}\delta(\mathbf{r}-\mathbf{r}_\mathrm{g})\phi_j(\mathbf{r}) + \phi_i(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}_\mathrm{g})K_{i\alpha} + \phi_i(\mathbf{r})\frac{\partial K_{ij}}{\partial \phi_\alpha(\mathbf{r}_\mathrm{g})}\phi_j(\mathbf{r})\Big]. \tag{76}$$

Thus

$$\begin{aligned}\frac{\partial E_\mathrm{H}}{\partial \phi_\alpha(\mathbf{r}_\mathrm{g})} &= \int d^3r\ \frac{\delta E_\mathrm{H}}{\delta n(\mathbf{r})}\frac{\partial n(\mathbf{r})}{\partial \phi_\alpha(\mathbf{r}_\mathrm{g})} = \int d^3r\ V_\mathrm{H}(\mathbf{r})\frac{\partial n(\mathbf{r})}{\partial \phi_\alpha(\mathbf{r}_\mathrm{g})} \\ &= 4[K_{\alpha\beta}\hat{V}_\mathrm{H} + (3LV_\mathrm{H}L - 2LSLV_\mathrm{H}L - 2LV_\mathrm{H}LSL)_{\alpha\beta}]\phi_\beta(\mathbf{r}_\mathrm{g}).\end{aligned} \tag{77}$$

4.4.3 Pseudopotential energy

As with the derivative of the pseudopotential energy with respect to $L_{\alpha\beta}$, the derivation below follows that of the kinetic energy, so we quote the result:

$$\frac{\partial E_\mathrm{ps}}{\partial \phi_\alpha(\mathbf{r}_\mathrm{g})} = 4[K_{\alpha\beta}\hat{V}_\mathrm{ps} + (3LV_\mathrm{ps}L - 2LSLV_\mathrm{ps}L - 2LV_\mathrm{ps}LSL)_{\alpha\beta}]\phi_\beta(\mathbf{r}_\mathrm{g}). \tag{78}$$

4.4.4 Exchange-correlation energy

The result below follows from the result for the Hartree energy.

$$\frac{\partial E_{\mathrm{XC}}}{\partial \phi_\alpha(\mathbf{r}_{\mathrm{g}})} = 4[K_{\alpha\beta}\hat{V}_{\mathrm{XC}} + (3LV_{\mathrm{XC}}L - 2LSLV_{\mathrm{XC}}L - 2LV_{\mathrm{XC}}LSL)_{\alpha\beta}]\phi_\beta(\mathbf{r}_{\mathrm{g}}) \tag{79}$$

4.4.5 Total energy

Summing up these results we arrive at the derivative of the total energy with respect to $\phi_\alpha(\mathbf{r})$:

$$\frac{\partial E_{\mathrm{tot}}}{\partial \phi_\alpha(\mathbf{r}_{\mathrm{g}})} = 4[K_{\alpha\beta}\hat{H} + (3LHL - 2LSLHL - 2LHLSL)_{\alpha\beta}]\phi_\beta(\mathbf{r}_{\mathrm{g}}) \tag{80}$$

where $\hat{H}$ is the Kohn-Sham Hamiltonian which operates on the support function ϕ_β.

4.4.6 Electron number

From equation 51 the following result is obtained by substitution from equation 75 and the last line by comparison with equation 66:

$$\begin{aligned}
\frac{\partial N_{\mathrm{e}}}{\partial \phi_\alpha(\mathbf{r}_{\mathrm{g}})} &= 4[K_{\alpha\beta} + (3LSL - 4LSLSL)_{\alpha\beta}]\phi_\beta(\mathbf{r}_{\mathrm{g}}) \\
&= 24(LSL - LSLSL)_{\alpha\beta}\phi_\beta(\mathbf{r}_{\mathrm{g}}) && (81) \\
&= 2\frac{\partial N_{\mathrm{e}}}{\partial L_{\alpha\beta}}\phi_\beta(\mathbf{r}_{\mathrm{g}}). && (82)
\end{aligned}$$

4.4.7 Grand potential

For completeness, we quote here the result for the derivative of the grand potential with respect to $\phi_\alpha(\mathbf{r})$:

$$\frac{\partial \Omega}{\partial \phi_\alpha(\mathbf{r}_{\mathrm{g}})} = 4[K_{\alpha\beta}\hat{H}' + (3LH'L - 2LSLH'L - 2LH'LSL)_{\alpha\beta}]\phi_\beta(\mathbf{r}_{\mathrm{g}}) \tag{83}$$

where the operator $\hat{H}'$ is defined in terms of the Kohn-Sham Hamiltonian $\hat{H}$ and the identity operator $\hat{I}$ as $\hat{H}' = \hat{H} - \mu\hat{I}$.

4.5 Practical minimisation

The practical problems of minimising a multi-dimensional energy functional [21] have already been addressed during the development of $O(N^3)$ codes and so we do not give a discussion here. Having obtained the steepest descent vectors for the appropriate functional, this vector can be conjugated [22] to previous search directions to minimise the functional more efficiently.

Since the electron number N_{e} will not in general remain constant during the minimisation, it is necessary to reconverge it to its correct value periodically during the minimisation. This can be achieved by displacing the auxiliary matrix $L_{\alpha\beta}$ in the direction of the electron number derivative, given by equation 66. Since the electron number is a cubic function in all search directions, this can be implemented without the need to iteratively converge because the cubic coefficients can be calculated from the traces of sparse matrix products, as outlined below.

We can also ensure that the electron number does not change to first order, by projecting the search direction perpendicular to the electron number gradient at the initial point. This was suggested by Hernández, Gillan and Goringe for the minimisation with respect to L. This can also be achieved with slightly more effort in the case of minimisation with respect to the support functions.

From equations 80 and 82 we can write

$$\begin{aligned} \frac{\partial E_{\rm tot}}{\partial \phi_\alpha(\mathbf{r}_{\rm g})} &= 4(K_{\alpha\beta}\hat{H} + (3LHL - 2LSLHL - 2LHLSL)_{\alpha\beta})\phi_\beta(\mathbf{r}_{\rm g}) \\ &= 4(K_{\alpha\beta}\hat{H} + M_{\alpha\beta})\phi_\beta(\mathbf{r}_{\rm g}), \quad (84) \\ \frac{\partial N_{\rm e}}{\partial \phi_\alpha(\mathbf{r}_{\rm g})} &= 24(LSL - LSLSL)_{\alpha\beta}\phi_\beta(\mathbf{r}_{\rm g}) = N_{\alpha\beta}\phi_\beta(\mathbf{r}_{\rm g}). \quad (85) \end{aligned}$$

The projection can be achieved simply by replacing $M_{\alpha\beta}$ by

$$M_{\alpha\beta} - \frac{\mathrm{Tr}(NHK + NSM)}{\mathrm{Tr}(NSN)} N_{\alpha\beta}. \quad (86)$$

We can also consider the electronic density and hence the Hamiltonian to be fixed during the minimisation with respect to $L_{\alpha\beta}$. In this case, we are no longer minimising $\Omega = E_{\rm tot} - \mu N_{\rm e}$ but $\Omega' = E' - \mu N_{\rm e}$, where $E' = 2\mathrm{Tr}(KH)$. This is equivalent to iteratively diagonalising H in the representation of the current support functions. The electronic density and Hamiltonian will naturally be updated at the start of the minimisation with respect to the support functions. This scheme has the added advantage that, like the electron number, this functional is also a cubic polynomial in all search directions, and so the exact location of line minima can be calculated.

Below we quote the results for the cubic coefficients of the electron number and grand potential when minimising with respect to $L_{\alpha\beta}$ in the search direction $\Delta_{\alpha\beta}$ i.e. when $L_{\alpha\beta} \to L_{\alpha\beta} + \delta\Delta_{\alpha\beta}$ then $N_{\rm e} \to N_{\rm e} + a_{N_{\rm e}}\delta + b_{N_{\rm e}}\delta^2 + c_{N_{\rm e}}\delta^3$ and $\Omega' \to \Omega' + a_{\Omega'}\delta + b_{\Omega'}\delta^2 + c_{\Omega'}\delta^3$ where the coefficients are given by:

$$\begin{aligned} a_{N_{\rm e}} &= 2\mathrm{Tr}[3(\Delta SLS + LS\Delta S) - 2(\Delta SLSLS + LS\Delta SLS + LSLS\Delta S)] \\ b_{N_{\rm e}} &= 2\mathrm{Tr}[3\Delta S\Delta S - 2(\Delta S\Delta SLS + \Delta SLS\Delta S + LS\Delta S\Delta S)] \\ c_{N_{\rm e}} &= -4\mathrm{Tr}[\Delta S\Delta S\Delta S] \quad (87) \end{aligned}$$

and

$$\begin{aligned} a_{\Omega'} &= 2\mathrm{Tr}[3(\Delta SLH' + LS\Delta H') - 2(\Delta SLSLH' + LS\Delta SLH' + LSLS\Delta H')] \\ b_{\Omega'} &= 2\mathrm{Tr}[3\Delta S\Delta H' - 2(\Delta S\Delta SLH' + \Delta SLS\Delta H' + LS\Delta S\Delta H')] \\ c_{\Omega'} &= -4\mathrm{Tr}[\Delta S\Delta S\Delta H']. \quad (88) \end{aligned}$$

4.6 Preliminary results

Ground-state pseudo-electronic densities are shown in figures 3 and 4 for 64 and 216 atoms of silicon respectively. The density generated by the plane-wave $O(N^3)$ CASTEP code is shown in figure 6 for comparison. We find in general that the linear-scaling code concentrates the density more in the bonding regions: this is reflected in the higher Hartree energy which is compensated for by a lower kinetic energy, which suggests that the real-space grid method may not be accurate enough for the kinetic energy.

Nevertheless, with a support region radius of 2.04Å and density matrix cut-off of 5.09Å we also reproduce the total ground-state energy to within 2% of the value calculated by the plane-wave $O(N^3)$ CASTEP code. Unfortunately, when comparing with such a result, it is difficult to determine how much of the discrepancy is due to the use of a real-space grid, and how much is the effect of the localisation restriction. In order to unravel these two effects, it is necessary to converge the total energy separately with respect to the grid spacing and the spatial cut-offs. We have analysed the accuracy of the kinetic energy using the FFT method described above, and the results suggest that a large fraction of the discrepancy could be due to the finite difference method underestimating the kinetic energy.

5 Linking $O(N)$ and $O(N^3)$ methods

$O(N)$ codes typically suffer from slow rates of convergence when compared to traditional $O(N^3)$ methods. This manifests itself in a large *cross-over* i.e. the number of atoms at which the $O(N)$ code reaches a specified level of convergence before the $O(N^3)$ code. In the density matrix formulation, the wave functions are replaced by the density matrix K and support functions, and it so it should be possible to make a reasonably accurate initial estimate of these quantities. In particular, if one could obtain some of these quantities from a converged $O(N^3)$ calculation on a smaller system, then the $O(N)$ calculation would initially start off closer to its fully converged value. To illustrate this idea, we consider two systems, where one might suppose that such a strategy would be worthwhile.

Consider a single vacancy in an otherwise ideal crystal. By running an $O(N^3)$ calculation on a small section of the bulk material, we could use this to estimate the initial density matrix and support functions for an $O(N)$ calculation on a large system containing a single vacancy. It would not then be unreasonable to assume that the majority of the large system would then start off quite close to its fully converged state and subsequently require fewer iterations to converge.

Secondly, we might consider a molecule interacting with the surface of a solid. We could separately converge the electronic density for the solid surface and for the molecule and use this information to generate an estimate for the complete interacting system. Obviously this method would be more successful in the case of a physisorped rather than a chemisorped molecule.

In section 5.2 we describe such a method for obtaining the density matrix from a fully converged plane-wave calculation. First, however, we consider estimating the density matrix from the approach of the linear combination of hybridised atomic orbitals, in order to provide a benchmark against which to compare other results.

5.1 Linear Combination of Hybridised Atomic Orbitals

We consider a bulk crystal such as silicon, with normalised pseudo-atomic orbitals $|s\rangle$, $|p_x\rangle$, $|p_y\rangle$, $|p_z\rangle$. In the coordinate representation, $\langle \mathbf{r}|s\rangle = R(r)$, a spherically symmetric function. We

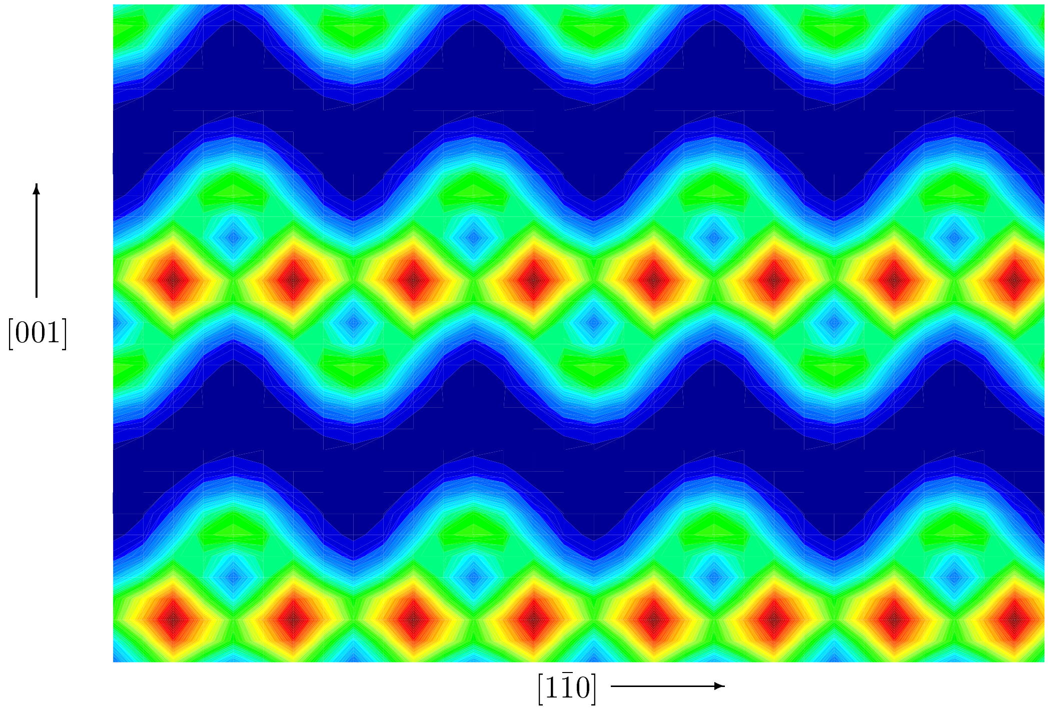

Figure 3: Pseudo-electronic density for 64-atom silicon unit cell in a (110) bonding plane

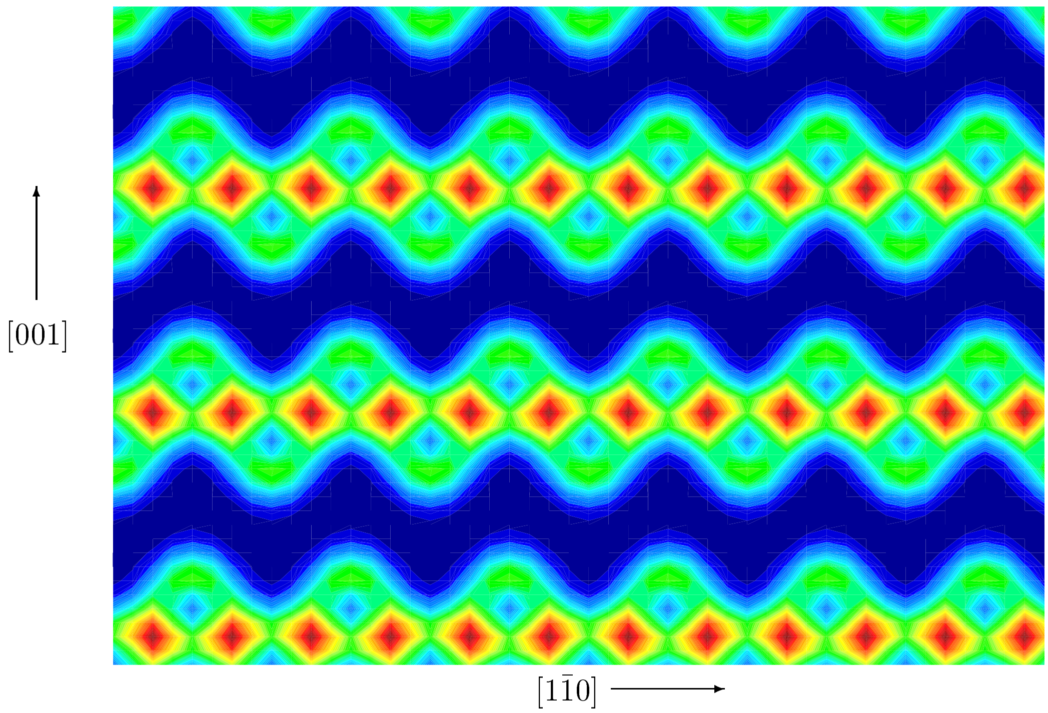

Figure 4: Pseudo-electronic density for 216-atom silicon unit cell in a (110) bonding plane

then assume a similar form for the remaining wave functions e.g. $\langle \mathbf{r}|p_x\rangle = NxR(r)$, where N is a constant to maintain normalisation.

We now consider hybridising these atomic orbitals by taking linear combinations of the form

$$\alpha\,|s\rangle + \beta\,|p_x\rangle + \gamma\,|p_y\rangle + \delta\,|p_z\rangle \tag{89}$$

which are normalised so that

$$\alpha^2 + \beta^2 + \gamma^2 + \delta^2 = 1. \tag{90}$$

The hybridised wave function is thus of the form

$$[\alpha + N(\beta x + \gamma y + \delta z)]R(r). \tag{91}$$

This wave function has a nodal plane given by

$$\alpha + N(\beta x + \gamma y + \delta z) = 0 \tag{92}$$

which by comparison with the vector equation for a plane $\mathbf{r}\cdot\mathbf{n} = d|\mathbf{n}|$ shows that the orbital points in the direction (β, γ, δ). We can thus make linear combinations of the orbitals on each atom to point in the directions of its nearest neighbours. Atoms at $(0,0,0)$ in the diamond structure motif have nearest neighbours in the $[111], [1\bar{1}\bar{1}], [\bar{1}\bar{1}1], [\bar{1}1\bar{1}]$ directions whereas atoms at $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ in the motif have nearest neighbours in the $[11\bar{1}], [1\bar{1}1], [\bar{1}11], [\bar{1}\bar{1}\bar{1}]$ directions.

These hybridised atomic orbitals are orthogonal to the other hybridised atomic orbitals on the same atom, and so we can construct a block diagonal unitary matrix to transform from the basis of atomic orbitals to that of the hybridised atomic orbitals.

We now take linear combinations of two of these hybrid atomic orbitals centred on neighbouring atoms to form bonding and anti-bonding molecular orbitals. Consider one hybrid atomic orbital $|1\rangle$ on atom 1 and another $|2\rangle$ on atom 2. Bonding $|b'\rangle$ and anti-bonding $|a'\rangle$ orbitals are defined by

$$|b'\rangle = |1\rangle + |2\rangle \tag{93}$$

$$|a'\rangle = |1\rangle - |2\rangle\,. \tag{94}$$

The inner product of these orbitals[2] is

$$\langle a'|b'\rangle = \langle 1|1\rangle - \langle 2|2\rangle - \langle 2|1\rangle + \langle 1|2\rangle = 2i\Im(\langle 1|2\rangle) \tag{95}$$

and this expression vanishes because the hybridised atomic orbitals are real. Defining $S_{12} = \Re(\langle 1|2\rangle)$ we can obtain the normalised orbitals:

$$|b\rangle = \frac{1}{\surd(2(1+S_{12}))}(|1\rangle + |2\rangle) \tag{96}$$

$$|a\rangle = \frac{1}{\surd(2(1-S_{12}))}(|1\rangle - |2\rangle). \tag{97}$$

These results can be used to construct another unitary matrix to transform between the hybridised atomic and molecular orbitals. In the representation of the molecular orbitals, we

[2] $\Re$ and $\Im$ denote taking the real and imaginary parts of a complex quantity respectively

take the density matrix to be diagonal, with fully occupied bonding orbitals and unoccupied anti-bonding orbitals. Using the unitary transformations derived above we can then transform this density matrix to the representation of the original atomic orbitals.

For simplicity we take the function $R(r)$ to be Gaussian, and having calculated S_{12} we construct the density matrix for these support functions. The resulting charge density is shown in figure 5. The support region radius was 2.35Å and the support functions were chosen to be highly localised.[3] The original method involved choosing the auxiliary matrix to be approximately the inverse of the overlap matrix (calculated according to $L = 2I - S \simeq [I - (I-S)]^{-1} = S^{-1}$ since $S \simeq I$ where I is the identity matrix) which effectively gives a diagonally dominant density matrix with all support functions half-occupied, and gave a total energy for the 8-atom unit cell of -75eV, whereas this method, using the new density matrix gave a total energy of -705eV which compares much better with the fully converged value of about -850eV. This suggests that this method provides a reasonable baseline to compare future results against.

5.2 Wave function projection

This method is based upon the work by Sanchez-Portal, Artacho and Soler [23] on the projection of plane-wave calculations into atomic orbitals. Here we review the method, for the case of a single **k**-point. The plane-wave eigenstates are denoted $|\psi_\alpha\rangle$ and we introduce a set of localised atomic orbitals (which correspond to support functions in this work) denoted ϕ_μ. The states obtained by projecting the plane-wave eigenstates onto the space spanned by the localised basis are denoted $|\chi_\alpha\rangle$. We also introduce the *dual states* $|\phi^\mu\rangle$ and $|\chi^\alpha\rangle$ with the properties outlined below:

$$\begin{aligned} S_{\mu\nu} &= \langle\phi_\mu|\phi_\nu\rangle &(98)\\ |\phi^\mu\rangle &= \sum_\nu S^{-1}_{\nu\mu}\,|\phi_\nu\rangle &(99)\\ \langle\phi^\mu|\phi_\nu\rangle &= \langle\phi_\mu|\phi^\nu\rangle = \delta_{\mu\nu} &(100)\end{aligned}$$

$$\begin{aligned} R_{\mu\nu} &= \langle\chi_\mu|\chi_\nu\rangle &(101)\\ |\chi^\mu\rangle &= \sum_\nu R^{-1}_{\nu\mu}\,|\chi_\nu\rangle &(102)\\ \langle\chi^\mu|\chi_\nu\rangle &= \langle\chi_\mu|\chi^\nu\rangle = \delta_{\mu\nu}. &(103)\end{aligned}$$

The projection operator onto the subspace spanned by the localised orbitals is defined by

$$\hat{P} = \sum_\mu |\phi_\mu\rangle\langle\phi^\mu| = \sum_{\mu\nu} |\phi_\mu\rangle S^{-1}_{\mu\nu}\langle\phi_\nu|\,. \tag{104}$$

A *spilling parameter* S can be defined to measure how much the subspace spanned by the plane-wave eigenstates falls outside the subspace spanned by the localised basis. Maximising this quantity gives a method for optimising the choice of support functions.

$$S = \frac{1}{N_\alpha}\sum_{\alpha=1}^{N_\alpha}\langle\psi_\alpha|(1-\hat{P})|\psi_\alpha\rangle \tag{105}$$

[3]The Gaussian standard deviation was 0.55Å

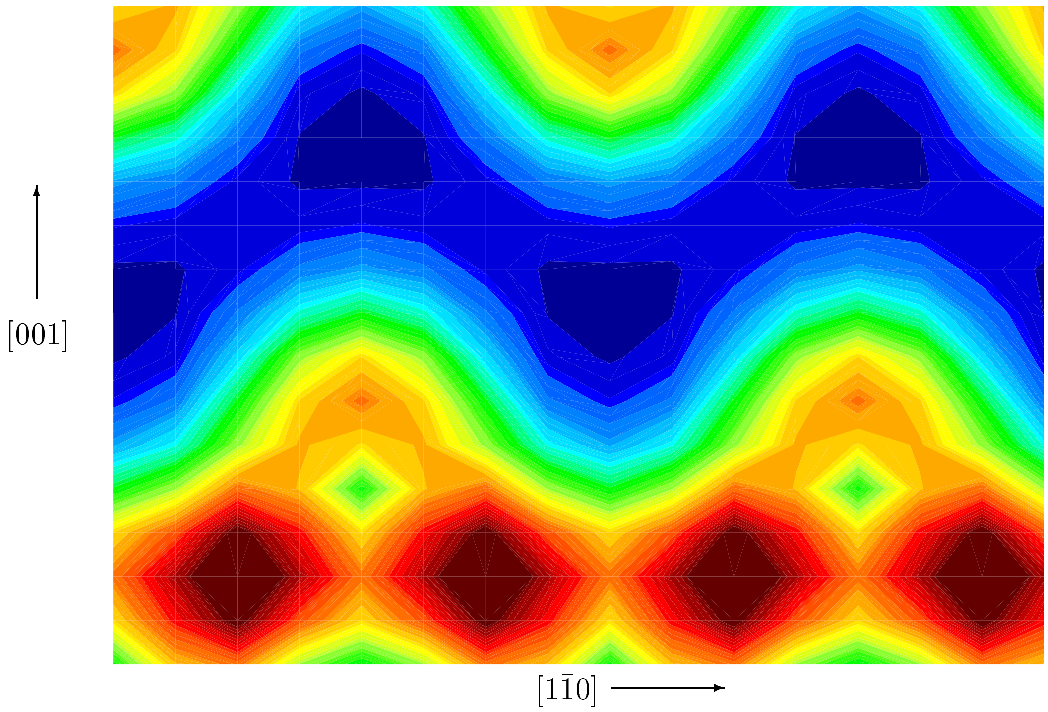

Figure 5: Initial pseudo-electronic density for 8-atom silicon unit cell in a (110) bonding plane with a density matrix calculated by the linear combination of hybridised atomic orbitals

where N_α is the number of bands considered. The density operator is then defined by

$$\hat{\rho} = \sum_{\alpha}^{\text{occ}} |\chi_\alpha\rangle \langle\chi^\alpha| \tag{106}$$

where the sum is taken over occupied bands only. Substitution of the results given above then yields the following expression for the density matrix in the representation of the localised functions:

$$K_{mn} = \langle\phi^m|\,\hat{\rho}\,|\phi^n\rangle = \sum_{\alpha\beta}^{\text{occ}} \sum_{\lambda\mu} S^{-1}_{m\lambda} \langle\phi_\lambda|\psi_\alpha\rangle R^{-1}_{\alpha\beta} \langle\psi_\beta|\phi_\mu\rangle S^{-1}_{\mu n} = (S^{-1}TR^{-1}T^\dagger S^{-1})_{mn}, \tag{107}$$

defining the rectangular matrix T as

$$T_{\lambda\alpha} = \langle\phi_\lambda|\psi_\alpha\rangle \tag{108}$$

Finally, we give an expression for the matrix R in terms of T and S

$$R = T^\dagger S^{-1} T. \tag{109}$$

The results obtained here were generated using a slightly adapted version of the code by Segall, Pickard and Shah [24]. Given the radial wave functions, this code calculates the overlap matrix and its inverse, as well as the density matrix K. To use this result with the purifying transformation, it is necessary to reverse that transformation to obtain the auxiliary matrix L (as was also the case for the linear combination of hybridised orbitals method) i.e. given K and S we need to solve the matrix equation 37, $K = 3LSL - 2LSLSL$ for the matrix L. We achieve this by post-multiplying the equation by the matrix S and then defining matrices $K' = KS$ and $L' = LS$ so that the equation to be solved becomes

$$K' = 3L'^2 - 2L'^3. \tag{110}$$

We then consider minimising the function $F(L')$ defined by

$$F(L') = \text{Tr}(L'^4 - 2L'^3 + 2K'L') \tag{111}$$

whose derivative with respect to L' is

$$\frac{\partial F(L')}{\partial L'_{ij}} = 2(2L'^3 - 3L'^2 + K')_{ji}. \tag{112}$$

This derivative vanishes at the minimum, and so we find that the matrix L' which minimises $F(L')$ also satisfies equation 110. We therefore choose to minimise $F(L')$ by the conjugate gradients method, to obtain L' from which we obtain L by post-multiplying by S^{-1} at no extra computational cost since this inverse was calculated by the projection code.

In figure 6 the pseudo-electronic density from a CASTEP calculation is shown, which yields a total energy for the 8-atom unit cell of -846.9eV. The wave functions generated by the same run were used in the projection process. The highly localised support functions used in the LCHAO work gave a very high spilling parameter, and so less localised support functions had to be used, which gave a spilling parameter of $S = 0.089$, which is still quite large. Using these less localised support functions[4] with the LCHAO density matrix did not significantly lower the total energy however. From the resulting electronic density in figure 7, we see that the method tends to smear out the density over the atomic sites, rather than in the bonding regions. The total energy was still a small improvement on the LCHAO value at -726eV. This work suggests that we need to minimise the spilling parameter with respect to the support functions, or use a more complete set of localised functions, which is followed up in the final section of this report.

6 New Localised Basis

In the previous sections we have noted some of the limitations of the real-space grid method, in which the basis consists of a uniform grid of discrete δ-functions. However, in section 2.5, we noted that one of the principal advantages of the pseudopotential approximation was that the weaker potential meant that the wave functions could be expanded in a much smaller set of plane-wave basis states, as one would expect from perturbation theory given that the plane-wave states are just the eigenstates for free electrons. In this section, we consider one candidate basis set which has the advantage of being similar to the plane-wave basis, but still retains the localisation properties of the previous basis set.

6.1 Formal Justification

In the absence of a potential, Schrödinger's time-independent equation simply reduces to the Helmholtz equation

$$(\nabla^2 + k^2)\psi(\mathbf{r}) = 0 \tag{113}$$

where $k^2 = 2E$. From the self-adjoint property of the Laplacian operator (given appropriate boundary conditions) we recognise this as a Sturm-Liouville eigenvalue problem.

Under periodic boundary conditions, we choose to expand the general solution in terms of discrete plane-wave eigenstates, which is efficient for describing the extended Kohn-Sham orbitals. In this case however, the boundary conditions we would like to choose are that the functions vanish outside a spherical region of radius R_{reg}. We make the conventional separation of variables and write the function as the product $\psi(\mathbf{r}) = R(r)Y_{lm}(\vartheta, \varphi)$ where Y_{lm} is a spherical harmonic (actually we choose the real linear combinations of the spherical harmonics, but this makes no difference to the following analysis) and $R(r)$ is some radial function which vanishes for $r \geq R_{\text{reg}}$ and obeys the equation

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR(r)}{dr}\right) - \frac{l(l+1)}{r^2}R(r) = -k^2R(r). \tag{114}$$

This is Bessel's equation, and the solutions are the spherical Bessel functions $j_l(kr)$ and $n_l(kr)$. The $n_l(kr)$ solutions are discarded because they are divergent at $r = 0$. Applying the boundary conditions leads to the condition that $j_l(kR_{\text{reg}}) = 0$ for all l. This condition only holds for certain discrete values of k, which we label k_n. Application of Sturm-Liouville theory therefore

[4] Gaussian standard deviation of 0.83Å

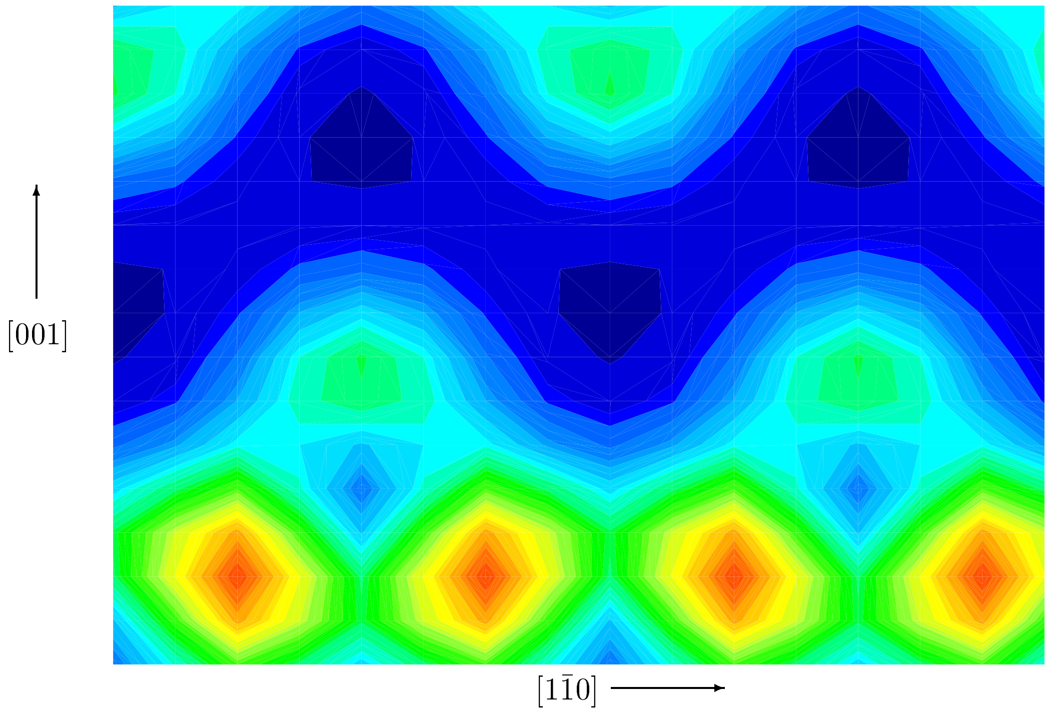

Figure 6: Pseudo-electronic density for 8-atom silicon unit cell in a (110) bonding plane generated by the CASTEP code

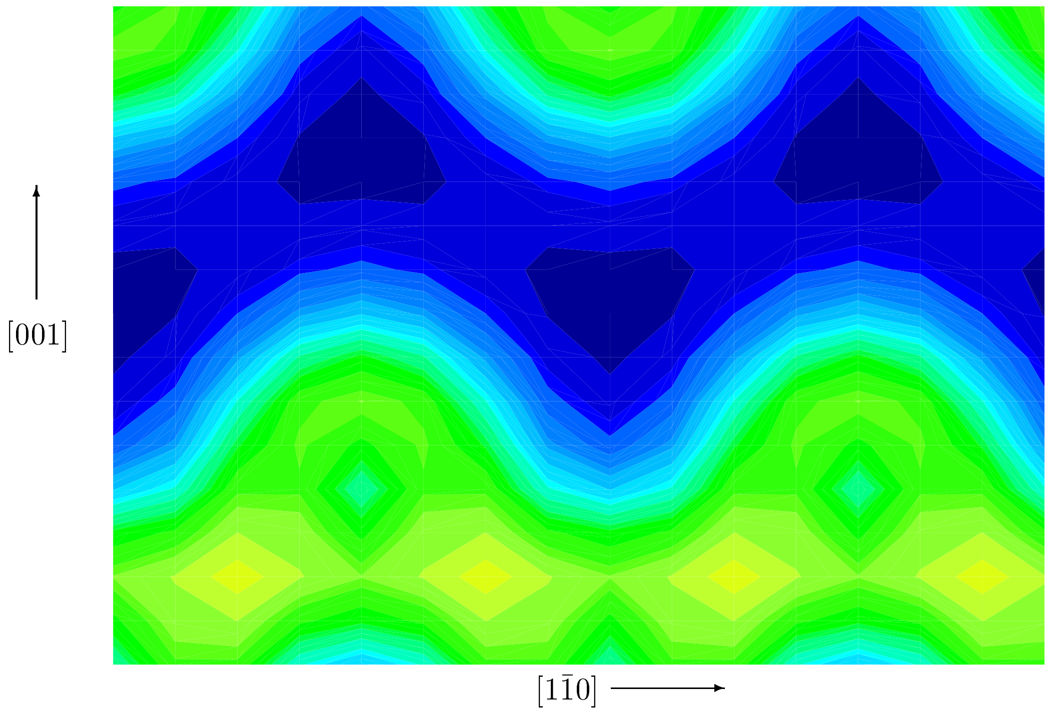

Figure 7: Pseudo-electronic density for 8-atom silicon unit cell in a (110) bonding plane generated by the CASTEP code and projected onto a localised basis

shows that we can expand any function which is non-zero only within a sphere of finite radius as

$$f(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \sum_{n} c_{nlm} j_l(k_n r) Y_{lm}(\vartheta, \varphi), \qquad r < R_{\text{reg}} \tag{115}$$

and this is a complete expansion, so that the coefficients c_{nlm} are unique. From equation 113 we can associate kinetic energy $\frac{1}{2}k_n^2$ with each state, and so as with the plane-wave basis, we can truncate the basis set according to some kinetic energy cut-off. We can also truncate the basis by considering only the lowest values of l, typically $l = 0, 1, 2$. The usual practice is to introduce a separate support function for each Y_{lm} included, and so instead of representing each support function by a value on $\sim 10^3$ grid points, we now represent them by $\sim 10^1$ coefficients.

6.2 Further development

The previous section demonstrates that the basis being considered is valid in the sense that it is complete, and can be truncated in a controlled manner. However, spherical Bessel functions are cumbersome to work with numerically, and so this basis set is only likely to be of practical use if some progress can be made analytically with some of the computation. In this section, we show that this is indeed the case.

In equation 38 we defined the overlap matrix, and we can rewrite this equation as

$$S_{\alpha\beta} = \int d^3 r_1 \int d^3 r_2 \; \phi_\alpha(\mathbf{R}_\alpha + \mathbf{r}_1) \phi_\beta(\mathbf{R}_\beta + \mathbf{r}_2) \delta(\mathbf{R}_\alpha + \mathbf{r}_1 - \mathbf{R}_\beta - \mathbf{r}_2) \tag{116}$$

where $\mathbf{R}_\alpha$ is the centre of support region $\mathcal{R}_\alpha$ of support function ϕ_α. This equation simplifies to

$$S_{\alpha\beta} = \int d^3 r \; \phi_\alpha(\mathbf{r}) \phi_\beta(\mathbf{R}_{\alpha\beta} - \mathbf{r}), \tag{117}$$

where $\mathbf{R}_{\alpha\beta} = \mathbf{R}_\beta - \mathbf{R}_\alpha$, which we recognise as a convolution, which can therefore be evaluated as the inverse Fourier transform of the product of the support function Fourier transforms.

The support function Fourier transform is defined by

$$\tilde{\phi}_\alpha(\mathbf{k}) = \int d^3 r \; \phi_\alpha(\mathbf{r}) \exp[i\mathbf{k} \cdot \mathbf{r}] \tag{118}$$

which, by equation 115 becomes the sum of terms such as

$$\int d^3 r \; j_l(k_n r) Y_{lm}(\vartheta, \varphi) \exp[i\mathbf{k}_n \cdot \mathbf{r}]$$

The angular part of this integral is a well-known result which can be obtained by partial differentiation of the spherically symmetric case ($l = m = 0$), and the result turns out to be related to spherical Bessel functions. The radial integral can then be performed using a generalised form of the orthogonality relationship for Bessel functions, so that we obtain the support function Fourier transforms in closed form.

We then multiply two of these Fourier transforms together, and calculate the inverse Fourier transform for the case of $\mathbf{r} = \mathbf{R}_{\alpha\beta}$

$$S_{\alpha\beta} = \frac{1}{(2\pi)^3} \int d^3 k \; \tilde{\phi}_\alpha(\mathbf{k}) \tilde{\phi}_\beta(\mathbf{k}) \exp[-i\mathbf{k} \cdot \mathbf{R}_{\alpha\beta}]. \tag{119}$$

The resulting integral can be separated into singular terms, which can then be regularised and calculated by the calculus of residues. There are several special cases to consider, and the results are quite involved, so we will not state them here.

The kinetic energy matrix elements can also be calculated in the same manner:

$$T_{\alpha\beta} = \frac{1}{(2\pi)^3} \int d^3k \; (-\tfrac{1}{2}k^2) \tilde{\phi}_\alpha(\mathbf{k}) \tilde{\phi}_\beta(\mathbf{k}) \exp[-i\mathbf{k} \cdot \mathbf{R}_{\alpha\beta}] \tag{120}$$

These integrals can be calculated in the same manner as the overlap matrix elements, and have also been obtained in closed form. This means that the kinetic energy can be calculated exactly which is a great improvement over the real-space grid method.

The electronic density can be constructed on a real-space grid efficiently by tabulating the spherical Bessel functions and other data. From this we can calculate the Hartree energy, pseudopotential energy and exchange-correlation energy. We can save some computational effort by adopting the neutral-atom potential of Sankey and Niklewski [25]. In this case we choose some reference electronic density $n_0(\mathbf{r})$ (which integrates to the correct number of electrons for the unit cell) which is typically concentrated around the atomic sites. We thus write

$$n(\mathbf{r}) = n_0(\mathbf{r}) + \delta n(\mathbf{r}). \tag{121}$$

Since the Hartree potential is linear in the electronic density, we have

$$V_{\mathrm{H}}(\mathbf{r}) = \int d^3r' \; \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} = \int d^3r' \; \frac{n_0(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \int d^3r' \; \frac{\delta n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} = V_{\mathrm{H}}^0(\mathbf{r}) + \delta V_{\mathrm{H}}(\mathbf{r}). \tag{122}$$

Given the reference density, which remains the same throughout the calculation, we can calculate matrix elements of $V_{\mathrm{H}}^0(\mathbf{r})$ and the local pseudopotential just once. These two terms both contain large contributions from the Coulomb tails, and these contributions cancel, so by adding these terms we obtain a weaker neutral-atom potential, whose matrix elements can also be calculated just once. Matrix elements for $\delta V_{\mathrm{H}}(\mathbf{r})$ and the exchange-correlation potential can be calculated by summing over grid points.

The final advantage of this basis set lies in calculating the non-local pseudopotential energy. The non-local components, like the support functions, are only non-zero within a spherical region with a given core radius. These components can be given as values on a radial grid, which can then be decomposed into spherical Bessel functions. The matrix elements of the non-local pseudopotential operator can then be calculated using the result for the overlap matrix elements.

The total energy and Hamiltonian matrix elements can therefore be calculated more cheaply using this basis set than with the real-space grid method, and the analytic results remove the need in many cases to deal directly with the spherical Bessel functions.

7 Conclusions and Further Work

In this report I have shown how it is possible to reformulate the Kohn-Sham equations in terms of the single-particle density matrix and thus to write an energy functional of that density matrix. This functional can therefore be minimised to obtain the self-consistent solutions to the Kohn-Sham equations. By imposing the restrictions of separability and localisation on the density matrix, an $O(N)$ method results. I have reviewed two methods of imposing the difficult idempotency constraint, and given details of a scheme to implement one of these methods on a real-space grid. Some of the problems associated with this scheme have been overcome by changing to a localised basis, and I have demonstrated how to calculated the total energy efficiently using this basis. Further work to be done would obviously include implementing this scheme fully, perhaps using the penalty functional suggested by Kohn instead of the McWeeny purifying transformation.

Another issue raised is that of linking linear scaling methods to the results of conventional $O(N^3)$ codes, and I have reviewed a method which allows this to be done. If the new localised basis scheme can be implemented successfully, further work would also need to be done evaluating this method, given that the new basis set is likely to span the subspace of the Kohn-Sham orbitals much more fully.

Acknowledgements

I would like to thank several members of the Theory of Condensed Matter Group for helpful discussions throughout the year, particularly Dr. Jim White for helpful discussions on direct minimisation techniques and penalty functions, Matthew Segall, Chris Pickard and Rajiv Shah for help with the wave function projection work, and Murray Jarvis and Duncan Kerr for other useful discussions. Finally, I would like to thank my supervisor, Dr. Mike Payne for suggesting this topic, particularly the projection technique, and giving me the freedom to pursue this work on my own. This work was supported by an EPSRC studentship.

References

[1] R. M. Dreizler and J. da Providencia, *Density Functional Methods in Physics* (Plenum, New York, 1985).

[2] R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).

[3] M. Born and R. Oppenheimer, Ann. Physik **84**, 457 (1927).

[4] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[5] T. A. Arias, M. C. Payne, and J. D. Joannopoulos, Phys. Rev. B **45**, 1538 (1992).

[6] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[7] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[8] O. Gunnarson and B. I. Lundqvist, Phys. Rev. B **13**, 4274 (1976).

[9] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).

[10] J. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[11] V. Heine, *Solid State Physics, Vol. 24* (Academic Press, New York, 1970), p. 1.

[12] N. W. Ashcroft and N. D. Mermin, in *Solid State Physics* (Saunders College, Philadelphia, 1976), Chap. 8, pp. 132–141.

[13] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[14] J. Ihm, A. Zunger, and M. L. Cohen, J. Phys. C: Solid State Physics **12**, 4409 (1979).

[15] P. J. H. Denteneer and W. van Haeringen, J. Phys. C: Solid State Physics **18**, 4127 (1985).

[16] X. P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10891 (1993).

[17] R. McWeeny, Rev. Mod. Phys. **32**, 335 (1960).

[18] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).

[19] E. Hernández, M. J. Gillan, and C. M. Goringe, Phys. Rev. B **53**, 7147 (1996).

[20] R. D. King-Smith, M. C. Payne, and J. S. Lin, Phys. Rev. B **44**, 13063 (1991).

[21] M. C. Payne *et al.*, Rev. Mod. Phys. **64**, 1045 (1992).

[22] M. P. Teter, M. C. Payne, and D. C. Allan, Phys. Rev. B **40**, 12255 (1989).

[23] D. Sanchez-Portal, E. Artacho, and J. M. Soler, Solid State Communications **95**, 685 (1995).

[24] M. D. Segall, C. J. Pickard, R. Shah, and M. C. Payne, To appear in *Mol. Phys.*

[25] O. F. Sankey and D. J. Niklewski, Phys. Rev. B **40**, 3979 (1989).