UNIVERSITY OF SOUTHAMPTON

Faculty Of Engineering, Science & Mathematics

School of Chemistry

Development of more accurate computational methods within linear-scaling density functional theory

by

Quintin Hill

Thesis for the degree of Doctor of Philosophy

December 2010

2

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS

SCHOOL OF CHEMISTRY

Doctor of Philosophy

DEVELOPMENT OF MORE ACCURATE COMPUTATIONAL METHODS WITHIN
LINEAR-SCALING DENSITY FUNCTIONAL THEORY

by Quintin Owen Hill

Kohn-Sham Density Functional Theory (DFT) provides a method for electronic structure calculations applicable to a wide variety of systems. Traditional implementations of DFT are cubic-scaling which limits the size of the systems that can be studied. However recently developed linear-scaling methods, such as ONETEP, are available which allow much larger systems to be considered.

Regardless of scaling DFT has limitations as the exact exchange-correlation functional (a key term in the Kohn-Sham equations) is not known and so approximations have to be made. These approximate functionals generally describe dispersion interactions poorly. In this thesis empirical corrections for dispersion have been developed with parameters optimised for a large set of dispersion bound complexes for the ONETEP code. This provides a much improved description of dispersion forces which are especially important for biological systems.

There is a hierarchy of exchange-correlation functionals available the most accurate of which include a portion of Hartree-Fock exchange. Methods for calculating Hartree-Fock exchange energy in ONETEP have been developed and are described in this thesis. A quadratic-scaling method using Fourier transforms has been implemented as a benchmark for other implementations. Hartree-Fock exchange may be calculated in a linear-scaling manner by using a numerical pointwise or auxiliary basis set method. Spherical waves have been used as an auxiliary basis set. Linear-scaling has been demonstrated for a polythene chain for these methods. Several hybrid functionals have also been implemented in ONETEP. These have been validated by comparison with a Gaussian basis set approach in calculations on the reaction paths of an organometallic system.

Contents

List of Tables

List of Figures

Declaration of authorship

I, Quintin Owen Hill, declare that the thesis entitled Development of more accurate computational methods within linear-scaling density functional theory and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;

- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

- where I have consulted the published work of others, this is always clearly attributed;

- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

- I have acknowledged all main sources of help;

- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

- parts of this work have been published as a paper [3].

Signed: ....
Date: March 9, 2011

I lift up my eyes to the mountains—
    where does my help come from?
My help comes from the LORD,
    the Maker of heaven and earth.

He will not let your foot slip—
    he who watches over you will not slumber;
indeed, he who watches over Israel
    will neither slumber nor sleep.

The LORD watches over you
    the LORD is your shade at your right hand;
the sun will not harm you by day,
    nor the moon by night.

The LORD will keep you from all harm—
    he will watch over your life;
the LORD will watch over your coming and going
    both now and forevermore.

<div align="right">Psalm 121 (NIV)</div>

Acknowledgements

I would like to express my gratitude to the following:

- Dr. Chris-Kriton Skylaris, my supervisor, for the support and encouragement he has provided to me.

- Professor John Dyke, my advisor, for providing advice when I have needed it.

- Dr. Andreas Danopoulos for his provision of and insights into the organometallic system studied in this work.

- The EPSRC, who provided the studentship that financially support the research presented in this thesis and also provided access to HECToR.

- iSolutions at Southampton University for maintaining the clusters that have been used for much of the computational work in this dissertation. The computational capacity provided by Iridis 3 has been of great value by allowing so many calculations to be completed quickly.

- The other postgraduate researchers in the group: Alvaro, Chris, Nikos and Stephen, for making the office a stimulating place to be.

- Dr Jacek Dziedzic, a post-doctoral fellow in the research group, for his help with optimising the code.

- My friends at Southampton University for making my time here enjoyable outside of studying.

- The people I've played sport alongside for providing this welcome break from work.

- Tamsin, for making the time spent writing up much more fun and enjoyable.

- My parents for loving and supporting me throughout my life thus far.

Chapter 1

Introduction

1.1 Quantum mechanics

Quantum mechanics is the set of laws governing the behaviour of very small particles including electrons and atomic nuclei. At this small scale the usual laws of classical mechanics fail to provide an accurate description of the behaviour of matter. For example, electrons cannot be fully described if treated as particles or waves, because they possess properties of both waves (for example diffraction) and particles (for example mass). In quantum mechanics it is not possible to simultaneously measure the momentum and position of an object exactly. The Heisenberg uncertainty principle [4] states that,

$$\Delta x \Delta p_x \geq \frac{\hbar}{2} \tag{1.1.1}$$

where Δx is the uncertainty in the position, Δp_x is the uncertainty in the momentum and $\hbar = \frac{h}{2\pi}$, h is Planck's constant. This non-deterministic description is another way that quantum mechanics differs from classical mechanics. Quantum particles can also have a spin, an intrinsic angular momentum, for example electrons have a spin of either $+\frac{1}{2}$ or $-\frac{1}{2}$.

1.1.1 The wavefunction

The wavefunction (Ψ, sometimes known as the state function) of a system is a function of each particle's coordinates (position and spin) and time; it provides a complete description of a quantum system. For a single particle system the probability density of that particle is defined as $|\Psi(\mathbf{r}, s, t)|^2$ (assuming a normalised wavefunction). Integrating this in a region of space gives the probability of the particle being in that region (at time t, with spin s). The probability density integrated over all space clearly has to equal one (as the particle must reside somewhere) [5]. Similarly the probability density for a system is defined as $|\Psi|^2$. A well-behaved wavefunction is quadratically integrable, i.e. the integral over all space of the probability density is finite. As electrons are fermions the wavefunction must also be antisymmetric with respect to the interchange of the coordinates $\mathbf{x}$ (where $\mathbf{x} = \mathbf{r}, s$) of any

two electrons [6],

$$\Psi(\mathbf{x}_1, \cdots, \mathbf{x}_i, \cdots, \mathbf{x}_j, \cdots) = -\Psi(\mathbf{x}_1, \cdots, \mathbf{x}_j, \cdots, \mathbf{x}_i, \cdots). \tag{1.1.2}$$

1.1.2   Operators

In quantum mechanics observable properties are defined by operators: every observable has a corresponding (quantum-mechanical) operator. For example to find the x coordinate of a particle the operator $\hat{x} = x\cdot$ would be used. However the operator for the momentum parallel to the x-axis is $\hat{p}_x = \frac{\hbar}{i}\frac{\partial}{\partial x}$. The result of a measurement of the observable O is one of the eigenvalues o_i of the operator $\hat{O}$. If the wavefunction (or state function) Ψ of the system is an eigenfunction of the operator $\hat{O}$ with eigenvalue o_j, then o_j is guaranteed to result from a measurement of O. However if Ψ is not an eigenfunction of $\hat{O}$ then any of the eigenvalues o_i may be obtained by a measurement of the system. What can be considered is this case is the expectation (or average) value of observable O given by,

$$\langle O \rangle = \int_V \Psi^* \hat{O} \Psi d\tau \tag{1.1.3}$$

where V denotes that the integral is over all (3N-dimensional) space and τ has the 3N-dimensional coordinates of the wavefunction [5].

The operators corresponding to physical observables both linear and Hermitian. A linear operator $\hat{A}$ satisfies the following properties,

$$\hat{A}[f(x) + g(x)] = \hat{A}f(x) + \hat{A}g(x) \tag{1.1.4}$$

$$\hat{A}[cf(x)] = c\hat{A}f(x) \tag{1.1.5}$$

where $f(x)$ and $g(x)$ are arbitrary functions and c is an arbitrary, potentially complex constant. Hermitian operators obey the following relation, obtained by the requirement that the expectation value of an observable be a real number,

$$\langle A \rangle = \langle A \rangle^* \tag{1.1.6}$$

$$\int_V \Psi^* \hat{A} \Psi d\tau = \int_V \Psi(\hat{A}\Psi)^* d\tau \tag{1.1.7}$$

for all well-behaved Ψ. It can be shown using equation 1.1.7 that a Hermitian operator also satisfies,

$$\int_V \Psi_i^* \hat{A} \Psi_j d\tau = \int_V \Psi_j(\hat{A}\Psi_i)^* d\tau \tag{1.1.8}$$

for all well-behaved Ψ_i and Ψ_j. Hermitian operators have some useful properties: all eigenvalues are real; eigenfunctions corresponding to different eigenvalues are orthogonal (and eigenfunctions with degenerate eigenvalues may be orthogonalised); the set of eigenfunctions form a complete (orthonormal) set [5].

1.1.3 The Schrödinger equation

The Schrödinger equation [7] for a system of nuclei and electrons provides the wavefunction of the system. For simplicity relativity effects are neglected in this work, however in heavy elements these effects are significant. The time-dependent Schrödinger equation is,

$$i\hbar\frac{\partial\Psi}{\partial t} = \hat{H}\Psi \qquad (1.1.9)$$

where Ψ is the wavefunction and $\hat{H}$ is the Hamiltonian or energy operator,

$$\hat{H} = \hat{T} + \hat{V} \qquad (1.1.10)$$

where $\hat{T}$ is the kinetic energy operator and $\hat{V}$ is the potential energy operator. The kinetic energy operator depends on the positions of the particles. The potential energy operator depends on the positions of the particles but can also be time-dependent, for example a charged particle in an oscillating electric field, would experience a time-dependent potential.

1.1.3.1 Time-independent Schrödinger equation

If the potential is time-independent then the time-dependent Schrödinger equation reduces to the time-independent Schrödinger equation. To show this consider a wavefunction, for a single particle, of the form $\Psi(\mathbf{r}, t) = f(t)\psi(\mathbf{r})$. Inserting this into the time-dependent Schrödinger equation (with a potential, $v(\mathbf{r})$, that is time-independent) gives,

$$i\hbar\frac{\partial\Psi}{\partial t} = \left(-\frac{\hbar^2}{2m}\hat{\nabla}^2 + \hat{V}(\mathbf{r})\right)\Psi(\mathbf{r}, t) \qquad (1.1.11)$$

$$i\hbar\frac{df(t)}{dt}\psi(\mathbf{r}) = -\frac{\hbar^2}{2m}f(t)\hat{\nabla}^2\psi(\mathbf{r}) + v(\mathbf{r})f(t)\psi(\mathbf{r}) \qquad (1.1.12)$$

dividing by $\Psi(\mathbf{r}, t)$ then gives,

$$i\hbar\frac{1}{f(t)}\frac{df(t)}{dt} = -\frac{\hbar^2}{2m}\frac{1}{\psi(\mathbf{r})}\hat{\nabla}^2\psi(\mathbf{r}) + v(\mathbf{r}). \qquad (1.1.13)$$

The left hand side of this equation depends only on t whereas the right hand side depends only on $\mathbf{r}$ therefore both sides must be constant (independent of $\mathbf{r}$ and t). Taking the left side first and equating it to a constant E gives,

$$i\hbar\frac{1}{f(t)}\frac{df(t)}{dt} = E \qquad (1.1.14)$$

$$\frac{df(t)}{f(t)} = -\frac{iE}{\hbar}dt \qquad (1.1.15)$$

$$\ln f(t) = -\frac{iEt}{\hbar} + C \qquad \text{by integration} \qquad (1.1.16)$$

$$f(t) = e^C e^{-\frac{iEt}{\hbar}} \qquad (1.1.17)$$

$$= Be^{-\frac{iEt}{\hbar}} \qquad (1.1.18)$$

where B (like C) is an arbitrary constant. B is usually omitted from $f(t)$ by including it as a factor in $\psi(\mathbf{r})$.

Consideration of the right side of equation 1.1.13 gives,

$$-\frac{\hbar^2}{2m}\frac{1}{\psi(\mathbf{r})}\hat{\nabla}^2\psi(\mathbf{r}) + v(\mathbf{r}) = E \tag{1.1.19}$$

$$-\frac{\hbar^2}{2m}\hat{\nabla}^2\psi(\mathbf{r}) + \hat{V}(\mathbf{r})\psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{1.1.20}$$

$$\Leftrightarrow \quad \hat{H}\psi = E\psi \tag{1.1.21}$$

which is the the time-independent Schrödinger equation [5]. $\psi(\mathbf{r})$ is an eigenfunction (or eigenvector) of the Hamiltonian and E is an eigenvalue. The energy eigenstates of the form $\Psi_n(\mathbf{r},t) = e^{-\frac{iEt}{\hbar}}\psi_n(\mathbf{r})$ are known as stationary states. Stationary states have the property that the physical properties of the system are constant in time. Consider the expectation value of the observable A at time t for the stationary state $\Psi_n(\mathbf{r},t)$,

$$\langle A\rangle_t = \int \Psi_n^*(\mathbf{r},t)\hat{A}\Psi_n(\mathbf{r},t)d\mathbf{r} \tag{1.1.22}$$

$$= \int e^{\frac{iEt}{\hbar}}\psi_n^*(\mathbf{r})\hat{A}e^{-\frac{iEt}{\hbar}}\psi_n(\mathbf{r})d\mathbf{r} \tag{1.1.23}$$

$$= \int \psi_n^*(\mathbf{r})\hat{A}\psi_n(\mathbf{r})d\mathbf{r} \tag{1.1.24}$$

which is time-independent if A is time-independent.

Above only separable wavefunctions were considered, but a linear superposition of stationary states is also a solution to the time-dependent Schrödinger equation. Such a wavefunction would have the form,

$$\Psi(\mathbf{r},t) = \sum_n c_n\psi_n(\mathbf{r})e^{-\frac{iE_nt}{\hbar}} \tag{1.1.25}$$

where c_n is an expansion coefficient. With a superposition of states $\langle A\rangle_t$ is, in general, time-dependent. This can be seen by considering the superposition of two stationary states,

$$\Psi(\mathbf{r},t) = c_1\psi_1(\mathbf{r})e^{-\frac{iE_1t}{\hbar}} + c_2\psi_2(\mathbf{r})e^{-\frac{iE_2t}{\hbar}} \tag{1.1.26}$$

where $|c_1|^2 + |c_2|^2 = 1$, and ψ_1 and ψ_2 are orthonormal, the expectation value of A is then,

$$\langle A\rangle_t = \int \Psi_n^*(\mathbf{r},t)\hat{A}\Psi_n(\mathbf{r},t)d\mathbf{r} \tag{1.1.27}$$

$$= |c_1|^2 \int \psi_1^*(\mathbf{r})\hat{A}\psi_1(\mathbf{r})d\mathbf{r} + c_1^*c_2 e^{\frac{i(E_1-E_2)t}{\hbar}} \int \psi_1^*(\mathbf{r})\hat{A}\psi_2(\mathbf{r})d\mathbf{r}$$

$$+ c_2^*c_1 e^{\frac{i(E_2-E_1)t}{\hbar}} \int \psi_2^*(\mathbf{r})\hat{A}\psi_1(\mathbf{r})d\mathbf{r} + |c_2|^2 \int \psi_2^*(\mathbf{r})\hat{A}\psi_2(\mathbf{r})d\mathbf{r} \tag{1.1.28}$$

since A is an observable $\hat{A}$ is a Hermitian operator and has a real expectation value so,

$$
\begin{aligned}
=&|c_1|^2 \int \psi_1^*(\mathbf{r})\hat{A}\psi_1(\mathbf{r})d\mathbf{r} + 2\Re\left[c_1^*c_2 e^{\frac{i(E_1-E_2)t}{\hbar}} \int \psi_1^*(\mathbf{r})\hat{A}\psi_2(\mathbf{r})d\mathbf{r}\right] \\
&+ |c_2|^2 \int \psi_2^*(\mathbf{r})\hat{A}\psi_2(\mathbf{r})d\mathbf{r}
\end{aligned}
\tag{1.1.29}
$$

which is clearly time dependent [8].

1.1.3.2 Molecular Schrödinger equation

The time-independent Schrödinger equation for a system of nuclei and electrons is,

$$
\hat{H}\Psi = \left(\hat{T}_e + \hat{T}_n + \hat{V}_{ee} + \hat{V}_{ne} + \hat{V}_{nn}\right)\Psi
\tag{1.1.30}
$$

$$
\begin{aligned}
=&\left[-\sum_{i=1}^{N_e}\frac{\hbar^2}{2m_i}\hat{\nabla}_{\mathbf{r}}^2 - \sum_{K=1}^{N_{\text{nuc}}}\frac{\hbar^2}{2M_K}\hat{\nabla}_{\mathbf{R}}^2 + \frac{1}{2}\sum_{i=1}^{N_e}\sum_{j\neq i}^{N_e}\frac{e^2}{4\pi\varepsilon_0 r_{ij}}\right. \\
&\left.-\sum_{i=1}^{N_e}\sum_{K=1}^{N_{\text{nuc}}}\frac{Z_K e^2}{4\pi\varepsilon_0 R_{iK}} + \frac{1}{2}\sum_{K=1}^{N_{\text{nuc}}}\sum_{P\neq K}^{N_{\text{nuc}}}\frac{Z_K Z_P e^2}{4\pi\varepsilon_0 R_{KP}}\right]\Psi
\end{aligned}
\tag{1.1.31}
$$

$$
=E\Psi\left(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_e}, \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_{\text{nuc}}}, s_1, s_2, \ldots, s_{N_e}, S_1, S_2, \cdots, S_{N_{\text{nuc}}}\right)
\tag{1.1.32}
$$

where r_{ij} is the distance between electron i and electron j, R_{iK} is the distance between electron i and nucleus K and R_{KP} is the distance between nucleus K and nucleus P; Z_n is the atomic number of nucleus n; N_{nuc} is the number of nuclei; N_e is the number of electrons; E is the total energy of the system.

The first pair of terms in the Hamiltonian operator are the kinetic energy operators for the electrons and nuclei respectively. The remaining three terms are the potential energy operators for the electron-electron repulsion, electron-nucleus attraction and nucleus-nucleus repulsion respectively. This equation is the foundation for electronic structure calculations.

1.1.4 Born-Oppenheimer approximation

In the Born-Oppenheimer approximation [9] the positions of the nuclei are regarded as fixed. This allows us to simplify the molecular Schrödinger equation. This approximation works well because the mass of a nucleus is at least three orders of magnitude larger than the mass of an electron (for the heaviest elements this difference approaches six orders of magnitude) and hence the electrons move much faster than the nuclei. When the Born-

Oppenheimer approximation is made, equation 1.1.32 simplifies to become,

$$\hat{H}\Phi = \left(\hat{T}_e + \hat{V}_{\text{ee}} + \hat{V}_{\text{ne}} + \hat{V}_{\text{nn}}\right)\Phi \tag{1.1.33}$$

$$= \left[-\sum_i^n \frac{\hbar^2}{2m_i}\hat{\nabla}_{\mathbf{r}}^2 + \frac{1}{2}\sum_i^n\sum_{j\neq i}^n \frac{e^2}{4\pi\varepsilon_0 r_{ij}} - \sum_i^n\sum_K^M \frac{Z_k e^2}{4\pi\varepsilon_0 R_{iK}} + \frac{1}{2}\sum_K^M\sum_{P\neq K}^M \frac{Z_K Z_P e^2}{4\pi\varepsilon_0 R_{KP}}\right]\Phi \tag{1.1.34}$$

$$= E\left(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M\right)\Phi\left(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n, s_1, s_2, \ldots, s_n; \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M\right). \tag{1.1.35}$$

The wavefunction and the Hamiltonian now have a parametric dependence on the positions of the nuclei while the energy, and also the last term in equation 1.1.34, are constants depending parametrically on the positions of the nuclei. The Born-Oppenheimer approximation is crucial as it is the basis of chemistry in terms of defining and understanding structures. This equation cannot be solved analytically for anything other than some simple one electron systems, such as the hydrogen atom. Various approximations have been developed through the years to allow solutions to the equation to be obtained.

1.1.5   Atomic units

In SI units the electronic Schrödinger equation for the hydrogen atom is,

$$\left(-\frac{\hbar^2}{2m_e}\nabla^2 - \frac{e^2}{4\pi\varepsilon_0 r}\right)\Phi = E\Phi. \tag{1.1.36}$$

If the Cartesian coordinates of the system are now transformed $x, y, z \rightarrow \lambda x', \lambda y', \lambda z'$, then equation 1.1.36 can be written,

$$\left(-\frac{\hbar^2}{2m_e\lambda^2}\nabla'^2 - \frac{e^2}{4\pi\varepsilon_0\lambda r'}\right)\Phi' = E\Phi'. \tag{1.1.37}$$

The constants in front of the kinetic energy and potential terms can be factored out with a suitable choice of λ such that,

$$\frac{\hbar^2}{m_e\lambda^2} = \frac{e^2}{4\pi\varepsilon_0\lambda} = E_h \tag{1.1.38}$$

where E_h is the atomic unit of energy the Hartree. The value of λ can be determined by solving equation 1.1.38 for λ,

$$\lambda = \frac{4\pi\varepsilon_0\lambda}{m_e e^2} = a_0 \tag{1.1.39}$$

where a_0 is the Bohr radius, and hence the atomic unit of length the Bohr. Equation 1.1.36 can now be rewritten in atomic units as,

$$E_h \left(-\frac{1}{2}\nabla'^2 - \frac{1}{r'} \right) \Phi' = E\Phi' \tag{1.1.40}$$

$$\left(-\frac{1}{2}\nabla'^2 - \frac{1}{r'} \right) \Phi' = E'\Phi' \tag{1.1.41}$$

by setting $E' = E/E_h$ [10]. Atomic units are used in all the subsequent expressions in this work.

1.1.6   Variational principle

The variational principle [5] states that a normalised approximate wavefunction satisfying the boundary conditions of the given problem gives an expectation value of the Hamiltonian that is an upper bound to the exact ground state energy. This can be expressed by,

$$\left\langle \tilde{\psi} \, \middle| \, \tilde{\psi} \right\rangle = 1 \Rightarrow \left\langle \tilde{\psi} \, \middle| \, \hat{H} \, \middle| \, \tilde{\psi} \right\rangle \geq \epsilon_0 \tag{1.1.42}$$

where $\tilde{\psi}$ is the trial wavefunction and ϵ_0 is the ground state energy (the lowest eigenvalue of the Hamiltonian). This can be proved by expanding the trial wavefunction in the orthonormal set of eigenfunctions ($\{|\psi_\alpha\rangle\}$ of the Hamiltonian. This is possible because the Hamiltonian is Hermitian and so its eigenfunctions form a complete set spanning the solution space.

$$\left\langle \tilde{\psi} \, \middle| \, \tilde{\psi} \right\rangle = 1 \tag{1.1.43}$$

$$= \sum_{\alpha\beta} \left\langle \tilde{\psi} \, \middle| \, \psi_\alpha \right\rangle \langle \psi_\alpha \, | \, \psi_\beta \rangle \left\langle \psi_\beta \, \middle| \, \tilde{\psi} \right\rangle \tag{1.1.44}$$

since the eigenfunctions are orthonormal,

$$= \sum_{\alpha\beta} \left\langle \tilde{\psi} \, \middle| \, \psi_\alpha \right\rangle \delta_{\alpha\beta} \left\langle \psi_\beta \, \middle| \, \tilde{\psi} \right\rangle \tag{1.1.45}$$

$$= \sum_{\alpha} \left\langle \tilde{\psi} \, \middle| \, \psi_\alpha \right\rangle \left\langle \psi_\alpha \, \middle| \, \tilde{\psi} \right\rangle \tag{1.1.46}$$

$$= \sum_{\alpha} \left| \left\langle \psi_\alpha \, \middle| \, \tilde{\psi} \right\rangle \right|^2 \tag{1.1.47}$$

Now expanding the expectation value of the Hamiltonian in a similar manner,

$$\left\langle \tilde{\psi} \,\middle|\, \hat{H} \,\middle|\, \tilde{\psi} \right\rangle = \sum_{\alpha\beta} \left\langle \tilde{\psi} \,\middle|\, \psi_\alpha \right\rangle \left\langle \psi_\alpha \,\middle|\, \hat{H} \,\middle|\, \psi_\beta \right\rangle \left\langle \psi_\beta \,\middle|\, \tilde{\psi} \right\rangle \tag{1.1.48}$$

$$= \sum_{\alpha\beta} \left\langle \tilde{\psi} \,\middle|\, \psi_\alpha \right\rangle \epsilon_\beta \delta_{\alpha\beta} \left\langle \psi_\beta \,\middle|\, \tilde{\psi} \right\rangle \tag{1.1.49}$$

$$= \sum_{\alpha} \epsilon_\alpha \left| \left\langle \psi_\alpha \,\middle|\, \tilde{\psi} \right\rangle \right|^2 \tag{1.1.50}$$

now since $\epsilon_\alpha \geq \epsilon_0$,

$$\left\langle \tilde{\psi} \,\middle|\, \hat{H} \,\middle|\, \tilde{\psi} \right\rangle \geq \sum_{\alpha} \epsilon_0 \left| \left\langle \psi_\alpha \,\middle|\, \tilde{\psi} \right\rangle \right|^2 \tag{1.1.51}$$

$$= \epsilon_0 \sum_{\alpha} \left| \left\langle \psi_\alpha \,\middle|\, \tilde{\psi} \right\rangle \right|^2 \tag{1.1.52}$$

$$= \epsilon_0 \tag{1.1.53}$$

using equation 1.1.47. Methods that obey the variation principle provide a systematic route for convergence to an estimate of the ground state wavefunction. This is done by varying the parameters of a normalised wavefunction (for example the coefficients of a linear expansion in a basis set) in order to minimise $\left\langle \tilde{\psi} \,\middle|\, \hat{H} \,\middle|\, \tilde{\psi} \right\rangle$ [10].

1.2 Hartree-Fock theory

The Hartree-Fock method is an approximate method for solving the molecular electronic Schrödinger equation. The computationally intractable many-electron problem is broken into a series of single electron problems in this approximation. In the Hartree-Fock method Ψ is expressed as an anti-symmetrized product of N (N is the number of electrons) orthonormal spin orbitals, known as a Slater determinant. A spin orbital is the product of a spatial orbital $\psi(\mathbf{r})$ and a spin function,

$$\chi(\mathbf{x}) = \begin{cases} \psi(\mathbf{r})\alpha(s) \\ \text{or} \\ \psi(\mathbf{r})\beta(s) \end{cases} \tag{1.2.1}$$

where $\mathbf{x} = \mathbf{r}, s$ represents both the spatial coordinates and the spin. This approximation for Ψ is expressed in the form,

$$\Psi_{\mathrm{HF}} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \cdots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \cdots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & \vdots & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix} \tag{1.2.2}$$

$$= \frac{1}{\sqrt{N!}} \det[\chi_1 \chi_2 \cdots \chi_N]. \tag{1.2.3}$$

The Hartree-Fock approximation provides a method to find (approximate) solutions for Ψ_0 (which are limited to be Slater determinants), the wavefunction that minimises the energy,

$$E_0 = \left\langle \Psi_0 \,\middle|\, \hat{H} \,\middle|\, \Psi_0 \right\rangle \tag{1.2.4}$$

where $\hat{H}$ is the full electronic Hamiltonian [6]. The variational theorem is used to obtain the optimum Slater determinant. In the Hartree-Fock method instead of solving the full multi-electron problem, a series of one-electron problems with electron-electron interaction is solved. This one-electron problem is expressed using the Fock operator,

$$\hat{f}_i = -\frac{1}{2}\hat{\nabla}_i^2 - \sum_{K=1}^{N_{\mathrm{nuc}}} \frac{Z_K}{R_{iK}} + \hat{v}_{\mathrm{HF},i} \tag{1.2.5}$$

where atomic units have been used and $\hat{v}_{\mathrm{HF},i}$ is the average potential felt by electron i due to the presence of the other electrons. This energy is normally expressed in terms of the Coulomb integrals,

$$J_{ab} = \int \int \chi_a^*(\mathbf{x}_1)\chi_b^*(\mathbf{x}_2) r_{12}^{-1} \chi_a(\mathbf{x}_1)\chi_b(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \tag{1.2.6}$$

$$= \langle ab \,|\, ab \rangle \tag{1.2.7}$$

and exchange integrals,

$$K_{ab} = \int \int \chi_a^*(\mathbf{x}_1)\chi_b^*(\mathbf{x}_2) r_{12}^{-1} \chi_b(\mathbf{x}_1)\chi_a(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \tag{1.2.8}$$

$$= \langle ab \,|\, ba \rangle. \tag{1.2.9}$$

The energy can now be expressed as,

$$E_0 = \sum_a^N \int \chi_a^*(\mathbf{x}) \left(-\frac{1}{2}\hat{\nabla}_a^2 - \sum_{K=1}^{N_{\mathrm{nuc}}} \frac{Z_K}{R_{aK}} \right) \chi_a(\mathbf{x}) d\mathbf{x} + \frac{1}{2}\sum_{ab}^N \left(\langle ab \,|\, ab \rangle - \langle ab \,|\, ba \rangle \right). \tag{1.2.10}$$

This expression can be simplified by introducing the core Hamiltonian operator defined as,

$$\hat{h}_i = -\frac{1}{2}\hat{\nabla}_i^2 - \sum_{K=1}^{N_{\text{nuc}}} \frac{Z_K}{R_{iK}}. \tag{1.2.11}$$

This simplifies the expression in equation 1.2.10 to,

$$E_0 = \sum_a^N \left\langle a \left| \hat{h} \right| a \right\rangle + \frac{1}{2} \sum_{ab}^N \left(\langle ab \,|\, ab \rangle - \langle ab \,|\, ba \rangle \right). \tag{1.2.12}$$

The Coulomb term in this expression is also known as the Hartree energy. The exchange term appears because of the use of antisymmetric products in the wavefunction.

The Fock operator for the one-electron problem in terms of spatial orbitals is,

$$\hat{f}_i(\mathbf{r}_i) = -\frac{1}{2}\hat{\nabla}_i^2 - \sum_{K=1}^{N_{\text{nuc}}} \frac{Z_K}{R_{iK}} + \sum_b^{N/2} \left(2\hat{J}_b(\mathbf{r}_i) - \hat{K}_b(\mathbf{r}_i) \right) \tag{1.2.13}$$

and the energy is,

$$E_0 = 2\sum_a^{N/2} \left\langle a \left| \hat{h} \right| a \right\rangle + \sum_{ab}^{N/2} \left(2J_{ab} - K_{ab} \right). \tag{1.2.14}$$

So the set of equations to be solved can be written,

$$\hat{f}_i(\mathbf{r}_i)\psi_i(\mathbf{r}_i) = \varepsilon_i \psi_i(\mathbf{r}_i). \tag{1.2.15}$$

These Hartree-Fock equations can be solved self-consistently with the use of a basis set [10].

1.2.1   Roothaan's equations

Roothaan's equations result from the expansion of the one-electron wavefunctions, ψ, in the Hartree Fock equations in terms of a basis set, $\{\phi_j\}$,

$$\psi_i(\mathbf{r}) = \sum_\mu^{N_{\text{BF}}} C_{\mu i}\phi_\mu(\mathbf{r}) \tag{1.2.16}$$

where N_{BF} is the number of basis functions, and $C_{\mu i}$ is an expansion coefficient. This expansion is only exact in general if $\{\phi_j\}$ is a complete set. However this would require N_{BF} to be infinite which is computationally impractical, so an incomplete finite basis set is used. The choice of basis set is important and this will be discussed in section 1.3 below. To obtain the molecular orbitals, the expansion coefficients now need to be calculated. A matrix equation for the expansion coefficients can be obtained by combining equations 1.2.15 and 1.2.16,

$$\hat{f}_i(\mathbf{r}_i)\sum_\nu C_{\nu i}\phi_\nu(\mathbf{r}_i) = \varepsilon_i \sum_\nu C_{\nu i}\phi_\nu(\mathbf{r}_i) \tag{1.2.17}$$

and then multiplying from the left by $\phi_\mu^*(\mathbf{r}_i)$ and integrating,

$$\sum_\nu C_{\nu i} \int \phi_\mu^*(\mathbf{r}_i) \hat{f}_i(\mathbf{r}_i) \phi_\nu(\mathbf{r}_i) d\mathbf{r}_i = \varepsilon_i \sum_\nu C_{\nu i} \int \phi_\mu^*(\mathbf{r}_i) \phi_\nu(\mathbf{r}_i) d\mathbf{r}_i. \tag{1.2.18}$$

From this expression it is natural to define the overlap, $\mathbf{S}$, and Fock, $\mathbf{F}$, matrices,

$$S_{\mu\nu} = \int \phi_\mu^*(\mathbf{r}_i) \phi_\nu(\mathbf{r}_i) d\mathbf{r}_i \tag{1.2.19}$$

$$F_{\mu\nu} = \int \phi_\mu^*(\mathbf{r}_i) \hat{f}_i(\mathbf{r}_i) \phi_\nu(\mathbf{r}_i) d\mathbf{r}_i. \tag{1.2.20}$$

With these definitions equation 1.2.18 can be simplified to give the Roothaan equations,

$$\sum_\nu F_{\mu\nu} C_{\nu i} = \varepsilon_i \sum_\nu S_{\mu\nu} C_{\nu i} \tag{1.2.21}$$

or in terms of matrices,

$$\mathbf{FC} = \mathbf{SC}\varepsilon \tag{1.2.22}$$

where ε is a diagonal matrix (with elements ε_i). Equation 1.2.22 can be solved to obtain the molecular orbitals and orbital energies. If an orthonormal basis set is used then $\mathbf{S}$ is the identity matrix and so the equation 1.2.22 reduces to,

$$\mathbf{FC} = \mathbf{C}\varepsilon \tag{1.2.23}$$

which can be solved by diagonalising $\mathbf{F}$. Since in general the basis sets used are nonorthogonal, the basis set is first orthogonalised [10].

The Fock matrix elements, defined in equation 1.2.20, also depend on the expansion coefficients; this can be seen by using the explicit form for the Fock operator given in equation 1.2.13,

$$F_{\mu\nu} = \int \phi_\mu^*(\mathbf{r}_i) \left(-\frac{1}{2}\hat{\nabla}_i^2 - \sum_{K=1}^{N_{\text{nuc}}} \frac{Z_K}{R_{iK}} + \sum_b^{N/2} \left(2\hat{J}_b(\mathbf{r}_i) - \hat{K}_b(\mathbf{r}_i) \right) \right) \phi_\nu(\mathbf{r}_i) d\mathbf{r}_i \tag{1.2.24}$$

$$= H_{\mu\nu}^{\text{core}} + G_{\mu\nu}. \tag{1.2.25}$$

The components of the core Hamiltonian (the kinetic and nuclear-electron attraction energy) have the simple matrix element form,

$$H_{\mu\nu}^{\text{core}} = \int \phi_\mu^*(\mathbf{r}_i) \hat{h}(\mathbf{r}_i) \phi_\nu(\mathbf{r}_i) d\mathbf{r}_i \tag{1.2.26}$$

$$= \int \phi_\mu^*(\mathbf{r}_i) \left(-\frac{1}{2}\hat{\nabla}_i^2 \right) \phi_\nu(\mathbf{r}_i) d\mathbf{r}_i + \int \phi_\mu^*(\mathbf{r}_i) \left(-\sum_{K=1}^{N_{\text{nuc}}} \frac{Z_K}{R_{iK}} \right) \phi_\nu(\mathbf{r}_i) d\mathbf{r}_i \tag{1.2.27}$$

$$= T_{\mu\nu} + V_{\mu\nu}^{\text{nuc}}. \tag{1.2.28}$$

The two electron integral part is more complicated,

$$G_{\mu\nu} = \int \phi_\mu^*(\mathbf{r}_i) \left(\sum_b^{N/2} \left(2\hat{J}_b(\mathbf{r}_i) - \hat{K}_b(\mathbf{r}_i) \right) \right) \phi_\nu(\mathbf{r}_i) d\mathbf{r}_i \qquad (1.2.29)$$

$$= \sum_b^{N/2} \sum_{\lambda\sigma} C_{\lambda b} C_{\sigma b}^* \left(2 \int \int \phi_\mu^*(\mathbf{r}_i) \phi_\nu(\mathbf{r}_i) r_{ij}^{-1} \phi_\sigma^*(\mathbf{r}_j) \phi_\lambda(\mathbf{r}_j) d\mathbf{r}_i d\mathbf{r}_j \right.$$
$$\left. - \int \int \phi_\mu^*(\mathbf{r}_i) \phi_\lambda(\mathbf{r}_i) r_{ij}^{-1} \phi_\sigma^*(\mathbf{r}_j) \phi_\nu(\mathbf{r}_j) d\mathbf{r}_i d\mathbf{r}_j \right). \qquad (1.2.30)$$

The dependence of $G_{\mu\nu}$, and therefore $F_{\mu\nu}$, on the expansion coefficients means that an iterative method is required to solve equation 1.2.22 [10].

1.2.2   Post Hartree-Fock methods

The difference between the exact Hartree-Fock energy (the Hartree-Fock limit, the result of a fully converged complete basis set calculation) and the exact ground state energy is known as the correlation energy. There are many methods which seek to improve the energy obtained by calculating some of the correlation energy [10]. These include methods such as configuration interaction and coupled cluster [11] where the wavefunction is expanded in Slater determinant space and approaches such as Møller-Plesset perturbation theory [12] for expanding the correlation energy in a perturbation series. Two commonly used methods are CCSD(T) [13] and MP2 [14].

1.3   Basis sets

Electronic structure theories, including Hartree-Fock, configuration interaction, coupled cluster and (Kohn-Sham) density functional theory, are all based on some form of one-electron wavefunction (a molecular orbital). In practical calculations a linear combination of basis functions is used to express each one-electron wavefunction, as discussed in section 1.2.1. The coefficients in the linear combination are then optimised, using linear algebra to solve the equations. To obtain the highest level of accuracy a basis set would have to be infinite in size, however this is clearly impractical. In Hartree-Fock theory (and its derivatives) large basis sets have a large computational cost as the number of two electron integrals that need to be calculated scales as N^4 where $N_{\rm BF}$ is the number of basis functions used [15]. Therefore compromises have to be made and basis sets are designed to give a good accuracy for a limited number of functions. It is also important that the form of the basis functions chosen are convenient to use in calculations. There are a number of different classes of basis function some of which will be described below.

1.3.1   Basis set superposition error

Basis set superposition error is an artificial increase in binding energy that arises due to the use of incomplete (atom-centred) basis sets. The binding energy between two molecules

may be calculated using,

$$\Delta E_{\text{bind}} = E^{a \cup b}(A - B) - E^a(A) - E^b(B) \qquad (1.3.1)$$

where $A - B$ is the complex of the two molecules and the superscripts a and b are the basis sets associated with molecules A and B respectively. If a and b are finite sets of atom-centred functions then basis set used in the energy calculation of the complex $a \cup b$ will be larger than a or b. Since the larger basis set provides more variational freedom the energy of the complex may be artificially lowered relative to the individual molecules [14].

The counterpoise correction is one method of correcting for this error. Applying one form of the correction [16, 17] to equation 1.3.1 gives,

$$\Delta E_{\text{bind}}^{\text{CP}} = E^{a \cup b}(A - B) - E^{a \cup b}(A) - E^{a \cup b}(B). \qquad (1.3.2)$$

In implementations of this form some of the basis functions are present without the corresponding atom. More complicated formulations of the counterpoise correction also exist [14, 17].

1.3.2 Slater-type orbitals

Slater-type orbitals [18] were commonly used as basis functions in earlier calculations. They have a form close to that of the atomic orbitals of the hydrogen atom. Slater orbitals have a cusp at the nucleus and have exponential decay of the form e^{-ar}. The main advantage of this is that, unlike Gaussian-type orbitals, they possess the correct radial form. The main disadvantage is that their use prevents the analytic solution of the Coulomb and exchange integrals, which have to be solved numerically. This makes them computationally inefficient and so generally restricts their use to small systems (for example diatomics). Also the long range behaviour is only correct if the smallest exponent is less than $\zeta_{\text{min}} = \sqrt{2I_{\text{min}}}$ where I_{min} is the lowest ionisation potential and atomic units are used, however smaller values than this are required for accurate work on molecules [19]. The Cartesian form for a Slater-type orbital is [20],

$$\phi_S(x, y, z; \zeta, i, j, k) = N x^i y^j z^k e^{-\zeta \sqrt{(x^2 + y^2 + z^2)}} \qquad (1.3.3)$$

where N is a normalisation constant.

1.3.3 Gaussian-type orbitals

Gaussian-type orbitals [21] were developed as a solution to the limitations of Slater functions. By choosing the orbitals to have the form of a Gaussian function it is possible to solve the Coulomb and exchange integrals analytically. A Gaussian-type orbital has the general form,

$$\phi_G(x, y, z; \alpha, i, j, k) = N x^i y^j z^k e^{-\alpha(x^2 + y^2 + z^2)} \qquad (1.3.4)$$

where α is an exponent controlling the width of the orbital and i, j and k are non negative integer parameters that determine the Cartesian nature of the orbital. The disadvantage is that the radial shape of a Gaussian-type orbital is not correct: at the origin a Gaussian function has a zero slope whereas a Slater function has a finite slope, producing a cusp at the nucleus. Also as the distance from the origin increases Gaussian functions decay far more quickly than Slater functions. This problem can be overcome by the use of contracted Gaussian functions where a linear combination of Gaussian-type orbitals is used to approximate the form of Slater-type orbitals [14].

1.3.4   Plane Waves

Plane waves are solutions to the Schrödinger equation for a particle in a periodic box (and also for non-interacting jellium). Hence they satisfy periodic boundary conditions. The form of a plane wave for a cubic box with side length l is,

$$\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{l^{\frac{3}{2}}} e^{i(k_x x + k_y y + k_z z)} \tag{1.3.5}$$

$$= \frac{1}{V^{\frac{1}{2}}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{1.3.6}$$

where $k_x = \frac{2\pi}{l} n_x$, $k_y = \frac{2\pi}{l} n_y$, $k_z = \frac{2\pi}{l} n_z$, and $n_x, n_y, n_z \in \mathbb{Z}$, and V is the volume of the box.

When plane waves are used in density functional calculations a much larger number (thousands or millions, depending on simulation cell size and the fineness of the grid) of plane wave basis functions are employed compared to the number of Gaussian-type orbitals used in a typical calculation (hundreds or thousands). The cost of evaluating the exchange integrals (see equation 1.2.8) in a wavefunction based calculation directly would be prohibitive; some solutions to this problem are discussed in section 4.1.2. Recently several post-Hartree-Fock methods have been implemented in a plane wave basis set code to study solids [22, 23]. One of the advantages of using a plane wave basis set is that the calculation does not suffer from basis set superposition error, since there is uniform coverage of the simulation cell by the basis functions. A disadvantage is that for calculations on isolated molecules the supercell approximation must be made, this entails using a large simulation cell to isolate the molecule from its periodic image. There is a trade off to be made however as in conventional codes this empty space is very computationally expensive. Another problem is that the connection with atomic orbitals is lost so it is harder to chemically interpret the final set of molecular orbitals [24].

1.3.5   Periodic Sinc Functions

Periodic sinc functions (psincs) are a new type of basis set that are used in the ONETEP linear-scaling code (discussed in section 2.2).

1.4   Periodic systems

Many systems of interest have translational symmetry for example bulk solids. To accurately model these periodic boundary conditions are required. Also as discussed above (section 1.3.4) the use of plane waves forces the use of periodic boundary conditions.

1.4.1   Cells

The cell is the periodically repeating unit in the system. In such systems there is a small primitive or unit cell which defines the symmetry of the system. For practical calculations it can be convenient to use a larger simulation cell containing multiple unit cells. A cell is defined by three primitive lattice vectors $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_3$. For a $1a_0 \times 2a_0 \times 3a_0$ orthorhombic unit cell the lattice vectors in Cartesian coordinates would be,

$$\mathbf{A}_1 = (1, 0, 0) \tag{1.4.1}$$

$$\mathbf{A}_2 = (0, 2, 0) \tag{1.4.2}$$

$$\mathbf{A}_3 = (0, 0, 3). \tag{1.4.3}$$

The volume of the cell is defined as,

$$V_c = \mathbf{A}_1 \cdot (\mathbf{A}_2 \times \mathbf{A}_3). \tag{1.4.4}$$

The right hand ordering of the lattice vectors is assumed such that a positive cell volume results [24].

1.4.2   Reciprocal lattice

The k points introduced in section 1.3.4 were restricted to a lattice of discrete points,

$$\mathbf{k} = \frac{2\pi}{l}\mathbf{n} \qquad \mathbf{n} \in \mathbb{Z}^3 \tag{1.4.5}$$

where l is the length of the cell. This lattice is known as the reciprocal lattice. The primitive reciprocal lattice vectors, $\mathbf{B}_i$, are defined by,

$$\mathbf{B}_1 = \frac{2\pi(\mathbf{A}_2 \times \mathbf{A}_3)}{\mathbf{A}_1 \cdot (\mathbf{A}_2 \times \mathbf{A}_3)} \tag{1.4.6}$$

$$= \frac{2\pi}{V_c}(\mathbf{A}_2 \times \mathbf{A}_3) \tag{1.4.7}$$

$$\mathbf{B}_2 = \frac{2\pi}{V_c}(\mathbf{A}_3 \times \mathbf{A}_1) \tag{1.4.8}$$

$$\mathbf{B}_3 = \frac{2\pi}{V_c}(\mathbf{A}_1 \times \mathbf{A}_2) \tag{1.4.9}$$

where $\mathbf{A}_i$ are the primitive direct lattice vectors. The primitive lattice vectors, $\mathbf{A}_i$, and the primitive reciprocal lattice vectors, $\mathbf{B}_i$, clearly satisfy the orthogonality relation,

$$\mathbf{B}_i \cdot \mathbf{A}_j = 2\pi\delta_{ij}. \tag{1.4.10}$$

Now consider two vectors $\mathbf{R}$ and $\mathbf{G}$, defined as integer sums of the primitive lattice and primitive reciprocal lattice vectors,

$$\mathbf{R} = \sum_{i=1}^{3} n_i \mathbf{A}_i \qquad n_i \in \mathbb{Z} \tag{1.4.11}$$

$$\mathbf{G} = \sum_{i=1}^{3} g_i \mathbf{B}_i \qquad g_i \in \mathbb{Z}. \tag{1.4.12}$$

The scalar product of $\mathbf{R}$ and $\mathbf{G}$ can be calculated using equation 1.4.10,

$$\mathbf{G} \cdot \mathbf{R} = 2\pi \sum_{i=1}^{3} g_i n_i \tag{1.4.13}$$

which is an integer multiple of 2π, hence,

$$e^{-i\mathbf{G} \cdot \mathbf{R}} = 1. \tag{1.4.14}$$

The first Brillouin zone is defined as the region of $\mathbf{k}$-space that is closer to the origin than any other lattice point [24].

1.4.3 Bloch's theorem

If the atoms in a system form a periodic structure (with lattice vectors $\mathbf{R}$) then the Hamiltonian for this system will also have this periodicity. In this case the effective potential will have the property that,

$$V_{\text{eff}}(\mathbf{r} + \mathbf{R}) = V_{\text{eff}}(\mathbf{r}) \tag{1.4.15}$$

where $\mathbf{R}$ is as defined in 1.4.11. The electron density is also periodic but the wavefunction of such a system is not [24]. Bloch's theorem arises from a consideration of the Schrödinger equation for a single-particle in such a periodic potential. It states that the solutions are of the form,

$$\psi_n(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}} u_n(\mathbf{r}) \tag{1.4.16}$$

$$\mathbf{k} = \sum_{i=1}^{3} c_i \mathbf{B}_i \qquad c_i \in \mathbb{R} \tag{1.4.17}$$

$$u_n(\mathbf{r}) = u_n(\mathbf{r} + \mathbf{R}). \tag{1.4.18}$$

These solutions are also eigenstates of the translation operator, $\hat{T}_{\mathbf{R}}$,

$$\hat{T}_{\mathbf{R}} \psi_n(\mathbf{r}) = \psi_n(\mathbf{r} + \mathbf{R}) \tag{1.4.19}$$

$$= e^{i\mathbf{k} \cdot (\mathbf{r} + \mathbf{R})} u_n(\mathbf{r} + \mathbf{R}) \tag{1.4.20}$$

$$= e^{i\mathbf{k} \cdot \mathbf{R}} \psi_n(\mathbf{r}). \tag{1.4.21}$$

It is appropriate to label these simultaneous eigenstates with an additional suffix specifying the wavevector $\mathbf{k}$,

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r}). \tag{1.4.22}$$

Consider the wavevector $\mathbf{k}' = \mathbf{k} + \mathbf{G}$ where $\mathbf{G}$ is as defined in equation 1.4.12. The Bloch state of $\mathbf{k}'$ is,

$$\psi_{n\mathbf{k}'}(\mathbf{r}) = e^{i\mathbf{k}'\cdot\mathbf{r}}u_{n\mathbf{k}'}(\mathbf{r}) \tag{1.4.23}$$

$$= e^{i\mathbf{k}\cdot\mathbf{r}}\left(e^{i\mathbf{G}\cdot\mathbf{r}}u_{n\mathbf{k}'}(\mathbf{r})\right) \tag{1.4.24}$$

$$= e^{i\mathbf{k}\cdot\mathbf{r}}u_{n'\mathbf{k}}(\mathbf{r}) \qquad \text{using equation 1.4.14} \tag{1.4.25}$$

$$= \psi_{n'\mathbf{k}}(\mathbf{r}). \tag{1.4.26}$$

This result shows that the infinite number of wavefunctions in a periodic system may be represented by a finite number of occupied wavefunctions at each $\mathbf{k}$ point located in the first Brillouin zone [24, 25].

1.4.4 Brillouin zone sampling

For an infinite system while there is a finite number of occupied eigenstates of the Hamiltonian for each $\mathbf{k}$ point there are an infinite number of $\mathbf{k}$ points inside the first Brillouin zone. In theory all of these eigenstates should be calculated; however in practice, only a finite number of $\mathbf{k}$ points need be considered, since the wavefunctions and eigenvalues of the Hamiltonian do not vary abruptly for small changes in $\mathbf{k}$ [26]. Various methods have been developed and are used in plane wave codes to accurately approximate integrals over the Brillouin zone by calculating the electronic states over special sets of $\mathbf{k}$ points. Examples of such sets include the Baldereschi (or mean-value) point [27] and the Monkhorst and Pack uniform set [28] with points defined by,

$$\mathbf{k}_{n_1,n_2,n_3} = \sum_{i=1}^{3} \frac{2n_i - N_i - 1}{2N_i}\mathbf{B}_i. \tag{1.4.27}$$

This formula is valid for any crystal [29]. Increasing the number of $\mathbf{k}$ points in the sample is a way to reduce errors, $\mathbf{k} \cdot \mathbf{p}$ perturbation theory can also be used [25]. In the Γ point approximation only the single point at $\mathbf{k} = \mathbf{0}$ is sampled.

1.4.4.1 Minimum image convention

If the periodic images are neglected while calculating a potential, a non-periodic potential results. In this case the form of the potential will depend on how the unit cell is defined relative to the atoms in the system. One method used to restore the periodicity of the potential is the minimum image convention. In this convention the interactions are between the closest periodic images [30]. Figure 1.1 illustrates this method.
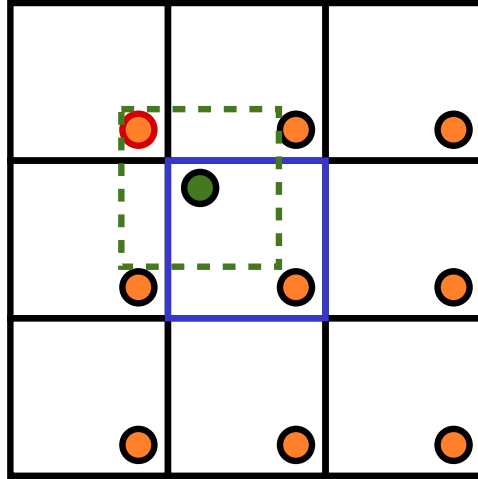
Figure 1.1: In the minimum image convention the interactions are between the nearest periodic neighbours. The green centre in the unit cell will interact with the periodic image in the upper left corner (the red ringed orange centre) as this is the closest. The green dotted square is the size of the cell but shifted so that it is centred on the green centre.

1.5 Density Functional Theory (DFT)

Density Functional Theory (DFT) takes a different approach to the solution of the Schrödinger equation using the density as the basic variable rather than the wavefunction. The advantage of using the density in preference to the wavefunction is that for a system with N electrons the wavefunction is a 3N-dimensional quantity whereas the density remains a 3-dimensional quantity. Instead of using operators to compute the properties of a system from the wavefunction, functionals of the density are used. A functional is a function that maps a function to a number. The origins of DFT can be traced back to the empirical Thomas-Fermi model [31, 32] which was developed a few years after the publication of the Schrödinger equation [7]. The Thomas-Fermi energy functional contains a kinetic energy term that is exact for the uniform non-interacting electron gas. Dirac augmented this model with an exchange energy term, leading to the Thomas-Fermi-Dirac energy functional [33]. A major shortcoming of these models is their failure to predict chemical bonds between atoms because of the inaccuracies in the kinetic energy functional, which made them of limited use to chemists [24]. Density functional theory was developed as a formally exact theory in the 1960s, with the publication of the Hohenberg and Kohn theorems.

1.5.1 The Hohenberg and Kohn theorems

HK Theorem 1 (Hohenberg and Kohn [34]) *The external potential $v_{\text{ext}}(\mathbf{r})$ is a unique functional of the density $n(\mathbf{r})$ (apart from a trivial additive constant).*

This can be proved by assuming that there exists an alternative potential $v'_{\text{ext}}(\mathbf{r})$ (with ground state wavefunction Ψ') that also produces $n(\mathbf{r})$. Ψ and Ψ' are distinct since:

$$v_{\text{ext}}(\mathbf{r}) - v'_{\text{ext}}(\mathbf{r}) \neq constant \Rightarrow \Psi' \neq \Psi. \tag{1.5.1}$$

(This implication arises because Ψ and Ψ' satisfy different Schrödinger equations, $\hat{H}\Psi = E_0\Psi$ and $\hat{H}'\Psi' = E_0'\Psi'$, and so will only be equal if the Hamiltonians only differ by a constant.) Then using the variational property of the ground state electronic energy,

$$E_0' = \langle\Psi'|\,\hat{H}'\,|\Psi'\rangle < \langle\Psi|\,\hat{H}'\,|\Psi\rangle = \langle\Psi|\,\hat{H}\,|\Psi\rangle + \langle\Psi|\,\hat{H}' - \hat{H}\,|\Psi\rangle \tag{1.5.2}$$

$$E_0' < E_0 + \int [v_{\text{ext}}'(\mathbf{r}) - v_{\text{ext}}(\mathbf{r})]n(\mathbf{r})d\mathbf{r}. \tag{1.5.3}$$

The primed and unprimed quantities in inequality 1.5.3 can be interchanged to obtain,

$$E_0 < E_0' - \int [v_{\text{ext}}'(\mathbf{r}) - v_{\text{ext}}(\mathbf{r})]n(\mathbf{r})d\mathbf{r}. \tag{1.5.4}$$

Adding these two inequalities together gives,

$$E_0' + E_0 < E_0 + E_0' \tag{1.5.5}$$

which is a contradiction and therefore the assumption that such a $v_{\text{ext}}'(\mathbf{r})$ exists is wrong. $n(\mathbf{r})$ therefore fixes $v_{\text{ext}}(\mathbf{r})$ which in turn fixes $\hat{H}$ [6,34].

Therefore the ground state wavefunction Ψ is a functional of $n(\mathbf{r})$ (since $n(\mathbf{r})$ fixes $\hat{H}$). Therefore kinetic energy and electronic repulsion energy are also functionals of the density. A universal functional of the density, the Hohenberg-Kohn functional, can be defined for these parts of the Hamiltonian,

$$F_{\text{HK}}[n(\mathbf{r})] = \langle\Psi|\,\hat{T}_e + \hat{V}_{\text{ee}}\,|\Psi\rangle \tag{1.5.6}$$

where $\hat{T}_e$ is the kinetic energy operator and $\hat{V}_{\text{ee}}$ is the electron-electron potential energy operator; both were defined in equation 1.1.34 . This functional is universal, it is valid for any number of particles and any external potential. However its explicit form is unknown. Using this functional the total energy functional can be defined as,

$$E_v[n(\mathbf{r})] \equiv \int v_{\text{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + F_{\text{HK}}[n(\mathbf{r})]. \tag{1.5.7}$$

HK Theorem 2 (Hohenberg and Kohn [34]) *The energy functional, $E_v[n(\mathbf{r})]$, has as its minimum the correct ground state energy associated with $v_{\text{ext}}(\mathbf{r})$ if the density is constrained to preserve the number of particles ($N[n(\mathbf{r})] = \int n(\mathbf{r})d\mathbf{r} = N_e$).*

The (electronic) energy functional of the density,

$$E_v[n'(\mathbf{r})] \equiv \int v_{\text{ext}}(\mathbf{r})n'(\mathbf{r})d\mathbf{r} + F_{\text{HK}}[n'(\mathbf{r})] \tag{1.5.8}$$

is by definition equal to the energy functional of the wavefunction,

$$\varepsilon_v[\Psi'] \equiv \langle\Psi'|\,V_{\text{ne}}\,|\Psi'\rangle + \langle\Psi'|\,\hat{T}_e + \hat{V}_{\text{ee}}\,|\Psi'\rangle. \tag{1.5.9}$$

It is clear that $\varepsilon_v[\Psi']$ attains as its minimum the ground state electronic energy, when Ψ' is the correct ground state ($= \Psi$) for our system of N particles. If Ψ' denotes the ground state wavefunction corresponding to external potential $v'(\mathbf{r}) \neq v(\mathbf{r})$ then,

$$\varepsilon_v[\Psi'] > \varepsilon_v[\Psi] = E_0 \tag{1.5.10}$$

and hence,

$$E_v[n'(\mathbf{r})] > E_v[n(\mathbf{r})] = E_0. \tag{1.5.11}$$

This proves that $E_v[n(\mathbf{r})] < E_v[n'(\mathbf{r})]$ for all $n'(\mathbf{r})$ associated with $v'_{\text{ext}}(\mathbf{r})$ and with $n(\mathbf{r})$ being the exact density for the ground state. This theorem provides a variational principle for the ground state in terms of the density [34]. This allows the density to be used as the basic variable in quantum chemical calculations.

1.5.2 Kohn-Sham theory

Kohn and Sham [35] further developed density functional theory by providing a form for the Hohenberg and Kohn functional that avoided the use of an explicit kinetic energy functional. Instead they used the non-interacting particle kinetic energy, whose expression is known in terms of orbitals from quantum mechanics. Hence orbitals were introduced. Starting from the definition of the Hohenberg and Kohn functional, Kohn and Sham modified this by repartitioning the terms,

$$F_{\text{HK}}[n] = T[n] + V_{\text{ee}}[n] \tag{1.5.12}$$
$$= T[n] + T_{\text{s}}[n] - T_{\text{s}}[n] + E_{\text{H}}[n] - E_{\text{H}}[n] + V_{\text{ee}}[n] \tag{1.5.13}$$
$$= T_{\text{s}}[n] + E_{\text{H}}[n] + (T[n] - T_{\text{s}}[n] + V_{\text{ee}}[n] - E_{\text{H}}[n]) \tag{1.5.14}$$
$$= T_{\text{s}}[n] + E_{\text{H}}[n] + E_{\text{xc}}[n] \tag{1.5.15}$$

where $E_{\text{H}}[n]$ is the Hartree energy (see equation 1.2.6) and $T_{\text{s}}[n]$ is the kinetic energy of a non-interacting system with density $n(\mathbf{r})$, the value of both of these terms can be calculated exactly. $E_{\text{xc}}[n]$ is defined as the exchange correlation energy and kinetic energy correction for the interacting system with density $n(\mathbf{r})$. Unfortunately a general exact explicit form for this term is unknown. The ground state density can be found by minimising the energy functional, that is setting its functional derivative to zero with the inclusion of the Lagrange multiplier μ to fix the number of electrons,

$$\frac{\delta E[n]}{\delta n} = \frac{\delta T_{\text{s}}}{\delta n} + v_{\text{H}}(\mathbf{r}) + v_{\text{xc}}(\mathbf{r}) + v_{\text{ext}}(\mathbf{r}) = \mu \tag{1.5.16}$$

$$\text{where} \quad v_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[n]}{\delta n} \tag{1.5.17}$$

and $v_{\mathrm{H}}(\mathbf{r})$ is the Hartree potential. If this is rewritten by collecting the potential terms,

$$\frac{\delta T_{\mathrm{s}}}{\delta n} + v_{\mathrm{eff}}(\mathbf{r}) = \mu \tag{1.5.18}$$

$$\text{where} \qquad v_{\mathrm{eff}}(\mathbf{r}) = v_{\mathrm{H}}(\mathbf{r}) + v_{\mathrm{xc}}(\mathbf{r}) + v_{\mathrm{ext}}(\mathbf{r}) \tag{1.5.19}$$

this is exactly the same as the equation for the ground state density of a non-interacting system with external potential v_{eff}. The density of this non-interacting system is also $n(\mathbf{r})$. A non-interacting system of electrons in an external potential v_{eff} is described by a set of single particle Schrödinger equations,

$$\left(-\frac{1}{2}\hat{\nabla}^2 + \hat{v}_{\mathrm{eff}}(\mathbf{r}) \right) \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}) \tag{1.5.20}$$

which can be solved (on a computer) as each involves only a single electron. The eigen-functions of equation 1.5.20 can be used to calculate the density. From the solution it is clear that $T_{\mathrm{s}}[n]$ can be computed by introducing Kohn-Sham orbitals, $\{\psi_i\}$,

$$T_{\mathrm{s}}[n] = \sum_i^N \int \psi_i^*(\mathbf{r}) \left(-\frac{1}{2}\nabla^2 \right) \psi_i(\mathbf{r})d\mathbf{r} \tag{1.5.21}$$

where the density is defined as

$$n(\mathbf{r}) = \sum_i^N |\psi_i(\mathbf{r})|^2. \tag{1.5.22}$$

So the energy functional can therefore be rewritten as,

$$E[n] = T_{\mathrm{s}}[n] + E_{\mathrm{H}}[n] + E_{\mathrm{xc}}[n] + \int v_{\mathrm{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} \tag{1.5.23}$$

$$= \sum_i^N \int \psi_i^*(\mathbf{r}) \left(-\frac{1}{2}\nabla^2 \right) \psi_i(\mathbf{r})d\mathbf{r} + E_{\mathrm{H}}[n] + E_{\mathrm{xc}}[n] + \int v_{\mathrm{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r}. \tag{1.5.24}$$

The goal is to minimise the value of the energy functional, however there is a need to constrain the orbitals to be orthonormal. So a functional that includes this constraint is defined and this functional is minimised instead,

$$\Omega[\{\psi_i\}] = E[n] - \sum_i^N \sum_j^N \epsilon_{ij} \left[\int \psi_i^*(\mathbf{r})\psi_i(\mathbf{r})d\mathbf{r} - \delta_{ij} \right], \tag{1.5.25}$$

where ϵ_{ij} are Lagrange multipliers. At the minimum,

$$\delta\Omega[\{\psi_i\}] = 0 \tag{1.5.26}$$

which leads to,

$$\hat{h}_{\mathrm{eff}}\psi_i = \left[-\frac{1}{2}\hat{\nabla}^2 + \hat{v}_{\mathrm{eff}} \right]\psi_i = \sum_j^N \epsilon_{ij}\psi_j \tag{1.5.27}$$
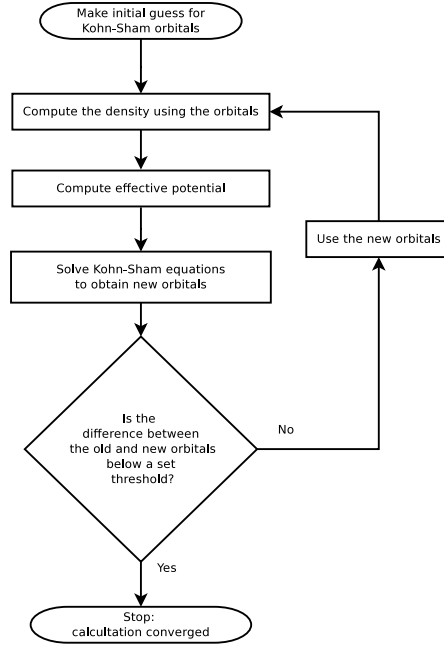
Figure 1.2: Flow chart for the self consistent solution of the Kohn-Sham equations.

where v_{eff} is as defined above in equation 1.5.19. Since $\hat{h}_{\text{eff}}$ is Hermitian, ϵ_{ij} is also Hermitian and as such can be diagonalised by a unitary transformation of the orbitals. The transformation does not cause the density to vary and therefore the Hamiltonian is also unchanged. So equation 1.5.27 can be simplified to give the canonical form of the Kohn-Sham orbital equations,

$$\left[-\frac{1}{2}\hat{\nabla}^2 + \hat{v}_{\text{eff}} \right] \psi_i = \epsilon_i \psi_i \tag{1.5.28}$$

$$v_{\text{eff}} = v_{\text{ext}}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{\text{xc}}(\mathbf{r}) \tag{1.5.29}$$

$$n(\mathbf{r}) = \sum_i^N |\psi_i(\mathbf{r})|^2. \tag{1.5.30}$$

where ψ_i are canonical wavefunctions which are the eigenfunctions of the Schrödinger equation. To obtain a solution for the ground state energy this set of equations need to be solved self consistently, a procedure for this is shown in figure 1.2. A self-consistent procedure for this is required because v_{eff} is a function of the density and therefore the wavefunctions also depend on the density.

Kohn-Sham theory created a mapping from the interacting system being studied to a non-interacting system with the same density [6].

1.5.3   Exchange-correlation functionals

Kohn-Sham theory introduces the exchange-correlation functional but does not provide an expression for it. The form of the exact exchange-correlation functional remains unknown and so approximations are required to calculate the exchange-correlation energy and po-

tential. It is common to separate the exchange-correlation functional into exchange and correlation parts,

$$E_{\text{xc}}[n] = E_{\text{X}}[n] + E_{\text{C}}[n]. \tag{1.5.31}$$

The two parts are then dealt with separately. There are several approximations of the exchange-correlation functional, some of these are discussed below.

1.5.3.1 Local density approximation

The Local Density Approximation (LDA) uses as its exchange energy functional an expression derived from the exact exchange energy for jellium, a collection of electrons with uniform charge density and an opposite uniform positive charge. The Schrödinger equation for an non-interacting electron in jellium is of the form,

$$\left(\nabla^2 + |\mathbf{k}|^2\right)\psi = 0. \tag{1.5.32}$$

Since periodic boundary conditions apply the solutions to this equation are plane waves.

The charge density of jellium can be expressed in terms of k_F, the $|\mathbf{k}|$ at the Fermi level, to be $n(\mathbf{r}) = \frac{k_F^3}{3\pi^2}$, this is a constant which is expected as the density is not position dependent. This expression for the density can be derived by considering that in reciprocal space there are $V/(2\pi)^3$ states per unit volume in a crystal of volume V, and all the occupied states are constrained to lie within the Fermi sphere of radius k_F, therefore,

$$N_e = 2\frac{V}{(2\pi)^3}\frac{4\pi k_F^3}{3} \tag{1.5.33}$$

which rearranges to give [24],

$$\frac{N_e}{V} = \frac{k_F^3}{3\pi^2} \tag{1.5.34}$$

$$= n(\mathbf{r}). \tag{1.5.35}$$

The LDA can be derived starting from the exact Hartree-Fock exchange energy expression,

$$E_X[\{\psi_i\}] = -\int\int\sum_{ij}\frac{\psi_i^*(\mathbf{r}_1)\psi_i(\mathbf{r}_2)\psi_j^*(\mathbf{r}_2)\psi_j(\mathbf{r}_1)d\mathbf{r}_1 d\mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2|} \tag{1.5.36}$$

and rewriting this in terms of plane waves giving,

$$= -\int\int\sum_{ij}\frac{e^{i\mathbf{k}_i\cdot\mathbf{r}_1}e^{i\mathbf{k}_i\cdot\mathbf{r}_2}e^{-i\mathbf{k}_j\cdot\mathbf{r}_2}e^{i\mathbf{k}_j\cdot\mathbf{r}_1}d\mathbf{r}_1 d\mathbf{r}_2}{V^2|\mathbf{r}_1 - \mathbf{r}_2|} \tag{1.5.37}$$

$$= -\int\int\sum_{\mathbf{k}_i}^{\mathbf{k}_F}\sum_{\mathbf{k}_j}^{\mathbf{k}_F}\frac{e^{-i\mathbf{k}_i\cdot(\mathbf{r}_2-\mathbf{r}_1)}e^{i\mathbf{k}_j\cdot(\mathbf{r}_1-\mathbf{r}_2)}d\mathbf{r}_1 d\mathbf{r}_2}{V^2|\mathbf{r}_1 - \mathbf{r}_2|}. \tag{1.5.38}$$

Since the $\mathbf{k}$ sums are very closely spaced they can be replaced by integrals (using $d\mathbf{n} = \frac{V}{8\pi^3}d\mathbf{k}$),

$$= -\int\int\int^{\mathbf{k}_F}\int^{\mathbf{k}_F} \frac{e^{i\mathbf{k}\cdot(\mathbf{r}_2-\mathbf{r}_1)}e^{i\mathbf{k}'\cdot(\mathbf{r}_1-\mathbf{r}_2)}d\mathbf{k}d\mathbf{k}'d\mathbf{r}_1 d\mathbf{r}_2}{64\pi^6|\mathbf{r}_1-\mathbf{r}_2|} \qquad (1.5.39)$$

now make the substitutions $\mathbf{r} = \frac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2)$ and $\mathbf{s} = (\mathbf{r}_1 - \mathbf{r}_2)$ to give

$$= -\int\int\int^{\mathbf{k}_F}\int^{\mathbf{k}_F} \frac{e^{-i\mathbf{k}\cdot\mathbf{s}}e^{i\mathbf{k}'\cdot\mathbf{s}}d\mathbf{k}d\mathbf{k}'d\mathbf{r}d\mathbf{s}}{64\pi^6|\mathbf{s}|} \qquad (1.5.40)$$

now split the integrals and integrate using polar coordinates,

$$= -\int\int \frac{1}{64\pi^6|\mathbf{s}|}\left(\left(\int_0^{k_F}\int_0^{\pi}\int_0^{2\pi} k^2 e^{-i\mathbf{k}\cdot\mathbf{s}}\sin\theta d\phi d\theta dk\right)\right.$$
$$\left.\left(\int_0^{k_F}\int_0^{\pi}\int_0^{2\pi} k'^2 e^{i\mathbf{k}'\cdot\mathbf{s}}\sin\theta' d\phi' d\theta' dk'\right)\right)d\mathbf{r}d\mathbf{s} \qquad (1.5.41)$$

$$= -\int\int \frac{1}{64\pi^6|\mathbf{s}|}\left(\left(\int_0^{k_F}\int_0^{\pi} k^2 e^{-ik\cos\theta|\mathbf{s}|}\sin\theta \int_0^{2\pi} d\phi d\theta dk\right)\right.$$
$$\left.\left(\int_0^{k_F}\int_0^{\pi} k'^2 e^{ik'\cos\theta'|\mathbf{s}|}\sin\theta' \int_0^{2\pi} d\phi' d\theta' dk'\right)\right)d\mathbf{r}d\mathbf{s} \qquad (1.5.42)$$

$$= -\int\int \frac{4\pi^2}{64\pi^6|\mathbf{s}|}\left(\left(\int_0^{k_F} k^2 \left[\frac{e^{-ik\cos\theta|\mathbf{s}|}}{ik|\mathbf{s}|}\right]_{\theta=0}^{\theta=\pi} dk\right)\right.$$
$$\left.\left(\int_0^{k_F} k'^2 \left[\frac{e^{ik'\cos\theta'|\mathbf{s}|}}{-ik'|\mathbf{s}|}\right]_{\theta'=0}^{\theta'=\pi} dk'\right)\right)d\mathbf{r}d\mathbf{s} \qquad (1.5.43)$$

$$= -\int\int \frac{1}{16\pi^4|\mathbf{s}|^3}\left(\left(\int_0^{k_F} ik\left[e^{-ik|\mathbf{s}|} - e^{ik|\mathbf{s}|}\right] dk\right)\right.$$
$$\left.\left(\int_0^{k_F} ik'\left[e^{-ik'|\mathbf{s}|} - e^{ik'|\mathbf{s}|}\right] dk\right)\right)d\mathbf{r}d\mathbf{s} \qquad (1.5.44)$$

which integrates to give,

$$= -\int\int \frac{1}{16\pi^4|\mathbf{s}|^5}\left(\left(\frac{2}{|\mathbf{s}|}\sin k_F|\mathbf{s}| - 2k_F\cos k_F|\mathbf{s}|\right)\right.$$
$$\left.\left(\frac{2}{|\mathbf{s}|}\sin k_F|\mathbf{s}| - 2k_F\cos k_F|\mathbf{s}|\right)\right)d\mathbf{r}d\mathbf{s} \qquad (1.5.45)$$

$$= -\int\int \frac{1}{4\pi^4|\mathbf{s}|^7}\left(\sin k_F|\mathbf{s}| - k_F|\mathbf{s}|\cos k_F|\mathbf{s}|\right)^2 d\mathbf{r}d\mathbf{s} \qquad (1.5.46)$$

now insert the density into this expression,

$$= -\int\int \left(\frac{k_F^3}{3\pi^2}\right)^2 \frac{9}{4k_F^6|\mathbf{s}|^7} \left(\sin k_F|\mathbf{s}| - k_F|\mathbf{s}|\cos k_F|\mathbf{s}|\right)^2 d\mathbf{r}d\mathbf{s} \tag{1.5.47}$$

$$= -\int n^2 \int \frac{9}{4k_F^6|\mathbf{s}|^7} \left(\sin k_F|\mathbf{s}| - k_F|\mathbf{s}|\cos k_F|\mathbf{s}|\right)^2 d\mathbf{s}d\mathbf{r} \tag{1.5.48}$$

$$= -\int n^2 \int_0^\infty \frac{9\pi}{k_F^6|\mathbf{s}|^5} \left(\sin k_F|\mathbf{s}| - k_F|\mathbf{s}|\cos k_F|\mathbf{s}|\right)^2 d|\mathbf{s}|d\mathbf{r} \tag{1.5.49}$$

using the substitution $t = k_F|\mathbf{s}|$ to perform the integration gives,

$$= -\int n^2 \int_0^\infty \frac{9\pi}{k_F^2} \frac{(\sin t - t\cos t)^2}{t^5} dt d\mathbf{r} \tag{1.5.50}$$

$$= -\int_V \frac{9\pi}{4k_F^2} n^2 d\mathbf{r} \tag{1.5.51}$$

now expressing k_F in terms of n results in,

$$= -\int \frac{9\pi}{4\left(3\pi^2 n\right)^{\frac{2}{3}}} n^2 d\mathbf{r} \tag{1.5.52}$$

$$= -\frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int n^{\frac{4}{3}} d\mathbf{r}. \tag{1.5.53}$$

In the LDA this expression for the exchange energy in terms of the density is adopted. In general the density is position dependent and so the uniform density in equation 1.5.53 is substituted for the true density to give the expression for the LDA,

$$E_X^{LDA}[n] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int n(\mathbf{r})^{\frac{4}{3}} d\mathbf{r}. \tag{1.5.54}$$

Correlation functionals have been developed to complement this exchange functional. These include CAPZ (Ceperley, Alder, Perdew and Zunger) [36, 37] and VWN (Vosko, Wilk and Nusair) [38]. Both of these correlation functionals are parametrisations of the correlation energy obtained for different values of n in numerical quantum Monte-Carlo simulations of the uniform electron gas published by Ceperley and Alder [36]. Different methods of interpolation were used to obtain each each parametrisation however the difference between the energies obtained is small (a maximum of 1.3% relative deviation at points representing a range of densities) [37].

The LDA has been used in many calculations over the years but has found more favour with physicists than chemists. The approximation made by the LDA produces solutions for calculations on systems of chemical interest with a level of accuracy unsatisfactory to many chemists. The systems of interest to physicists are closer to the uniform electron gas, such as, for example, bulk metals. More sophisticated functionals have therefore been developed.

1.5.3.2   Generalised Gradient Approximation (GGA) functionals

In the Generalised Gradient Approximation (GGA) both the value of the density at a point and the gradient of the density at this point are used. The consideration of the density gradient is important in systems that don't resemble the uniform electron gas [39,40]. BLYP is one early example of a GGA (or gradient corrected) functional. It is a combination of the Becke88 exchange functional [41] and the Lee, Yang and Parr (LYP) correlation functional [42, 43]. The Becke88 exchange formula is,

$$E_X^{\text{Becke88}} = E_X^{\text{LDA}} - \beta \sum_\sigma \int n_\sigma(\mathbf{r})^{\frac{4}{3}} \frac{x_\sigma(\mathbf{r})^2}{1 + 6\beta x_\sigma(\mathbf{r})\sinh^{-1}x_\sigma(\mathbf{r})} d^3\mathbf{r}, \tag{1.5.55}$$

where $\beta = 0.0042$ a.u. and

$$x_\sigma(\mathbf{r}) = \frac{|\nabla n_\sigma(\mathbf{r})|}{n_\sigma(\mathbf{r})^{\frac{4}{3}}}. \tag{1.5.56}$$

The formula for LYP correlation in a spin unpolarised system is,

$$\begin{aligned} E_C^{\text{LYP}} = & - a \int \frac{n(\mathbf{r})}{1 + dn(\mathbf{r})^{-\frac{1}{3}}} d^3\mathbf{r} \\ & - ab \int \omega(\mathbf{r}) n(\mathbf{r})^2 \left[C_F n(\mathbf{r})^{\frac{8}{3}} - |\nabla n(\mathbf{r})|^2 \left(\frac{1}{24} - \frac{7}{12}\delta(\mathbf{r}) \right) \right] d^3\mathbf{r} \end{aligned} \tag{1.5.57}$$

where,

$$\omega(\mathbf{r}) = \frac{e^{-cn(\mathbf{r})^{-\frac{1}{3}}}}{1 + dn(\mathbf{r})^{-\frac{1}{3}}} n(\mathbf{r})^{-\frac{11}{3}} \tag{1.5.58}$$

$$\delta(\mathbf{r}) = cn(\mathbf{r})^{-\frac{1}{3}} + \frac{dn(\mathbf{r})^{-\frac{1}{3}}}{1 + dn(\mathbf{r})^{-\frac{1}{3}}} \tag{1.5.59}$$

$$C_F = \frac{3}{10}(3\pi^2)^{\frac{2}{3}} \tag{1.5.60}$$

$$a = 0.04918 \ a.u.; \ b = 0.132 \ a.u.; \ c = 0.2533 \ a.u.; \ d = 0.349 \ a.u.. \tag{1.5.61}$$

Other examples of GGA functionals include B86 [44], PW91 [45,46] and PBE [47].

1.5.3.3   Meta-GGA functionals

Another kind of functional are meta-GGA functionals. These improve upon the GGA by additionally including a dependence on the second derivative of the density. This improves accuracy, however this is often at the expense of numeric stability in the calculations. Numerical stability can be improved by introducing a dependence on the kinetic energy density τ,

$$\tau(\mathbf{r}) = \sum_i^{\text{occupied}} \frac{1}{2}|\nabla\psi_i(\mathbf{r})|^2, \tag{1.5.62}$$

instead. PKZB [48] and TPSS [49] are examples of meta-GGA functionals.

1.5.3.4 Hybrid functionals

The final common class of functional are hybrid functionals. These are (usually semi-empirical) functionals that combine different amounts of other functionals and also a portion of exact exchange (as defined in Hartree-Fock). B3LYP [42, 50–52] and B97 [53] are examples of hybrid functionals [14]. Hybrid functionals are discussed further in chapter 6.

1.6 Density matrix theory

It is possible to formulate an exact expression for the energy of a system, the expectation value of the Hamiltonian, in terms of one and two particle density matrices [6]. The single particle density matrix is defined as,

$$\rho_1(\mathbf{r}_1', \mathbf{r}_1) = N \int \cdots \int \Psi(\mathbf{r}_1', s_1, \mathbf{r}_2, s_2, \cdots, \mathbf{r}_N, s_N)$$
$$\Psi^*(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2, \cdots, \mathbf{r}_N, s_N) d\mathbf{r}_2 \cdots d\mathbf{r}_N ds_1 ds_2 \cdots ds_N \tag{1.6.1}$$

where the factor N is the number of electrons and Ψ is the wavefunction of the system. For a system of non-interacting particles, the density matrix can also be represented in terms of orbitals as,

$$\rho_1(\mathbf{r}_1', \mathbf{r}_1) = 2 \sum_i^{\text{occupied}} \psi_i(\mathbf{r}_1') \psi_i^*(\mathbf{r}_1) \tag{1.6.2}$$

where ψ_i is a spatial Hartree-Fock or Kohn-Sham orbital. The diagonal elements are,

$$\rho_1(\mathbf{r}_1) = \rho_1(\mathbf{r}_1, \mathbf{r}_1) \tag{1.6.3}$$

$$= N \int \cdots \int |\Psi(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2, \cdots, \mathbf{r}_N, s_N)|^2 d\mathbf{r}_2 \cdots d\mathbf{r}_N ds_1 ds_2 \cdots ds_N \tag{1.6.4}$$

$$= n(\mathbf{r}_1). \tag{1.6.5}$$

Similarly the two particle density matrix is,

$$\rho_2(\mathbf{r}_1', \mathbf{r}_2', \mathbf{r}_1, \mathbf{r}_2) = \frac{N(N-1)}{2} \int \cdots \int \Psi(\mathbf{r}_1', s_1, \mathbf{r}_2', s_2, \cdots, \mathbf{r}_N, s_N)$$
$$\Psi^*(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2, \mathbf{r}_3, s_3, \cdots, \mathbf{r}_N, s_N) d\mathbf{r}_3 \cdots d\mathbf{r}_N ds_1 ds_2 ds_3 \cdots ds_N \tag{1.6.6}$$

where the factor of $\frac{N(N-1)}{2}$ is the number of equivalent pairs. The diagonal elements are,

$$\rho_2(\mathbf{r}_1, \mathbf{r}_2) = \rho_2(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_1, \mathbf{r}_2) \tag{1.6.7}$$

$$= \frac{N(N-1)}{2} \int \cdots \int |\Psi(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2, \mathbf{r}_3, s_3, \cdots, \mathbf{r}_N, s_N)|^2$$
$$d\mathbf{r}_3 \cdots d\mathbf{r}_N ds_1 ds_2 ds_3 \cdots ds_N. \tag{1.6.8}$$

Using the above expressions the energy may be written [54],

$$E = \int \left[-\frac{1}{2} \nabla_{\mathbf{r}}^2 \rho_1(\mathbf{r}', \mathbf{r}) \right]_{\mathbf{r}'=\mathbf{r}} d\mathbf{r} + \int v_{\text{ext}}(\mathbf{r}) \rho_1(\mathbf{r}) d\mathbf{r} + \int \int \frac{\rho_2(\mathbf{r}_1, \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \tag{1.6.9}$$

In the Kohn-Sham case the use of the two-particle density matrix can be avoided since it is an independent particle theory,

$$E = \int \left[-\frac{1}{2} \nabla_{\mathbf{r}}^2 \rho_1(\mathbf{r}', \mathbf{r}) \right]_{\mathbf{r}'=\mathbf{r}} d\mathbf{r} + \int v_{\text{ext}}(\mathbf{r}) \rho_1(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \int \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{\text{xc}}[n].$$

$$(1.6.10)$$

1.6.1   Idempotency

An idempotent matrix is equal to its square, i.e. $\mathbf{A} = \mathbf{A}^2$. It is necessary for the ground state one-particle density matrix to be idempotent (at zero temperature). This requirement is the same as restricting all the eigenvalues of the density matrix to either 0 or 1. The eigenvalues of the matrix correspond to the orbital occupation numbers [29]. The idempotency constraint also enforces the orbital orthogonality requirements since [25],

$$\rho_1^2(\mathbf{r}, \mathbf{r}') = \int d\mathbf{r}'' \rho_1(\mathbf{r}, \mathbf{r}'') \rho_1(\mathbf{r}'', \mathbf{r}') \tag{1.6.11}$$

$$= \sum_{ij} f_i f_j \psi_i(\mathbf{r}) \psi_j^*(\mathbf{r}') \int d\mathbf{r}'' \psi_i^*(\mathbf{r}'') \psi_j(\mathbf{r}'') \tag{1.6.12}$$

$$= \sum_{ij} f_i f_j \psi_i(\mathbf{r}) \psi_j^*(\mathbf{r}') \delta_{ij} \tag{1.6.13}$$

$$= \sum_i f_i^2 \psi_i(\mathbf{r}) \psi_i^*(\mathbf{r}') \tag{1.6.14}$$

$$= \rho_1(\mathbf{r}, \mathbf{r}') \qquad \Leftrightarrow \qquad f_i^2 = f_i \qquad \forall i. \tag{1.6.15}$$

Methods for enforcing this during a calculation will be discussed in section 2.6.

1.7   Linear-scaling methods

The computational scaling of electronic structure methods is very important as this determines the limit as to size of the system which can be studied using that method. Conventional DFT methods scale cubically with the number of atoms N_{at} [55]. With codes using wavefunction methods, cubic-scaling with respect to the number of atoms is the minimum scaling encountered, post-Hartree-Fock methods, for example CCSD(T) have order $\mathcal{O}(N_{\text{at}}^7)$ scaling. The calculation of the two electron integrals in wavefunction methods scales with $\mathcal{O}(N^4)$ with respect to basis set size (unless further approximations are made) [15]. In recent years, encouraged by advances in computer processing power, linear-scaling codes have been developed. Linear-scaling codes are less efficient with smaller systems than conventional codes, the crossover point is around 100-200 atoms.

With a linear-scaling code the study of molecules and materials involving thousands of atoms becomes practical. Examples of such systems that occur in important applications are protein fragments, large biomolecules including their solvation sphere and extended nanostructures [2].

1.7.1 Near-sightedness

The properties at one point in a quantum system can be considered to be independent (at least directly) of the state of the system far away from this point. This principle was developed by Walter Kohn and is known as locality or near-sightedness [56, 57]. If this principle can be exploited in electronic structure calculations, the resulting method could potentially scale linearly with system size. The divide and conquer approach discussed below is one such method. In classical mechanics methods that take advantage of locality and scale linearly are feasible. Even long ranged forces, for example Coulomb interactions, can be treated using methods that scale linearly. In contrast quantum mechanical systems are not local, for example, since all the electrons in a system are indistinguishable, the symmetry and antisymmetry conditions obeyed by wavefunctions must be satisfied by all the electrons in the system.

However while a full description of the state of the system is non local, certain properties of the system, notably the density matrix (and hence the density) and the total energy, can be calculated using local information. In a quantum system of many particles the long range effects on the density at a point tend to interfere with each other and so become very small. Exponential decay is found in the one-electron density matrix in insulators [58].

$$\rho_1(\mathbf{r}, \mathbf{r}') \sim e^{-\gamma |\mathbf{r} - \mathbf{r}'|}, \tag{1.7.1}$$

where γ depends on the band gap of the system. A method can take advantage of this by truncating the density matrix at some point [29]. This can be done by setting the density matrix elements with arguments separated by more than r_{cut} to zero,

$$\rho_1(\mathbf{r}, \mathbf{r}') = 0 \qquad \text{when} \qquad |\mathbf{r} - \mathbf{r}'| > r_{\text{cut}}. \tag{1.7.2}$$

The number of non-zero elements in such a truncated density matrix will scale (asymptotically) linearly with system size. Some variational freedom is lost by the imposition of a cutoff so the energy should also be converged with respect to r_{cut} [25].

1.7.2 Divide and conquer

One of the first linear-scaling methods was the divide and conquer (DVC) scheme [59]. DVC is conceptually very simple it achieves linear-scaling by breaking the system into subsystems each of which is then treated using conventional methods. In the initial implementation the density was partitioned [59], the method was subsequently improved by instead partitioning the density matrix [60]. Buffer regions need to be added around each subsystem to ensure that the properties in each system are calculated accurately. This makes the calculations costly, but does not preclude linear-scaling in large enough systems [29].

1.7.3 Density matrix approaches

The Li-Nunes and Vanderbilt (LNV) method [61] is an algorithm for the optimisation of the density matrix while ensuring idempotency. For a truncated density matrix this method

can scale linearly; a method of building the Hamiltonian in a linear-scaling manner is also required. Also the scaling of this method is determined by the number of basis functions so smaller basis sets are most suitable [29]. The LNV method is discussed further in section 2.6.3.

1.7.4   Localised function approaches

To calculate the Hamiltonian in linear-scaling manner localised functions or orbitals are usually required. When localised basis functions are used the localised functions to use are obvious; for delocalised basis sets such as plane waves further transformations are required to provide the necessary locality. These are the same orbitals that appear in the definition of the density matrix in equation 1.6.2. Therefore instead of directly optimising the density matrix the localised orbitals can be optimised instead [29].

Chapter 2

The ONETEP approach

ONETEP [62] is a linear-scaling density matrix method where the energy is minimised directly with respect to the density matrix. It is a reformulation of the plane wave pseudopotential method where high accuracy is achieved at linear-scaling cost with localised functions that are optimised self-consistently during the calculation. ONETEP is able to perform single-point energy calculations and also geometry optimisations and properties calculations.

2.1 Non-orthogonal Generalised Wannier Functions (NGWFs)

The localised functions used in ONETEP are Non-orthogonal Generalised Wannier Functions (NGWFs), these can be used to calculate the density matrix. Wannier functions [63] are obtained by a unitary transformation of the Bloch wave-functions [25]. Wannier functions are exponentially localised in the general one dimensional crystal [64]. In ONETEP the orthogonality of the Wannier functions is not enforced. As a result the NGWFs do not need to have "orthogonality tails" which allows more stringent localisation to be enforced. Therefore the localised functions used in ONETEP are not strictly Wannier functions.

Each NGWF is centred and localised around the nucleus of an atom. These functions are not fixed instead they are optimised during the calculation subject to localisation constraints so a large number of NGWFs are not required. Non-orthogonal functions are used since these can be made more localised than an equivalent set of orthogonal functions. The localised NGWFs (ϕ_α) used in ONETEP are related to the Kohn-Sham eigenfunctions $\{\psi_i\}$ by,

$$\psi_i(\mathbf{r}) = \phi_\alpha(\mathbf{r})M^\alpha{}_i \qquad (2.1.1)$$

where M is a linear transformation [65] and the implicit Einstein summation convention has been used with repeated Greek indices. Equation 2.1.1 can be used to express the

one-particle density matrix in terms of the localised orbitals,

$$\rho_1(\mathbf{r}, \mathbf{r}') = \sum_n \psi_n(\mathbf{r}) f_n \psi_n^*(\mathbf{r}') \tag{2.1.2}$$

$$= \sum_n \phi_\alpha(\mathbf{r}) M^\alpha{}_n f_n (M^\dagger)_n{}^\beta \phi_\beta^*(\mathbf{r}') \tag{2.1.3}$$

$$= \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta^*(\mathbf{r}') \tag{2.1.4}$$

where f_n is the occupancy of orbital $\psi_n(\mathbf{r})$ number and $K^{\alpha\beta}$ is the density kernel defined,

$$K^{\alpha\beta} = \sum_n M^\alpha{}_n f_n (M^\dagger)_n{}^\beta. \tag{2.1.5}$$

The density matrix is truncated by setting the elements of the density kernel $K^{\alpha\beta}$ to zero when the centres of the localised orbitals $\phi_\alpha(\mathbf{r})$ and $\phi_\beta^*(\mathbf{r}')$ are further apart than a preset cutoff radius [25, 66]. NGWFs are stored as values on a grid divided into small parallelepipeds (PPDs) [2]. The NGWFs are initialised to a set of Gaussians or to a precalculated set of pseudo-atomic orbitals known in ONETEP as fireballs.

2.2   Psinc basis set

In ONETEP periodic sinc (cardinal sine) functions are used as a basis set to express the NGWFs using the following formula,

$$\phi_\alpha(\mathbf{r}) = \sum_m D_m(\mathbf{r}) c_{m\alpha} \tag{2.2.1}$$

where $D_m(\mathbf{r})$ is a psinc function centred at the point $\mathbf{r}_m$ and the $c_{m\alpha}$ are expansion co-efficients [65]. Psinc functions are easily expressed as a sum of plane waves and by their nature allow use of efficient discrete Fourier transforms for the evaluation of integrals and other quantities such as densities [67]. The definition of a psinc centred on a point $\mathbf{r}_{KLM}$ using plane waves is given by,

$$D_{KLM}(\mathbf{r}) = D(\mathbf{r} - \mathbf{r}_{KLM}) \tag{2.2.2}$$

$$= \frac{1}{N_1 N_2 N_3} \sum_{p=-J_1}^{J_1} \sum_{q=-J_2}^{J_2} \sum_{s=-J_3}^{J_3} e^{i(p\mathbf{B}_1 + q\mathbf{B}_2 + s\mathbf{B}_3) \cdot (\mathbf{r} - \mathbf{r}_{KLM})} \tag{2.2.3}$$

where p, q and s are integers and $N_i = 2J_i + 1$ are the number of grid points along direction i. An odd number of grid points in each direction in the cell is required to allow the psinc functions to be both real and orthogonal [25]. With an even grid the requirement for periodicity prevents orthogonality. The $\mathbf{B}_i$ are the reciprocal lattice vectors defined in section 1.4.2. The primitive lattice vectors, $\mathbf{A}_i$, and the reciprocal lattice vectors, $\mathbf{B}_i$, satisfy the orthogonality relation,

$$\mathbf{B}_i \cdot \mathbf{A}_j = 2\pi \delta_{ij}. \tag{2.2.4}$$

The grid points of the simulation cell, $\mathbf{r}_{KLM}$, are given by,

$$\mathbf{r}_{KLM} = \frac{K}{N_1}\mathbf{A}_1 + \frac{L}{N_2}\mathbf{A}_2 + \frac{M}{N_3}\mathbf{A}_3 \qquad (2.2.5)$$

where K, L and M are non-negative integers lower than the corresponding N_i [66,68]. The vector $\mathbf{r}$ can be expressed in a similar fashion,

$$\mathbf{r} = \sum_{i=1}^{3} \frac{x_i}{N_i}\mathbf{A}_i, x_i \in \mathbb{R}. \qquad (2.2.6)$$

Using equation 2.2.4 and 2.2.5 equation 2.2.3 can be factorised to give,

$$D_{KLM}(\mathbf{r}) = \mathcal{D}_K^{(1)}(x_1)\mathcal{D}_L^{(2)}(x_2)\mathcal{D}_M^{(3)}(x_3) \qquad (2.2.7)$$

where

$$\mathcal{D}_I^{(j)}(x_j) = \frac{1}{N_j} \sum_{p=-J_j}^{J_j} e^{2i\pi p(x_j-I)/N_j}. \qquad (2.2.8)$$

It is easy to see that equation 2.2.8 consists of a geometric sum with first term,

$$a = \frac{1}{N_j} e^{-i\pi(x_j-I)(1-1/N_j)} \qquad (2.2.9)$$

and common ratio,

$$r = e^{2i\pi(x_j-I)/N_j}. \qquad (2.2.10)$$

Hence $\mathcal{D}_I^{(j)}(x_j)$ can be written as,

$$\mathcal{D}_I^{(j)}(x_j) = \frac{a(1-r^{N_j})}{1-r} \qquad (2.2.11)$$

$$= \frac{1}{N_j}\frac{\sin(\pi(x_j-I))}{\sin(\pi(x_j-I)/N_j)} \qquad (2.2.12)$$

which is real valued. Clearly when $\mathbf{r}$ corresponds to a grid point the x_j will be integers and $\sin(\pi x_j)$ will be zero and therefore $\mathcal{D}^{(j)}(x_j)$ and $D_{KLM}(\mathbf{r})$ will both be zero unless $\mathbf{r} = \mathbf{r}_{KLM}$ in which case they will have value one since,

$$\lim_{x_j \to I} \frac{1}{N_j}\frac{\sin(\pi(x_j-I))}{\sin(\pi(x_j-I)/N_j)} = \frac{1}{N_j}\frac{\pi(x_j-I)}{\pi(x_j-I)/N_j} = 1. \qquad (2.2.13)$$

The property,

$$D_{KLM}(\mathbf{r}_{NPQ}) = \delta_{KN}\delta_{LP}\delta_{MQ} \qquad (2.2.14)$$

is called cardinality. Figure 2.1 illustrates this and also compares the periodic psinc functions with a standard sinc function. A single psinc function with the periodicity of the simulation cell is centred on each point in the real space grid [69].

Equation 2.2.3 can also be written more simply as a sum of plane waves up to the cutoff wavevector, $|\mathbf{G}_{\mathrm{max}}|$,

$$D_m(\mathbf{r}) = \frac{1}{N} \sum_{\mathbf{G}_p}^{|\mathbf{G}_{\mathrm{max}}|} e^{i\mathbf{G}_p \cdot (\mathbf{r}-\mathbf{r}_m)} \tag{2.2.15}$$

where N is the total number of grid points in the simulation cell and $\mathbf{G}_p$ is a reciprocal lattice vector defined (c.f. equation 1.4.12) as,

$$\mathbf{G}_p = \sum_{i=1}^{3} g_i \mathbf{B}_i \qquad g_i \in \mathbb{Z}. \tag{2.2.16}$$

2.2.1   Overlap matrix elements

The overlap matrix, $\mathbf{s}$ of the psinc function is given by,

$$s_{ij} = \int_V D_i^*(\mathbf{r}) D_j(\mathbf{r}) d\mathbf{r} \tag{2.2.17}$$

$$= \frac{1}{N^2} \sum_p^{\mathrm{max}} \sum_q^{\mathrm{max}} e^{i\mathbf{G}_p \cdot \mathbf{r}_i - i\mathbf{G}_q \cdot \mathbf{r}_j} \int_V e^{i(\mathbf{G}_p - \mathbf{G}_q) \cdot \mathbf{r}} d\mathbf{r} \tag{2.2.18}$$

$$= \frac{V}{N^2} \sum_p^{\mathrm{max}} \sum_q^{\mathrm{max}} e^{i\mathbf{G}_p \cdot \mathbf{r}_i - i\mathbf{G}_q \cdot \mathbf{r}_j} \delta_{pq} \tag{2.2.19}$$

$$= \frac{V}{N^2} \sum_p^{\mathrm{max}} e^{i\mathbf{G}_p \cdot (\mathbf{r}_i - \mathbf{r}_j)} \tag{2.2.20}$$

$$= w \delta_{ij} \tag{2.2.21}$$

where $w = \frac{V}{N}$ is the grid point weight (V is the cell volume and N is the number of points) and the relations,

$$\int_V e^{i(\mathbf{G}_p - \mathbf{G}_q) \cdot \mathbf{r}} d\mathbf{r} = V \delta_{pq} \tag{2.2.22}$$

and,

$$\sum_p^{\mathrm{max}} e^{i\mathbf{G}_p \cdot (\mathbf{r}_i - \mathbf{r}_j)} = N D_j(\mathbf{r}_i) \tag{2.2.23}$$

$$= N \delta_{ij} \tag{2.2.24}$$

have been used. It can be seen from this that psinc functions on distinct grid points are orthogonal [25, 65].
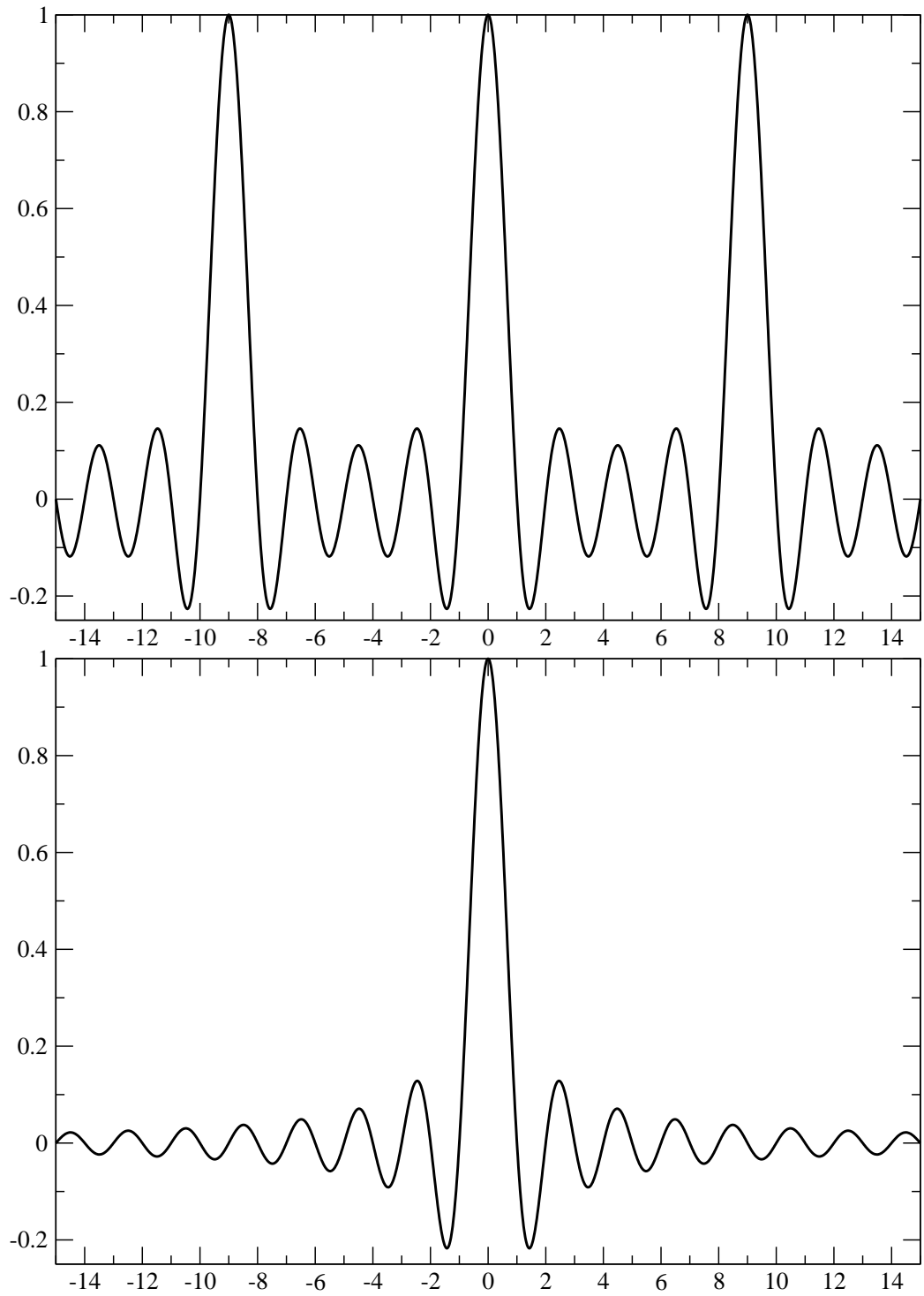
Figure 2.1: Top: One dimensional psinc function, $\mathcal{D}_0^{(i)}(x_i)$, with $N_i = 9$, see equation 2.2.12. Bottom: One dimensional sinc function. Note how both functions are either zero or one for integer values of x_i.

2.2.2   Fine grid psincs

In plane wave codes the electronic charge density must be represented by a set of plane waves with twice the cutoff frequency of the standard set used to describe the wavefunctions, to prevent aliasing problems. In ONETEP aliasing will also be a problem when NGWF products are calculated given that the psinc functions are a unitary transformation of the plane wave basis set. Consider the product of two identical standard psinc functions (both centred at $\mathbf{r}_m$),

$$D_m(\mathbf{r})D_m(\mathbf{r}) = \frac{1}{N}\sum_{\mathbf{G}_p}^{|\mathbf{G}_{\max}|} e^{i\mathbf{G}_p \cdot (\mathbf{r}-\mathbf{r}_m)} \frac{1}{N}\sum_{\mathbf{G}_p'}^{|\mathbf{G}_{\max}|} e^{i\mathbf{G}_p' \cdot (\mathbf{r}-\mathbf{r}_m)} \tag{2.2.25}$$

$$= \frac{1}{N^2}\sum_{\mathbf{G}_p}^{|\mathbf{G}_{\max}|}\sum_{\mathbf{G}_p'}^{|\mathbf{G}_{\max}|} e^{i\mathbf{G}_p \cdot (\mathbf{r}-\mathbf{r}_m)} e^{i\mathbf{G}_p' \cdot (\mathbf{r}-\mathbf{r}_m)} \tag{2.2.26}$$

$$= \frac{1}{N^2}\sum_{\mathbf{G}_p}^{|\mathbf{G}_{\max}|}\sum_{\mathbf{G}_p'}^{|\mathbf{G}_{\max}|} e^{i(\mathbf{G}_p + \mathbf{G}_p') \cdot (\mathbf{r}-\mathbf{r}_m)} \tag{2.2.27}$$

$$= \frac{1}{N^2}\sum_{\mathbf{G}_p}^{2|\mathbf{G}_{\max}|} a_p e^{i\mathbf{G}_p \cdot (\mathbf{r}-\mathbf{r}_m)} \tag{2.2.28}$$

where is a_p is a coefficient; clearly some accuracy will be lost if this were to be expressed using a standard psinc function. For ONETEP this means that a grid that is able to represent twice the $|\mathbf{G}_{\max}|$ of the standard grid must be used to calculate the charge density [2]. This is implemented by having a fine grid with twice as many points as the standard "coarse grid" in each direction. A set of fine grid psinc functions, $B_x(\mathbf{r})$, are used to express the density. The $B_x(\mathbf{r})$ are defined by,

$$B_x(\mathbf{r}) = B(\mathbf{r} - \mathbf{r}_x) \tag{2.2.29}$$

$$= \frac{1}{8N_1 N_2 N_3}\sum_{p=-2J_1}^{2J_1+1}\sum_{q=-2J_2}^{2J_2+1}\sum_{s=-2J_3}^{2J_3+1} e^{i(p\mathbf{B}_1 + q\mathbf{B}_2 + s\mathbf{B}_3)\cdot(\mathbf{r}-\mathbf{r}_x)} \tag{2.2.30}$$

with the set $\{\,\mathbf{r}_x\,\}$ of the fine grid points defined by,

$$\mathbf{r}_x = \sum_{i=1}^{3}\frac{x_i}{2N_i}\mathbf{A}_i, \qquad x_i \in \{\,0,1,\ldots,2N_i-1\,\}. \tag{2.2.31}$$

The fine grid psincs obey the cardinality condition on the fine grid points,

$$B_x(\mathbf{r}_y) = \delta_{xy}. \tag{2.2.32}$$

A fine grid psinc in the simpler form of equation 2.2.15 can be represented as,

$$B_m(\mathbf{r}) = \frac{1}{8N}\sum_{\mathbf{G}_p}^{|\mathbf{G}_{\max}'|} e^{i\mathbf{G}_p \cdot (\mathbf{r}-\mathbf{r}_m)} \tag{2.2.33}$$
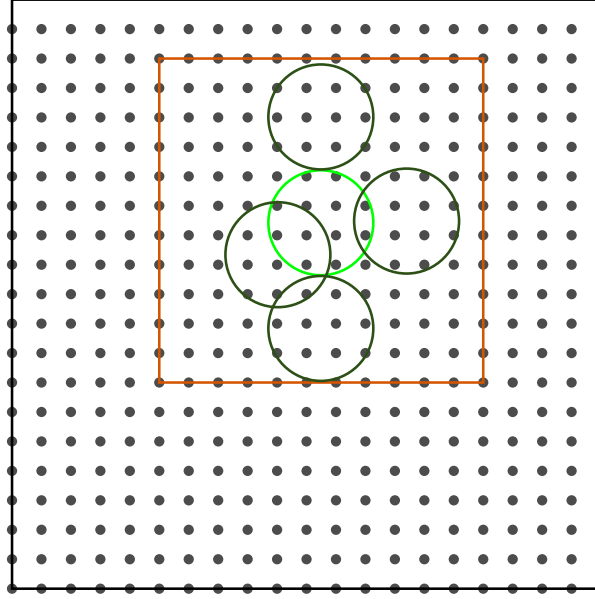
Figure 2.2: An FFT box (orange) inside the simulation cell (black) on the coarse grid. The green circle represents NGWF that is central to this particular FFT box. Any NGWF that overlaps with the central NGWF (highlighted) will fit into the FFT box (by construction).

where $|\mathbf{G}'_{max}| = 2|\mathbf{G}_{max}|$ (with $|\mathbf{G}_{max}|$ having the same value as in equation 2.2.15) [25].

2.3 FFT box

Many of the quantities that need to be calculated in ONETEP involve operations on overlapping NGWFs. For example, the kinetic energy is calculated by applying, in Fourier space, the Laplacian to an NGWF then taking the sum of the point-wise product of this with another (overlapping) NGWF [67]. An FFT box, as shown in figure 2.2, has dimensions that are six times the largest NGWF radius in the calculation (the width of three NGWFs), or if the simulation cell is small the FFT box will adopt the size of the cell. Therefore an FFT box is large enough to contain one central NGWF and all the NGWFs that overlap with it, since the NGWFs are localised in space. In some cases an FFT box will not be wholly located in the primary unit cell and will spill over into one of the periodic images as necessary.

The FFT box is useful for performing operations that require FFTs, including Fourier interpolation and Fourier filtering, because these operations can be performed in this smaller box. Performing the FFTs in these small boxes gives each operation a lower computational cost, compared to performing the operation on the entire simulation cell. Since the size of the FFT box is not related to the system's size (unless the simulation cell is small) the computational effort for such an operation is independent of the size of the system. This is very important for linear-scaling since a linear-scaling number of Fourier operations are required during the calculation of many quantities. In the FFT box the NGWFs are expanded in a basis set of psinc functions that are periodic with respect to the FFT box rather than cell periodic. The FFT box technique corresponds to a coarse sampling in
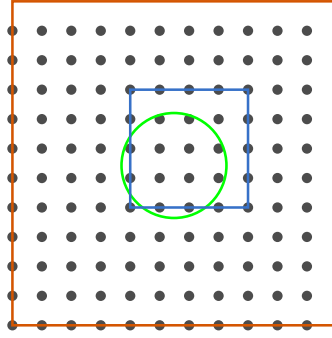
Figure 2.3: A tightbox (blue) inside an FFT box (orange) on the coarse grid. The tightbox can be used to hold a single NGWF (green).

Fourier space of the simulation cell.

2.4   Tightbox

In ONETEP tightboxes are boxes that are used to hold individual NGWFs. These boxes are on the same coarse grid as the simulation cell and FFT box, see figure 2.3. The universal tightbox is a tightbox large enough to hold any NGWF (and every PPD that partially contains that NGWF) in the system. When the FFT box is coincident with the simulation cell the tightbox is also set to the same size as the simulation cell. Currently tightboxes are used during the initialisation of the NGWFs, to store the optimised NGWFs to disk (to allow a calculation to be restarted) and to output the NGWFs in a format suitable for visualisation. Tightboxes were previously used as intermediate arrays when depositing NGWFs to FFT boxes. The NGWF stored in PPD representation was extracted to a tightbox which was subsequently deposited in the appropriate place in the FFT box. NGWFs are now deposited directly into FFT boxes for efficiency reasons [70].

2.5   The Kohn-Sham energy expression in ONETEP

In ONETEP the energy is minimised with respect to both the density kernel and the psinc expansion coefficients of the NGWFs. Starting from equation 1.6.10, the spin unpolarised

KS energy expression in terms of NGWFs and the density kernel can be written as,

$$
\begin{aligned}
E[n] &= \int \left[-\nabla_{\mathbf{r}}^2 \rho_1(\mathbf{r}', \mathbf{r}) \right]_{\mathbf{r}'=\mathbf{r}} d\mathbf{r} + \int v_{\mathrm{loc}}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} \\
&\quad + \frac{1}{2} \int \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{\mathrm{xc}}[n] + E_{\mathrm{nl}}[n]
\end{aligned}
\tag{2.5.1}
$$

$$
\begin{aligned}
&= \int \left[\nabla_{\mathbf{r}}^2 \phi_\alpha(\mathbf{r}') K^{\alpha\beta} \phi_\beta^*(\mathbf{r}) \right]_{\mathbf{r}'=\mathbf{r}} d\mathbf{r} + 2 \int v_{\mathrm{loc}}(\mathbf{r}) \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta^*(\mathbf{r}) d\mathbf{r} \\
&\quad + 2 \int \int \frac{\phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta^*(\mathbf{r}) \phi_\gamma(\mathbf{r}') K^{\gamma\delta} \phi_\delta^*(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{\mathrm{xc}}[n] + E_{\mathrm{nl}}[n]
\end{aligned}
\tag{2.5.2}
$$

$$
\begin{aligned}
&= 2 K^{\alpha\beta} \left\langle \phi_\beta \left| -\tfrac{1}{2}\nabla_{\mathbf{r}}^2 + \hat{v}_{\mathrm{loc}} + \hat{v}_{\mathrm{H}} + \hat{v}_{\mathrm{xc}} + \hat{v}_{\mathrm{nl}} \right| \phi_\alpha \right\rangle \\
&\quad - \frac{1}{2} \int_V n(\mathbf{r}) v_{\mathrm{H}} d\mathbf{r} - \int_V n(\mathbf{r}) v_{\mathrm{xc}} d\mathbf{r} + E_{\mathrm{xc}}[n]
\end{aligned}
\tag{2.5.3}
$$

where ϕ_α is an NGWF, $\hat{v}_{\mathrm{loc}}$ is the local pseudopotential, $\hat{v}_{\mathrm{nl}}$ is the non-local pseudopotential, $\hat{v}_{\mathrm{H}}$ is the Hartree potential and $\hat{v}_{\mathrm{xc}}$ is the exchange correlation potential. By defining the Hamiltonian matrix as,

$$
H_{\beta\alpha} = \left\langle \phi_\beta \left| -\tfrac{1}{2}\hat{\nabla}_{\mathbf{r}}^2 + \hat{v}_{\mathrm{loc}} + \hat{v}_{\mathrm{nl}} + \hat{v}_{\mathrm{xc}} + \tfrac{1}{2}\hat{v}_{\mathrm{H}} \right| \phi_\alpha \right\rangle
\tag{2.5.4}
$$

equation 2.5.3 simplifies to,

$$
E[n] = 2 K^{\alpha\beta} H_{\beta\alpha} - \tfrac{1}{2} \int_V n(\mathbf{r}) v_{\mathrm{H}} d\mathbf{r} - \int_V n(\mathbf{r}) v_{\mathrm{xc}} d\mathbf{r} + E_{\mathrm{xc}}[n].
\tag{2.5.5}
$$

2.5.1   Calculation of the exchange-correlation energy

In electronic structure codes that calculate the density on a grid, such as plane wave codes and ONETEP, the exchange and correlation energy for a generalised gradient approximation (GGA) functional,

$$
E_{\mathrm{xc}}[n] = \int f_{\mathrm{xc}}(n(\mathbf{r}), |\nabla n(\mathbf{r})|) d\mathbf{r}
\tag{2.5.6}
$$

can be expressed as discrete sum,

$$
\bar{E}_{\mathrm{xc}}[n] = \frac{\Omega_{\mathrm{cell}}}{N} \sum_{\mathbf{R}} f_{\mathrm{xc}}(n(\mathbf{R}), |\nabla n(\mathbf{R})|)
\tag{2.5.7}
$$

$$
= \frac{\Omega_{\mathrm{cell}}}{N} \sum_{\mathbf{R}} \left[f_{\mathrm{x}}(n(\mathbf{R}), |\nabla n(\mathbf{R})|) + f_{\mathrm{c}}(n(\mathbf{R}), |\nabla n(\mathbf{R})|) \right]
\tag{2.5.8}
$$

where Ω_{cell} is the unit cell volume, $\mathbf{R}$ are the grid point positions and N is the number of grid points [71].

As part of the current work several hybrid functionals and a pair of GGA functionals, BLYP [41, 42] and XLYP [72], have been implemented. In order to accommodate these additional functionals it was necessary to extensively restructure the ONETEP module that calculates the exchange correlation potential and energy. This will be discussed in more detail in section 6.3.

2.5.2   Calculation of the pseudopotential energy

The pseudopotential approximation is often applied in plane wave codes and is also applied in ONETEP. In the approximation only the valence electrons are treated explicitly; the core electrons are included with the attractive potential of the nucleus to form a smoother pseudopotential. This reduces the number of electrons involved in the calculation which reduces the computational demands. The use of this approximation is very desirable in grid based codes because it allows a coarser grid to be used to represent the orbitals and (nuclear) potential. Without this approximation the valence eigenstates are highly oscillatory in close proximity to the nucleus in order to retain orthogonality with the core states.

In ONETEP the pseudopotentials used are norm-conserving and can be split into local pseudopotential and non-local pseudopotential terms,

$$\hat{v}_{\text{ps}} = \hat{v}_{\text{loc}} + \hat{v}_{\text{nl}}. \tag{2.5.9}$$

The local pseudopotential is relatively simple in form and is calculated from the atomic positions to yield a potential v_{loc}. The form of the non-local pseudopotential is more complicated. The Kleinman- Bylander form [73] is used in ONETEP, this has the expression,

$$\hat{v}_{\text{nl}} = \sum_{I}^{N_{\text{at}}} \sum_{l=0}^{l_{\text{max}}} \sum_{m=-l}^{l} \frac{\left| \hat{v}_l^{(I)} \chi_{lm}^{(I)} \right\rangle \left\langle \chi_{lm}^{(I)} \hat{v}_l^{(I)} \right|}{\left\langle \chi_{lm}^{(I)} \left| \hat{v}_l^{(I)} \right| \chi_{lm}^{(I)} \right\rangle} \tag{2.5.10}$$

where $\chi_{lm}^{(I)}$ is an eigenstate of the atomic pseudo-Hamiltonian and $\hat{v}_l$ is an operator for a potential with angular momentum l. The projector states,

$$\left| \xi_{lm}^{(I)} \right\rangle = \left| \hat{v}_l^{(I)} \chi_{lm}^{(I)} \right\rangle \tag{2.5.11}$$

are strictly localised in real space within small spherical core regions that commonly have a radius of around $1.5a_0$. The calculation of the non-local pseudopotential energy involves integrals between projector states and NGWFs,

$$\mathfrak{s}_{\alpha,lm}^{(I)} = \left\langle \phi_\alpha \middle| \xi_{lm}^{(I)} \right\rangle \tag{2.5.12}$$

which can be calculated as a sum over grid points in an FFT box. The non-local pseudopotential energy can therefore be calculated [25],

$$E_{\text{nl}} = K^{\alpha\beta} \sum_{I}^{N_{\text{at}}} \sum_{l=0}^{l_{\text{max}}} \sum_{m=-l}^{l} \mathfrak{s}_{\beta,lm}^{(I)} \frac{1}{\left\langle \chi_{lm}^{(I)} \left| \hat{v}_l^{(I)} \right| \chi_{lm}^{(I)} \right\rangle} \left(\mathfrak{s}_{\alpha,lm}^{(I)} \right)^\dagger. \tag{2.5.13}$$

2.6 Idempotency

An idempotent matrix is equal to its square, i.e. $\mathbf{A} = \mathbf{A}^2$. The idempotency constraint (discussed in section 1.6.1) can be expressed using the density kernel as [66],

$$K^{\alpha\beta} = K^{\alpha\gamma} S_{\gamma\delta} K^{\delta\beta}. \tag{2.6.1}$$

This can be derived by writing the density matrix idempotency condition (from equation 1.6.11),

$$\rho_1^2(\mathbf{r}, \mathbf{r}') = \int d\mathbf{r}'' \rho_1(\mathbf{r}, \mathbf{r}'') \rho_1(\mathbf{r}'', \mathbf{r}') \tag{2.6.2}$$

in terms of the density kernel using equation 2.1.4,

$$
\begin{aligned}
\phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta^*(\mathbf{r}') &= \int d\mathbf{r}'' \phi_\alpha(\mathbf{r}) K^{\alpha\gamma} \phi_\gamma^*(\mathbf{r}'') \phi_\delta(\mathbf{r}'') K^{\delta\beta} \phi_\beta^*(\mathbf{r}') \\
&= \phi_\alpha(\mathbf{r}) K^{\alpha\gamma} \left(\int d\mathbf{r}'' \phi_\gamma^*(\mathbf{r}'') \phi_\delta(\mathbf{r}'') \right) K^{\delta\beta} \phi_\beta^*(\mathbf{r}') \\
&= \phi_\alpha(\mathbf{r}) K^{\alpha\gamma} S_{\gamma\delta} K^{\delta\beta} \phi_\beta^*(\mathbf{r}'),
\end{aligned} \tag{2.6.3}
$$

which can be simplified to give equation 2.6.1. It is not feasible in a linear-scaling scheme to enforce idempotency by calculating all the eigenvalues of the system (since the cost of matrix diagonalisation scales cubically with matrix size), and so therefore another approach is required.

2.6.1 Penalty functional

The following penalty functional first proposed by McWeeny [74] is the origin of many methods for imposing idempotency,

$$P[\rho] = \text{tr}\left[\left(\rho^2 - \rho \right)^2 \right] \tag{2.6.4}$$

$$= \sum_n \left(f_n^2 - f_n \right)^2, \tag{2.6.5}$$

where f_n are the orbital occupancies. This penalty functional can be added to the energy functional and then the resulting functional,

$$Q[\rho] = E[\rho] + \alpha P[\rho], \tag{2.6.6}$$

where α is a parameter controlling the strength of the penalty functional, can be minimised. The problem with such an implementation is that the minimum of $Q[\rho]$ cannot occur at $f_n = 1$ (since the slope of $E[\rho]$ here is given by the Kohn-Sham eigenvalue), so idempotency is only approximately imposed [75].

2.6.2   Purification

In ONETEP a purifying transformation is used to ensure idempotency. The McWeeny purifying transform [74], which has the form,

$$\mathbf{R}_{n+1} = 3\mathbf{R}_n^2 - 2\mathbf{R}_n^3 \qquad (2.6.7)$$

where $\mathbf{R}_n$ are matrices forming a sequence that converges quadratically to idempotency. Convergence to idempotency is guaranteed when the eigenvalues of $\mathbf{R}$ are in the intervals $\left(\frac{1-\sqrt{3}}{2}, \frac{1}{2}\right)$ and $\left(\frac{1}{2}, \frac{1+\sqrt{3}}{2}\right)$. There are two types of purification that are applied to the density kernel in ONETEP, adaptive purification and canonical purification.

In adaptive purification the extremal occupancies are checked and if they lie outside the desired range then P (defined in equation 2.6.4) is minimised using steepest descents (with a line search to find the optimal step length) until the occupancies are within the desired range.

Canonical purification [76] is a method to obtain the eigenvalues of the ground state density matrix of a fixed Hamiltonian in a non-self consistent manner. The purification transformation (equation 2.6.7) is applied repeatedly until the band-structure energy, the energy of the non-interacting system, $\mathrm{tr}[\mathbf{KH}]$ converges [76].

2.6.3   LNV functional

In the Li-Nunes and Vanderbilt (LNV) method [61] the density matrix, ρ, is defined in terms of an auxiliary matrix, σ, using the purification transformation,

$$\rho = 3\sigma^2 - 2\sigma^3 \qquad (2.6.8)$$

where σ is defined as

$$\sigma(\mathbf{r}, \mathbf{r}') = \phi_\alpha(\mathbf{r}) L^{\alpha\beta} \phi_\beta^*(\mathbf{r}'), \qquad (2.6.9)$$

and the auxiliary kernel $\mathbf{L}$ (of which $L^{\alpha\beta}$ is a matrix element) is related to the density kernel by,

$$\mathbf{K} = 3\mathbf{LSL} - 2\mathbf{LSLSL}. \qquad (2.6.10)$$

If the total energy $E[\sigma]$ is minimised by optimising the auxiliary kernel $\mathbf{L}$ then the density matrix tends towards idempotency [75].
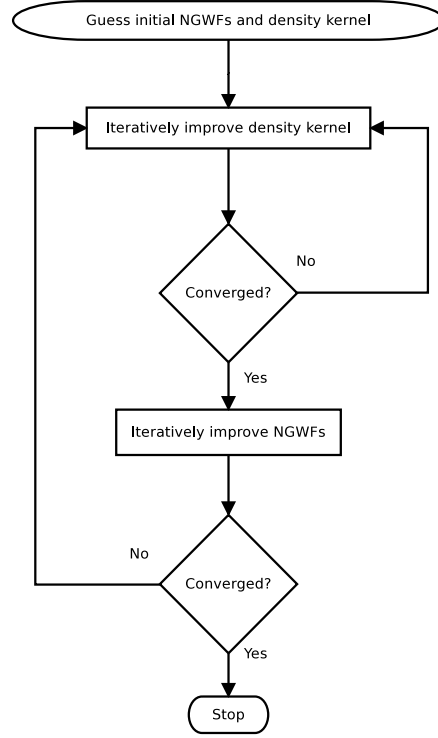
Figure 2.4: Flowchart of the optimisation of the energy in ONETEP.

2.7 Optimisation of the energy

As seen in equation 2.1.4, the density matrix can be expressed in terms of the density kernel and the NGWFs,

$$\rho_1(\mathbf{r}, \mathbf{r}') = \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta^*(\mathbf{r}'). \tag{2.7.1}$$

In ONETEP the density matrix is optimised self-consistently by minimising energy with respect to both the NGWFs and density kernel. Optimising both is necessary to obtain high accuracy while still retaining linear-scaling computational demands. Figure 2.4 illustrates the two loops of the energy optimisation scheme. Firstly the energy is minimised with respect to the density kernel, this procedure is discussed in section 2.8. The NGWFs are then optimised using the procedure in section 2.9. After each update of the NGWFs the density kernel is reoptimised.

2.8 Density kernel optimisation

At the start of a ONETEP calculation an initial guess for the density kernel is obtained from the inverse of the overlap matrix. To allow linear-scaling this operation is performed using the Hotelling algorithm [75,77]. This is improved by performing Palser-Manolopoulos canonical purification [76] on the initial guess. Further penalty functional minimisation is performed before the main LNV optimisation begins. The LNV functional implemented in ONETEP performs direct energy minimisation, rather than optimisation of the band

structure energy with density mixing.

2.8.1   Density kernel gradient

The density kernel gradient may be obtained by taking the derivative with respect to the density kernel of equation 2.5.3. Use will be made of the following result,

$$\frac{\partial K^{\alpha\beta}}{\partial K^{\eta\theta}} = \delta_\eta^\alpha \delta_\theta^\beta. \tag{2.8.1}$$

Since some of the components of the energy depend on the density an expression for the derivative of the density with respect to the density kernel is required,

$$\frac{\partial n(\mathbf{r})}{\partial K^{\eta\theta}} = \delta_\eta^\alpha \delta_\theta^\beta \phi_\alpha(\mathbf{r})\phi_\beta^*(\mathbf{r}). \tag{2.8.2}$$

For the density dependent parts of the energy,

$$E_{\text{Hxc}} = E_{\text{H}} + E_{\text{xc}} \tag{2.8.3}$$

the gradient is,

$$\frac{\partial E_{\text{Hxc}}}{\partial K^{\eta\theta}} = \int \frac{\delta E_{\text{Hxc}}}{\delta n(\mathbf{r})} \frac{\partial n(\mathbf{r})}{\partial K^{\eta\theta}} d\mathbf{r} \tag{2.8.4}$$

$$= 2 \int \delta_\eta^\alpha \delta_\theta^\beta \phi_\eta(\mathbf{r}) \left[v_{\text{H}} + v_{\text{xc}}\right] \phi_\theta^*(\mathbf{r}) d\mathbf{r} \tag{2.8.5}$$

$$= 2 \left\langle \phi_\beta \left| \hat{v}_{\text{H}} + \hat{v}_{\text{xc}} \right| \phi_\alpha \right\rangle. \tag{2.8.6}$$

The expression for the gradient is therefore [66],

$$\frac{\partial E[n]}{\partial K^{\eta\theta}} = 2\delta_\eta^\alpha \delta_\theta^\beta \left\langle \phi_\beta \left| -\tfrac{1}{2}\hat{\nabla}_\mathbf{r}^2 + \hat{v}_{\text{loc}} + \hat{v}_{\text{nl}} + \hat{v}_{\text{xc}} + \hat{v}_{\text{H}} \right| \phi_\alpha \right\rangle \tag{2.8.7}$$

$$= 2 \left\langle \phi_\theta \left| -\tfrac{1}{2}\hat{\nabla}_\mathbf{r}^2 + \hat{v}_{\text{loc}} + \hat{v}_{\text{nl}} + \hat{v}_{\text{xc}} + \hat{v}_{\text{H}} \right| \phi_\eta \right\rangle \tag{2.8.8}$$

$$= 2H_{\theta\eta}. \tag{2.8.9}$$

The expression in equation 2.8.9 is a simple "idealised" version of the gradient expressions that are used in ONETEP. In ONETEP more sophisticated expressions are used which include linear-scaling idempotency enforcing functionals for the energy such as the LNV energy functional [61]. In practice in ONETEP the gradient of the energy is obtained with respect to the auxiliary density kernel, **L** [25],

$$\frac{\partial E[n]}{\partial L^{\alpha\beta}} = 2 \frac{\partial K^{\eta\theta}}{\partial L^{\alpha\beta}} H_{\theta\eta} \tag{2.8.10}$$

where,

$$\frac{\partial K^{\eta\theta}}{\partial L^{\alpha\beta}} = 3\left(\delta_\alpha^\eta S_{\beta\gamma} L^{\gamma\theta} + L^{\eta\gamma} S_{\gamma\alpha} \delta_\beta^\theta\right)$$
$$- 2\left(\delta_\alpha^\eta S_{\beta\gamma} L^{\gamma\kappa} S_{\kappa\zeta} L^{\zeta\theta} + L^{\eta\gamma} S_{\gamma\alpha} S_{\beta\kappa} L^{\kappa\theta} + L^{\eta\gamma} S_{\gamma\kappa} L^{\kappa\zeta} S_{\zeta\alpha} \delta_\beta^\theta\right) \quad (2.8.11)$$

so,

$$\frac{\partial E[n]}{\partial L^{\alpha\beta}} = 6(SLH + HLS)_{\beta\alpha} - 4(SLSLH + SLHLS + HLSLS)_{\beta\alpha}. \quad (2.8.12)$$

As the NGWFs are non-orthogonal functions the tensor properties must be respected during the minimisation procedure. (With an orthogonal set of functions there is no distinction between covariant and contravariant quantities.) Since the auxiliary density kernel is a contravariant quantity a contravariant gradient is required for the search direction in the minimisation procedure [78,79]. This tensorially correct gradient is given by [25],

$$G^{\alpha\beta} = (S^{-1})^{\alpha\gamma} \frac{\partial E[n]}{\partial L^{\gamma\eta}} (S^{-1})^{\eta\beta}$$
$$= 6(LHS^{-1} + S^{-1}HL)^{\alpha\beta} - 4(LSLHS^{-1} + LHL + S^{-1}HLSL)^{\alpha\beta}. \quad (2.8.13)$$

2.8.2 Normalisation constraint

A normalisation constraint is applied to the density kernel to ensure that the number of electrons in the system remains correct. A Lagrange multiplier, the chemical potential μ, is used to ensure the correct number of electrons at the minimum of the functional used. Instead of minimising the energy, the grand potential ($\Omega = E - \mu N$) is minimised. In ONETEP three methods for imposing normalisation during the LNV steps are used [75]. Firstly constraining the purified electron number such that for a spin unpolarised system,

$$2\,\mathrm{tr}(\rho) = 2\,\mathrm{tr}(\mathbf{KS})$$
$$= 2\,\mathrm{tr}(3\mathbf{LSL} - 2\mathbf{LSLSL})$$
$$= N_e. \quad (2.8.14)$$

Secondly constraining the unpurified electron number such that,

$$2\,\mathrm{tr}(\sigma) = 2\,\mathrm{tr}(\mathbf{LS}) = N_e. \quad (2.8.15)$$

And finally using a modified purification transformation, where the kernel always remains scaled to the correct number of electrons,

$$\rho = N_e \frac{3\sigma^2 - 2\sigma^3}{2\,\mathrm{tr}\left(3\sigma^2 - 2\sigma^3\right)}. \quad (2.8.16)$$

The derivative of the electron number with respect to variation in the auxiliary density kernel can also be calculated,

$$\frac{\partial N_e}{\partial L^{\alpha\beta}} = 2\frac{\partial K^{\eta\theta}}{\partial L^{\alpha\beta}}S_{\theta\eta}$$

$$= 12(SLS - SLSLS)_{\beta\alpha} \tag{2.8.17}$$

using equation 2.8.11 [25]. The contravariant form is,

$$\Lambda^{\alpha\beta} = (S^{-1})^{\alpha\gamma}\frac{\partial N_e}{\partial L^{\gamma\eta}}(S^{-1})^{\eta\beta}$$

$$= 12(L - LSL)^{\alpha\beta}. \tag{2.8.18}$$

It is desirable that the electron number does not change during the calculation.

2.9   NGWF optimisation

The NGWFs in ONETEP are optimised by varying the psinc expansion coefficients. Using the simple steepest descents method this would have the form,

$$c'_{p\gamma} = c_{p\gamma} - \lambda\frac{\partial E[n]}{\partial c_{p\gamma}} \tag{2.9.1}$$

where λ is a parameter chosen to minimise the energy (for example by a line search). Since the psinc coefficients in ONETEP are real the complex conjugates have been omitted. Therefore, the gradient of the energy with respect to the psinc expansion coefficients is required. By calculating derivative of the energy with respect to an NGWF (ϕ_γ), the derivatives with respect to each of the psinc expansion coefficients for that NGWF ($c_{p\gamma}$) are obtained. The derivative of the kinetic and pseudopotential energies is [25],

$$\frac{\delta E_{\mathrm{kin,ps}}}{\delta\phi_\alpha^*(\mathbf{r})} = 2\frac{\delta K^{\mu\nu}}{\delta\phi_\alpha^*}\left\langle\phi_\nu^*\left|-\tfrac{1}{2}\hat{\nabla}^2 + \hat{v}_{\mathrm{loc}} + \hat{v}_{\mathrm{nl}}\right|\phi_\mu\right\rangle + 2\left(-\tfrac{1}{2}\hat{\nabla}^2 + \hat{v}_{\mathrm{loc}} + \hat{v}_{\mathrm{nl}}\right)\phi_\mu K^{\mu\alpha}. \tag{2.9.2}$$

The derivative of the density with respect to variations in the localised functions is given by [25],

$$\frac{\delta n(\mathbf{r}')}{\delta\phi_\alpha^*(\mathbf{r})} = 2\phi_\nu(\mathbf{r}')K^{\nu\alpha}\delta(\mathbf{r} - \mathbf{r}') + 2\phi_\nu(\mathbf{r}')\frac{\delta K^{\mu\nu}}{\delta\phi_\alpha^*}\phi_\mu^*. \tag{2.9.3}$$

The derivative of the Hartree and exchange-correlation energies (which depend on the density) is [25],

$$\frac{\delta E_{\mathrm{Hxc}}}{\delta\phi_\alpha^*(\mathbf{r})} = \int\frac{\delta E_{\mathrm{Hxc}}}{\delta n(\mathbf{r}'}\frac{\delta n(\mathbf{r}')}{\delta\phi_\alpha^*(\mathbf{r})}d\mathbf{r}' \tag{2.9.4}$$

$$= \int\left(v_{\mathrm{H}}(\mathbf{r}') + v_{\mathrm{xc}}(\mathbf{r}')\right)\frac{\delta n(\mathbf{r}')}{\delta\phi_\alpha^*(\mathbf{r})}d\mathbf{r}' \tag{2.9.5}$$

using equation 2.9.3 gives,

$$=2\left(v_{\mathrm{H}}(\mathbf{r}') + v_{\mathrm{xc}}(\mathbf{r}')\right)\phi_{\nu}(\mathbf{r}')K^{\nu\alpha} + 2\frac{\delta K^{\mu\nu}}{\delta\phi_{\alpha}^{*}}\langle\phi_{\nu}^{*}|\,\hat{v}_{\mathrm{H}} + \hat{v}_{\mathrm{xc}}\,|\phi_{\mu}\rangle. \tag{2.9.6}$$

Combining equations 2.9.2 and 2.9.6 gives the gradient of the total energy with respect to variations in the NGWFs [25],

$$\frac{\delta E}{\delta\phi_{\alpha}^{*}(\mathbf{r})} = 2\hat{H}\phi_{\mu}(\mathbf{r})K^{\mu\alpha} + 2\frac{\delta K^{\mu\nu}}{\delta\phi_{\alpha}^{*}}H_{\nu\mu}. \tag{2.9.7}$$

Since the purified density kernel is used, which depends on the overlap matrix which in turn depends on the NGWFs, a further term in the NGWF gradient is necessary,

$$\frac{\delta E}{\delta\phi_{\alpha}^{*}(\mathbf{r})} = 2\left[\hat{H}\phi_{\beta}(\mathbf{r})K^{\beta\alpha} + \phi_{\beta}(\mathbf{r})Q^{\beta\alpha}\right] \tag{2.9.8}$$

where,

$$\mathbf{Q} = 3\mathbf{LHL} - 2\mathbf{LSLHL} - 2\mathbf{LHLSL}. \tag{2.9.9}$$

As discussed in section 2.8.1 tensorially correct gradients are required. Since the NGWFs are covariant functions, the covariant gradient [25],

$$g_{\alpha}(\mathbf{r}) = \frac{\delta E}{\delta\phi_{\gamma}^{*}(\mathbf{r})}S_{\gamma\alpha} \tag{2.9.10}$$

$$= 2\left[\hat{H}\phi_{\beta}(\mathbf{r})(KS)^{\beta}{}_{\alpha} + \phi_{\beta}(\mathbf{r})(QS)^{\beta}{}_{\alpha}\right], \tag{2.9.11}$$

is required.

In ONETEP the energy is minimised in two nested loops. In the inner loop the NGWFs are fixed and the energy is minimised with respect to the density kernel (see section 2.8). In the outer loop the form of the NGWFs is optimised by varying the psinc expansion coefficients using a conjugate gradient method. The conjugate gradient method is a more sophisticated method than the steepest descent method which takes into account the search directions used in the previous iterations. This improves the convergence behaviour in many cases, since better search directions are chosen [25]. This minimisation is constrained by the requirements that the ground state one-particle density matrix remains idempotent and that the integral over all space of the density is equal to the number of electrons [66]. The derivative of the electron number with respect to variations in the localised functions is [25],

$$\frac{\delta N_{e}}{\delta\phi_{\alpha}^{*}(\mathbf{r})} = 2\left[\frac{\delta K^{\eta\theta}}{\delta\phi_{\alpha}^{*}(\mathbf{r})}S_{\theta\eta} + K^{\eta\theta}\frac{\delta S_{\theta\eta}}{\delta\phi_{\alpha}^{*}(\mathbf{r})}\right] \tag{2.9.12}$$

$$= 12\phi_{\beta}(\mathbf{r})(LSL - LSLSL)_{\beta\alpha}. \tag{2.9.13}$$

The tensorially correct covariant derivative of the electron number is [25],

$$\lambda_\alpha(\mathbf{r}) = \frac{\delta N_e}{\delta \phi_\gamma^*(\mathbf{r})} S_{\gamma\alpha} \tag{2.9.14}$$

$$= 12\phi_\beta(\mathbf{r})(LSLS - LSLSLS)^\beta{}_\alpha. \tag{2.9.15}$$

2.9.1 Reciprocal space preconditioning

Electronic structure calculations using large basis sets such as plane waves can suffer from kinetic energy ill-conditioning [80]. The efficiency of the steepest descents and conjugate gradient methods is related to the condition number which is the ratio between the extremal curvatures of the function around the minimum. If the condition number is small, the range of curvatures is small and minimisation is efficient. However if the condition number is large then the system is said to be ill-conditioned and the number of iterations that are required to reach convergence can be high [65].

The eigenvalues of the Hamiltonian determine the curvatures of the total energy functional. A wide range of eigenvalues, which is to be expected with a large systematic basis set, is going to lead to a large condition number. The kinetic energy dominates in the high energy eigenstates and is a significant source of ill-conditioning. These eigenstates will be unoccupied so do not significantly contribute to the total energy but do broaden the eigenspectrum. For lower energy eigenstates where the kinetic and potential energies are more balanced this is less of a problem. The ill-conditioning can be reduced by removing the effect of the kinetic energy operator for high energy states. This makes the eigenstates more degenerate and therefore reduces the width of the eigenspectrum, without significantly affecting the lower energy states. This process is known as preconditioning. In a plane wave basis preconditioning maybe applied by multiplying the steepest descents vector by a diagonal preconditioning matrix that behaves as the inverse kinetic energy at high wave-vectors and as a constant at low wave-vectors [81]. This is qualitatively similar to the exact preconditioner for the model Hamiltonian $\hat{X}$ given by,

$$\hat{X} = 1 - \frac{\nabla^2}{G_0^2} \tag{2.9.16}$$

where G_0 is a constant (a judiciously chosen wavevector) [80]. Preconditioners for this model Hamiltonian can also be derived for other basis sets [65].

2.9.1.1 Psinc basis preconditioning

Applying the model Hamiltonian in equation 2.9.16 to the psinc functions gives matrix elements,

$$x_{\mu\nu} = \int D_\mu^*(\mathbf{r})\hat{X}D_\nu(\mathbf{r})d\mathbf{r} \tag{2.9.17}$$

$$= s_{\mu\nu} + \frac{t_{\mu\nu}}{G_0^2} \tag{2.9.18}$$

where,

$$s_{\mu\nu} = \int D_\mu^*(\mathbf{r})D_\nu(\mathbf{r})d\mathbf{r} \qquad (2.9.19)$$

$$= w\delta_{\mu\nu} \qquad (2.9.20)$$

and,

$$t_{\mu\nu} = -\int D_\mu^*(\mathbf{r})\hat{\nabla}^2 D_\nu(\mathbf{r})d\mathbf{r} \qquad (2.9.21)$$

$$= \frac{w}{N}\sum_p^{\max} |\mathbf{G}_p|^2 \, e^{i\mathbf{G}_p\cdot(\mathbf{r}_\mu-\mathbf{r}_\nu)} \qquad (2.9.22)$$

where $w = V/N$, is the grid point weight.

The standard expression for the line minimisation for the psinc coefficients in equation 2.9.1 can be preconditioned by premultiplying the gradient by $\mathbf{x}^{-1}$ and postmultiplying by the overlap matrix, $\mathbf{S}$, to account for the non-orthogonality of the NGWFs, to give,

$$c_{i\gamma}' = c_{i\gamma} - \lambda\sum_j x_{ij}^{-1}\frac{\partial E[n]}{\partial c_{j\beta}}S_{\beta\gamma} \qquad (2.9.23)$$

$$= c_{i\gamma} - \lambda\sum_j x_{ij}^{-1}g_j^{\;\beta}S_{\beta\gamma}. \qquad (2.9.24)$$

For an orthogonal basis set a unitary transformation, $\mathbf{F}$ can be applied that diagonalises $\mathbf{x}$,

$$\tilde{\mathbf{x}} = \mathbf{F}\mathbf{x}\mathbf{F}^\dagger \qquad (2.9.25)$$

$$\tilde{x}_{pq} = \xi_p\delta_{pq} \qquad (2.9.26)$$

where the ξ_p are the eigenvalues of $\mathbf{x}$. Denoting transformed variables with a tilde, equation 2.9.24 becomes,

$$\tilde{c}_{p\alpha}' = \tilde{c}_{p\alpha} - \lambda\frac{1}{\xi_p}\tilde{g}_p^{\;\beta}S_{\beta\alpha}. \qquad (2.9.27)$$

In the psinc basis set the operator $\hat{F}$ that diagonalises $\mathbf{x}$ is the discrete Fourier transform,

$$\tilde{b}_p = \sum_j F_{pj}b_j \qquad (2.9.28)$$

$$= \frac{1}{\sqrt{N}}\sum_j b_j e^{i\mathbf{G}_p\cdot\mathbf{r}_j} \qquad (2.9.29)$$

$$b_i = \sum_p F_{ip}^\dagger\tilde{b}_p \qquad (2.9.30)$$

$$= \frac{1}{\sqrt{N}}\sum_p \tilde{b}_p e^{-i\mathbf{G}_p\cdot\mathbf{r}_i} \qquad (2.9.31)$$

where b_i are values on the real space grid and $\tilde{b}_p$ are values on the reciprocal space grid. Using the definitions above it can be shown that,

$$\tilde{x}_{pq} = \sum_{ij} F_{pi} x_{ij} F_{jq}^{\dagger} \tag{2.9.32}$$

$$= w \left(1 + \frac{|\mathbf{G}_p|^2}{G_0^2} \right) \delta_{pq}. \tag{2.9.33}$$

Therefore the eigenvalues ξ_p of $\mathbf{x}$ are given by,

$$\xi_p = w \left(1 + \frac{|\mathbf{G}_p|^2}{G_0^2} \right). \tag{2.9.34}$$

Using equation 2.9.33, the final expression for the preconditioned line minimisation is,

$$\tilde{c}'_{p\alpha} = \tilde{c}_{p\alpha} - \frac{\lambda}{w} \frac{G_0^2}{G_0^2 + |\mathbf{G}_p|^2} \tilde{g}_p^{\,\beta} S_{\beta\alpha}. \tag{2.9.35}$$

In ONETEP the preconditioning of the gradient is applied in reciprocal space using FFTs [25, 65].

2.10   Sparse matrices

Linear-scaling requires that the cost of the matrix operations scale linearly with system size. A dense matrix contains $\mathcal{O}(N^2)$ elements so linear-scaling is not possible. The matrices used in ONETEP are sparse (for sufficiently large systems with a kernel cutoff) and contain $\mathcal{O}(N)$ elements. Algorithms have been developed to operate on these matrices with $\mathcal{O}(N)$ computational cost (for large systems). In ONETEP the sparse matrices are structured in terms of atomblocks with only the atomblocks with non-zero elements being stored. This atomblocked structure is a natural because usually the elements of an atomblock are either all zero or all non-zero, for example in the overlap matrix all the NGWFs on atom A overlap with all those on atom B or none of them do, since the NGWF radii are equal. The notable exceptions to this are diagonal matrices (for example the matrix of Kleinman-Bylander denominators in equation 2.5.13). The elements of the non-zero atomblocks of a sparse matrix are stored in a linear array. There is a separate index, shared between sparse matrices with the same structure, that is used to access the arrays. For small systems, dense matrix algebra is faster as in these systems the matrices are less sparse and so the cost of the sparse indexing outweighs the benefit of ignoring zero atom blocks [70].

2.10.1   Sparsity patterns

The various sparse matrices have differing sparsity patterns. The sparsity pattern of the NGWF-NGWF overlap matrix is determined by the overlap of NGWFs, see figure 2.5. The NGWF-projector and projector-NGWF overlap matrices, which were defined in section 2.5.2, have non-zero elements for overlapping projectors and NGWFs. The non-local
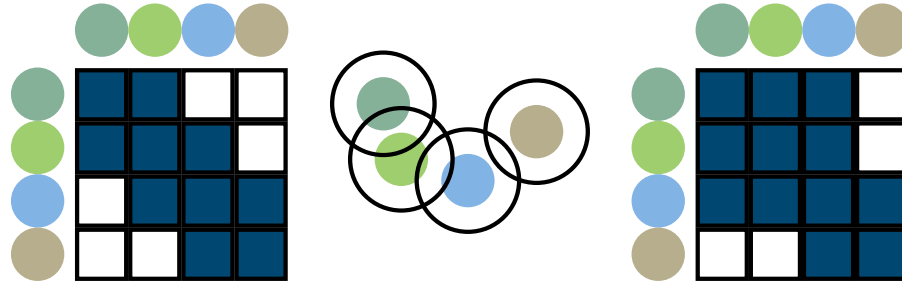
Figure 2.5: The origin of the sparsity pattern of the overlap (left) and Hamiltonian (right) matrices. The filled squares represent non-zero atom blocks. The coloured spheres represent the pseudopotential projectors (see section 2.5.2), with the colours labelling the atoms. The black circles represent the boundaries of the NGWF spheres on each atom. The overlap sparsity is defined by which NGWF spheres overlap. The Hamiltonian additionally has non-zero atomblocks where a pair of atoms with non-overlapping NGWFs both overlap with the same projector.

pseudopotential potential matrix is formed by the product of the NGWF-projector and projector-NGWF overlap matrices. Therefore the sparsity pattern is governed by the pairs of NGWFs that overlap with a common core projector (from a pseudopotential) but don't necessarily overlap themselves. The Hamiltonian is the union of the sparsity patterns of the overlap and non-local potential matrices, see figure 2.5. In the density kernel matrix only atomblocks corresponding to atoms separated by less than a given distance are non zero. This distance is known in ONETEP as the kernel cutoff and is an adjustable parameter. The truncation of the density kernel in this way is an approximation justified by the near-sightedness principle discussed in section 1.7.1. This approximation is required for linear-scaling to be obtained.

2.11 Parallel strategy

In a parallel ONETEP calculation the data and computational work are shared amongst the cores. Many parts of the calculation involve the communication of data between cores. Certain sparse matrix operations and whole simulation cell FFTs also involve all cores. An efficient parallel strategy is required to make best use of large numbers of cores. The parallel performance of ONETEP depends on the system, low density systems such as nanotubes scale better than dense solids, and the number of cores used [2, 70, 82].

2.11.1 Simulation cell

The simulation cell is equally divided into slabs along the z-axis with one slab being placed on each core. Quantities such as the electron density that are defined across the entire simulation cell are stored in this distributed way. The calculation of the Ewald energy and the Hartree potential involve FFTs of the entire simulation cell. In reciprocal space the simulation cell is divided into slabs along the x-axis [2]. During the construction of the electron density FFT boxes are deposited in the simulation cell; this operation in general involves multiple cores.

2.11.2   Atoms and NGWFs

The atoms and their corresponding NGWFs are distributed among the cores in the order
that a Peano space-filling fractal curve [83] visits each atom. Atoms that are close together
in space will be close together in the ordered list of atoms that results. Each core is
allocated a set of atoms in the order determined by the list such that on each core there is
a roughly equal number of NGWFs. This ordering helps to minimise communication since
many parts of a ONETEP calculation (such as the calculation of the density and the kinetic
energy) involve an NGWF and its overlapping neighbours [2].

2.11.2.1   Batches

There is a balance to be struck between the amount of communication and memory usage.
Communication could be minimised by sending an NGWF once and then performing all
operations that require that NGWF on the receiving core. To do this would require a
substantial amount of memory because it would require a large number of FFT boxes to be
held in memory at one time. For example in the calculation of the density if an NGWF is
sent to a core only once then an FFT box for each of the NGWFs local to the receiving core
that overlap with the received NGWF would need to be stored in memory. The opposite
extreme would be use the received NGWF only for a single operation and recommunicate
it each time it is required. In ONETEP a system of batches has been implemented where
the NGWFs on the local core are divided into batches. When an NGWF is received all the
operations for the current batch that require that NGWF are performed, so the NGWF
only needs to be communicated once per batch. The batch size is an adjustable parameter
in the input file of a ONETEP calculation. Small batches have low memory demands, but
require much communication; larger batches reduce the amount of communication but
increase memory usage [2].

2.11.2.2   Planned communication

The original communications model used in ONETEP had NGWFs being sent just before
they were required. This entailed a double loop: the outer loop over core-core blocks;
the inner loop over the NGWF pairs in the block. The outer loop was implemented by
considering first the diagonal blocks and then those blocks immediately below the these,
repeating until all blocks below the diagonal were processed. A similar procedure was then
followed with the blocks above the diagonal. A large degree of synchronisation between the
cores was enforced by this method which reduced the parallel efficiency [2]. An improvement
to this method has been developed which involves the construction of a plan, a list of the
NGWF pairs, determined from the sparse matrix structure, that the calculation for the
current batch on a particular core involves. The plan is then shared between the cores and
from this the required communication pattern can be determined. The plan can then be
used to determine at which NGWFs each core needs to send and receive at each step of the
plan. Since in ONETEP non-blocking sends are used (i.e. there is no need to wait for the
NGWF to be received) the NGWFs can be sent before they are required by looking ahead

in the plan. This further improves the parallel performance because it reduces the amount of time each core wastes waiting for another core to send an NGWF [70].

2.11.3 Sparse matrices

The sparse matrices used in ONETEP (see section 2.10) are distributed between the cores, such that each core holds all the non-zero atom blocks located in the block columns of the atoms located on that core (see above). The use of a space filling curve to distribute the atoms means that the non-zero atom blocks tend to be concentrated along the diagonal of the matrix. For example in the overlap matrix (where the non-zero atomblocks correspond to atoms that overlap), the space filling curve order means that overlapping atoms will be ordered close to each other in the list and therefore these atomblocks will be close to the diagonal [2]. Many quantities can be calculated using only the elements in atomblocks local to the core. However where matrix operations such as transposition or multiplication are required communication is necessary. Recent work has sought to increase the efficiency of parallel sparse matrix algebra in ONETEP [70]. One improvement has been to reorder the sequence in which portions of the matrix are manipulated taking into account the distributed sparsity pattern. This results in a more balanced load between the processors [70].

Chapter 3

Empirical corrections for dispersion in DFT

While Density Functional Theory (DFT) allows accurate quantum mechanical simulations from first principles in molecules and solids, commonly used exchange-correlation density functionals provide a very incomplete description of dispersion interactions. One way to include such interactions is to augment the DFT energy expression by damped London energy expressions. This is also referred to as the DFT-D approach [84]. Several variants of this have been developed for this task, which are discussed and compared below. These schemes have been implemented in the ONETEP program, which is capable of DFT calculations with computational cost that increases linearly with the number of atoms. All the parameters involved in this implementation of the dispersion correction have been optimized, with the aim of simulating biomolecular systems. Tests show that in cases where dispersion interactions are important this approach produces binding energies and molecular structures of a quality comparable with high-level wavefunction-based approaches. This work has been published [3].

3.1   Dispersion forces

Dispersion (or London [85]) forces between atoms arise due to the instantaneous dipoles, brought about by the fluctuations in the positions of the electrons. These induce dipoles in a nearby atom or molecule which then in turn interact with the dipole on the original atom or molecule. The energy of these attractive interactions can be approximated by the London formula, which may be derived by perturbation theory. For a pair of atoms the London formula has the following form:

$$E(R_{IJ}) = -\frac{C_{IJ}}{R_{IJ}^6} \tag{3.1.1}$$

$$C_{IJ} = \frac{3}{2}\alpha'_I\alpha'_J\frac{I_II_J}{I_I+I_J} \tag{3.1.2}$$

where I_J is the ionisation potential of atom J, α'_J is its polarisability volume and R_{IJ} is the (internuclear) distance between atoms I and J [86]. The potential between two atoms

due to dispersion forces can be modelled by the Lennard-Jones formula:

$$E(R_{IJ}) = \frac{C_{12,IJ}}{R_{IJ}^{12}} - \frac{C_{6,IJ}}{R_{IJ}^{6}}$$
(3.1.3)

The attractive R^{-6} term comes from the London formula [85,86], while the R^{-12} term represents a repulsive potential. This is required because at closer range the electrons on each atom repel each other. The form of the repulsive term however is mainly a computational convenience as for example an exponential form e^{-R/R_0} is more accurate approximation but also more costly [87]. Dispersion forces are very weak (for example, binding energies for noble gas atom pairs are vary between 0.022 kcal mol^{-1} for a Helium dimer and 0.56 kcal mol^{-1} for an Xenon dimer) but can collectively be responsible in determining the geometry of many molecules and solids. Important cases include the stacking interactions between π electron systems, such as between graphene sheets in graphite, and base pairs in DNA. In biomolecular simulations they are often described as "hydrophobic" interactions and often play an important role in determining structure and energetics, therefore they need to be described as well as other non-covalent interactions (such as ion-pairs or hydrogen bonding).

3.1.1   Description by DFT

Common DFT functionals are unable to describe dispersion interactions. The difficulty of describing dispersion with DFT is not an intrinsic failure of the theory as the exact exchange-correlation energy functional would be able to describe all such interactions correctly. However its form is unknown and approximations are required [88]. Commonly these approximations are based on the local electron density and its gradient, and therefore give a poor description of interactions occurring outside the area of electronic overlap, which is the case with dispersion interactions [89–91]. DFT does provide an adequate description of the repulsive interactions at closer range, where the electron densities overlap. The local density approximation will often appear to show binding (in, for example, a noble gas dimer [89, 92, 93]) but this binding is spurious as it results from the exchange part of the functional, whereas dispersion is a dynamical correlation effect [88, 94]. Gradient corrected functionals will usually show no binding at all, although basis set superposition error may often give rise to the appearance of weak binding [95].

3.2   Empirical correction schemes for dispersion in DFT

Density functionals capable of explicitly including dispersion interactions are being developed by several groups [96–99]. While these functionals are promising, they have a considerably higher computational cost than conventional DFT functionals as they include non-local terms and their description of binding due to dispersion is not yet consistently comparable with high level wavefunction based methods, such as coupled-cluster approaches. More pragmatic efforts to improve the treatment of dispersion in DFT have instead focused on empirical corrections, such as the inclusion of a damped London term

in the total energy expression. Following the pioneering application of such approaches by Ahlrichs et al [100] in Hartree-Fock calculations, which lack dispersion by definition, such schemes are implemented by summing the attractions between all distinct pairs of atoms,

$$E_{\text{disp}} = - \sum_{IJ, I > J} f_{\text{damp}}(R_{IJ}) \frac{C_{6, IJ}}{R_{IJ}^6} \tag{3.2.1}$$

where $f_{\text{damp}}(R_{IJ})$ is a damping function which decays to zero for small R_{IJ} and is one at large distances. This damping function is required because electronic structure calculations provide an adequate description of short range attractions and therefore the empirical correction becomes superfluous at small distances. If a damping function is not applied to the dispersion term then the total energy will be distorted, because of the resulting significant artificial strengthening of every covalent bond. As noted above in section 3.1.1 a repulsive potential is not required. The dispersion correction term, E_{disp}, is added to the electronic energy and nucleus-nucleus repulsion energy,

$$E_{\text{tot}} = E_{\text{el}} + E_{\text{disp}} + E_{\text{nuc}}. \tag{3.2.2}$$

The various dispersion correction schemes available differ in the form of the damping function $f_{\text{damp}}(R_{IJ})$ that they employ. Two major forms for this function have been widely used. The first form is that introduced by Mooij et al. [101] and later generalised by Elstner et al. [92],

$$f_{\text{damp}}(R_{IJ}) = \left(1 - e^{-c_{\text{damp}} \left(R_{IJ}/R_{0, IJ} \right)^N} \right)^M \tag{3.2.3}$$

and the second is a Fermi-like function introduced by Wu and Yang [93],

$$f_{\text{damp}}(R_{IJ}) = \frac{1}{1 + e^{-c_{\text{damp}}(R_{IJ}/R_{0, IJ} - 1)}} \tag{3.2.4}$$

where c_{damp} is a damping constant and $R_{0, IJ}$ is determined by the range of the overlap of atoms I and J [92]. Elstner et al. suggested values of 4 and 7 for M and N respectively in equation 3.2.3, which will be referred to here as damping function 1 (DF1). Mooij et al. used M = 2 and N = 3 with the damping function in equation 3.2.3 and this combination will henceforth be referred to as damping function 2 (DF2). Wu and Yang's damping function in equation 3.2.4 will be labelled in what follows as damping function 3 (DF3). It is possible to obtain heteroatomic $R_{0, IJ}$ from the homoatomic values using the following expression [92],

$$R_{0, IJ} = \frac{R_{0, I}^3 + R_{0, J}^3}{R_{0, I}^2 + R_{0, J}^2} \; . \tag{3.2.5}$$

The homoatomic $R_{0, I}$ can be estimated from atomic van der Waals radii [93, 101].

Figure 3.1 shows the three damping functions. While they have similar shapes, the range of r for which each damping function has values in the interval [0.01,0.99] varies. DF2 has a notably more gentle decay to zero, while DF3 decays particularly abruptly.
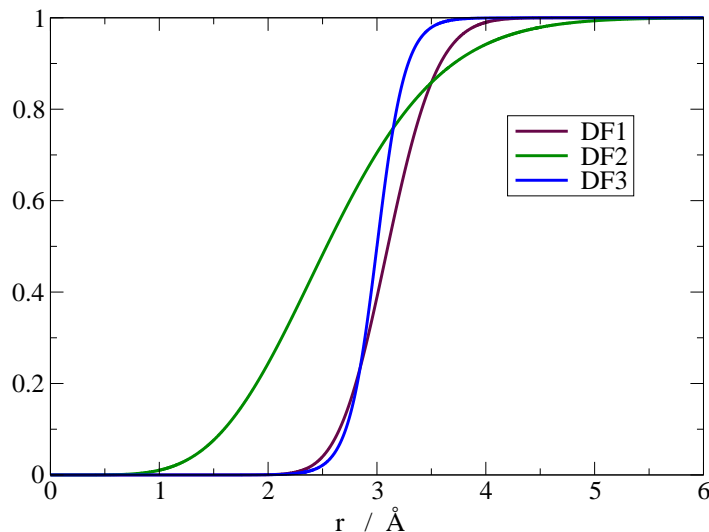
Figure 3.1: The three damping functions using the optimised parameters for carbon with the PBE exchange-correlation functional.

The $C_{6,IJ}$ coefficients can be calculated from the homoatomic $C_{6,I}$ coefficients which can in turn be obtained from experimental work or calculated from the atomic polarisabilities α_I using the expression,

$$C_{6,I} = \frac{3}{4}\sqrt{N_{\text{eff},I}\alpha_I^3} \qquad (3.2.6)$$

where $N_{\text{eff},I}$ is the effective number of electrons [92, 102]. Homoatomic $C_{6,I}$ coefficients can be combined to give heteroatomic $C_{6,IJ}$ coefficients using one of the two equivalent forms [92, 93] of the Slater-Kirkwood combination rule [103],

$$C_{6,IJ} = \frac{2C_{6,I}C_{6,J}\,\alpha_I\alpha_J}{\alpha_I^2 C_{6,J} + \alpha_J^2 C_{6,I}} \qquad (3.2.7)$$

$$= \frac{2\left(C_{6,I}^2 C_{6,J}^2 N_{\text{eff},I} N_{\text{eff},J}\right)^{\frac{1}{3}}}{(N_{\text{eff},I}^2 C_{6,J})^{\frac{1}{3}} + (N_{\text{eff},J}^2 C_{6,I})^{\frac{1}{3}}} \quad . \qquad (3.2.8)$$

3.2.1   Optimisation of parameters

The aim is to use the empirical dispersion correction schemes to improve the description of biomolecular systems in large-scale DFT calculations. Therefore the schemes described in the previous section have been implemented and tested in ONETEP. The approach taken involves also the optimisation of the parameters involved for each exchange-correlation functional in ONETEP. This is necessary because each functional behaves slightly differently; previous work by Grimme involved the optimisation of a single parameter for each functional [104], in the work of Chai a different selection of parameters were optimised for each functional [105]. In order to optimise the parameters, a benchmark set of complexes with dispersion interactions where the binding energies are known to high accuracy were required. Subsets of the JSCH-2005 and S22 sets [106] were chosen for this task. The S22 set is a set of 22 complexes designed to be used as a training set for the inclusion of dispersion corrections and consists of 7 hydrogen bonded complexes, 8 complexes with

predominant dispersion contribution and 7 complexes with significant contributions from both dispersion and hydrogen bonding to the binding. The reference binding energies have been calculated by a combination of MP2 and CCSD(T) methods and extrapolated to the complete basis set limit of CCSD(T). The geometries of the S22 set were obtained by geometry optimisations using MP2 (using a cc-pVTZ basis set and applying a counter-poise correction) for the larger complexes and CCSD(T) (using cc-pVTZ or cc-pVQZ basis sets) for the smaller complexes [106]. The JSCH-2005 set provides similarly high quality binding energies and reference geometries for sets of base pairs and amino acid pairs. The geometries of complexes in the subset of the JSCH-2005 set that are used in this work were obtained by hydrogen-only geometry optimisations of geometries obtained experimentally. A subset of the stacked base pairs and amino acid pairs from the JSCH-2005 set and the non-hydrogen bonded complexes from the S22 set were used as benchmarks. In addition to these, six sulphur-containing complexes from Morgado et al. [107], with binding energies calculated predominantly by MP2, were included so that the parameters for sulphur could also be optimised. The geometries of these complexes were obtained by BLYP-D (with a TZV basis set) optimisation [107]. In total 60 complexes were chosen for the optimisation of 11 parameters. The inclusion of further base pairs from the JSCH-2005 set was deemed undesirable as this could unbalance the chosen training set by giving a bias to base pairs. Also the hydrogen-bonded complexes in the above sets were omitted, as the empirical dispersion corrections are not designed to describe hydrogen bonding, therefore optimising the parameters in a way that causes them to do so (by imitating the CCSD(T) description of the hydrogen-bonds) could compromise their description of dispersion interactions.

Binding energies for the chosen set of complexes were obtained with the following GGA functionals that are available in ONETEP: BLYP [41, 42]; PBE [47]; PW91 [45, 46]; revPBE [108]; RPBE [109]; and XLYP [72], as differences of single-point energies of the bound complex and the two monomers. The geometries were used as provided by the literature and were not modified in these calculations. Subtracting from the reference "exact" binding energies gave the error in the binding energy (and ideal dispersion energy correction) for each complex. The goal of the optimisation was to adjust the parameters in the dispersion formula (3.2.1) to minimise the difference between the value of the dispersion energy and the error in the binding energy for each complex. The parameters optimised were the $C_{6,I}$ coefficients, the $R_{0,I}$ and the c_{damp} coefficients. The optimisation strategy involved the minimisation of the following object function,

$$Err = \sum_{A}^{\mathrm{complexes}} \left[\Delta E_{\mathrm{disp},A} - \left(E_{\mathrm{lit},A}^{\mathrm{bind}} - E_{\mathrm{uncorr},A}^{\mathrm{bind}} \right) \right]^2 \tag{3.2.9}$$

where the index A runs over the the complexes, $\Delta E_{\mathrm{disp},A}$ is the current dispersion energy contribution, $E_{\mathrm{uncorr},A}^{\mathrm{bind}}$ is the pure ONETEP DFT binding energy (without dispersion) and $E_{\mathrm{lit},A}^{\mathrm{bind}}$ is the literature CCSD(T) or MP2 binding energy. The method of steepest descents was then used to optimise the $C_{6,I}$ coefficients, the $R_{0,I}$ and the c_{damp} coefficients in order to minimise Err, the square sum of the errors in the binding energies. For the optimisation the N_{eff} and initial $C_{6,I}$ parameters were taken from Wu and Yang [93] for carbon, hydrogen,

nitrogen and oxygen, and from Halgren [102] for sulphur. The initial c_{damp} parameters used were: 3.0 for DF1, as used by Elstner [92]; 3.54 for DF2, the value Wu and Yang proposed [93], rather than Mooij's value of 7.19 [101]; 23.0 for DF3 following Wu and Yang [93]. $R_{0,I}$ values from Elstner et al. [92] were used. The optimisation was considered converged when either of the following two criteria were satisfied,

- The largest change in any parameter from its initial value exceeds 20%. Since the initial parameters are derived from physical quantities, the optimised parameters should not vary considerably in order to preserve transferability and avoid over-optimisation to the fitting set.

- An iteration satisfied the following inequality,

$$\frac{\text{Maximum percentage change in a parameter in the current step}}{\text{Percentage change in } Err \text{ in the current step}} < 0.5 \quad (3.2.10)$$

  which ensures that the parameters were only varied when this led to a significant reduction in the object function.

The c_{damp} parameter was not restricted by the former criterion as it is completely empirical, for example, Mooij's proposed c_{damp} for DF2 is double Wu and Yang's proposed value. The parameters for sulphur were further optimised by starting from the parameters obtained with the entire set (of 60 complexes) and optimising only the sulphur C_6 coefficient and R_0 with the set of sulphur-containing complexes. In this case the maximum parameter change was limited to 15%, with the latter convergence criterion the same as above. To eliminate possible effects of the basis set, the single point energy calculations with ONETEP were performed with a large kinetic energy cutoff of 1200eV, giving a near-complete psinc basis set. Also large NGWF radii of $8.0a_0$ were used for all elements (except hydrogen which had NGWF radii of $7.0a_0$).

Initial trials were conducted to guide the choice of suitable convergence conditions that incorporated the goals of the parameter optimisation. The trials showed that a small number of iterations (less than 12) gave a large decrease in Err with only small changes in the parameters and then a large number of further iterations (thousands) resulting in significant changes to the parameters but achieving only a small decrease in Err. For example taking the damping function in equation 3.2.4 (with $M = 4$ and $N = 7$) after four iterations Err is reduced by 54% with the largest change in a single parameter being 9%. A further 5063 iterations further reduced Err by only 39% (giving a total reduction by 72%) however, to achieve this one of the parameters was changed by 90% (other parameters changed by 56% and 46%). As noted above such large changes are undesirable. As a result of this trial the minimisation was stopped when any single parameter changed by more than 15% from its initial value or when the percentage change in Err made in the current iteration was less than half the change in the first iteration. When the progress of the minimisation is plotted on a graph, it is seen that the point where these criteria stop the minimisation is a visually intuitive place to stop.

3.2.2 Atomic forces and geometry optimisation

ONETEP is able to compute atomic forces (as analytic derivatives of the total energy) and use these to perform geometry optimisations. The contribution from the dispersion interactions has been included in the forces so that their effect on determining molecular structure can be taken into account during geometry optimisations. As dispersion interactions dominate only in very weakly-bound complexes, a very accurate calculation of all the forces is required. This is possible as the NGWFs are essentially expressed in a plane wave basis and therefore the "egg-box" effect [110] of energy variation with respect to the real-space grid, which is typically observed in real-space techniques, is negligible in ONETEP calculations.

3.2.3 Dispersion forces with periodic boundary conditions

In ONETEP periodic boundary conditions are applied and need to be accounted for in the calculation of the dispersion correction to the energy and atomic forces. This has been implemented by applying a spherical cutoff to the dispersion interaction. A spherical cutoff with a radius of $20a_0$ has been applied since at this distance the dispersion interactions are negligible. To prevent discontinuities in the energy and energy gradient (and hence force) at this cutoff a smoothing function of the form,

$$f_{\text{smooth}}(R_{IJ}) = 1 - e^{-(R_{IJ}-R_{\text{cutoff}})^2}, \tag{3.2.11}$$

where R_{IJ} is the interatomic distance and R_{cutoff} is the cutoff spherical radius. The expression for the dispersion energy is now,

$$E_{\text{disp}} = - \sum_{IJ,I>J} f_{\text{damp}}(R_{IJ}) f_{\text{smooth}}(R_{IJ}) \frac{C_{6,IJ}}{R_{IJ}^6}, \tag{3.2.12}$$

where $f_{\text{damp}}(R_{IJ})$ is a damping function. For each pair of atoms A and B, if atom B or any of its periodic images are within $20a_0$ of atom A, then the dispersion contributions between this pair are included.

3.3 Results and discussion

3.3.1 Validation of energies

Table 3.1: Binding energies (in kcal mol^{-1}) for the complexes used in the fitting of the parameters. The ONETEP results are given with and without dispersion interactions with the optimised parameters for the PBE functional and are compared with the "benchmark" values from the literature.

Complex	PBE	Corrected PBE			Lit [106, 107]
		DF1	DF2	DF3	
2CH$_3$SH (C_1) a3	−1.61	−2.49	−3.36	−2.40	−2.68
2CH$_3$SH (C_1) a5	−1.76	−2.79	−3.66	−2.68	−2.50

Table 3.1: (continued)

| Complex | PBE | Corrected PBE | | | Literature |
		DF1	DF2	DF3	
2CH$_3$SH (C_i) a4	−1.77	−2.68	−3.15	−2.59	−2.00
AA0-3.24 A–As	2.41	−6.08	−6.02	−6.22	−6.25
AA0-3.24 T–Ts	2.81	−4.32	−4.96	−4.19	−3.86
AA20-3.05 AAs2005	3.03	−5.85	−5.86	−5.99	−6.06
AA20-3.05 TTs2005	−0.68	−2.25	−2.47	−2.21	−4.18
A...C S	1.77	−6.48	−6.28	−6.44	−6.70
adenine thymine stack	−1.12	−11.40	−12.23	−11.43	−12.23
AG08-3.19 A–Gs	−0.09	−7.47	−7.32	−7.45	−7.58
AG08-3.19 T–Cs	−0.64	−6.31	−6.45	−6.19	−6.07
A...G S	2.35	−6.30	−6.59	−6.34	−6.50
AT10-3.26 A–Ts	0.88	−6.91	−7.00	−6.84	−6.64
A...T S	1.02	−8.42	−8.37	−8.34	−8.10
benzene ammonia (C_s)	−0.70	−2.38	−2.69	−2.36	−2.35
benzene dimer (C_{2h})	2.02	−3.34	−3.08	−3.41	−2.73
benzene dimer (C_{2v})	−0.05	−2.66	−3.20	−2.59	−2.74
benzene DMS (C_{2v}) a8	−0.15	−3.37	−3.28	−3.39	−3.00
benzene DMS (C_{2v}) a9	−0.66	−1.12	−1.28	−1.10	−1.21
benzene H$_2$S (C_{2v}) a7	−0.74	−2.43	−2.99	−2.30	−2.74
benzene HCN (C_s)	−3.02	−4.77	−5.55	−4.69	−4.46
benzene methane (C_3)	0.08	−1.67	−1.87	−1.67	−1.50
benzene water (C_s)	−1.82	−3.35	−3.83	−3.27	−3.28
CG0-3.19 G–Cs	−1.79	−6.83	−7.09	−6.68	−7.88
C...G S	−2.97	−10.58	−10.53	−10.60	−12.40
ethene dimer (D_{2d})	−0.39	−1.96	−2.30	−1.89	−1.51
ethene ethine (C_{2v})	−1.32	−2.04	−2.20	−2.01	−1.53
F30-F49	−0.16	−3.20	−3.34	−3.21	−3.30
F30-K46	−1.07	−3.66	−3.82	−3.62	−3.10
F30-L33	−0.40	−5.47	−6.44	−5.25	−5.00
F30-Y13	−1.05	−4.87	−5.05	−4.79	−3.90
F30-Y4	0.85	−6.05	−6.02	−6.03	−7.00
F49 C39	0.28	−2.12	−2.94	−2.03	−2.10
F49 C6	0.70	−5.01	−5.33	−4.90	−5.00
F49-K46	−1.34	−4.95	−5.83	−4.85	−4.80
F49-PB V5-C6	−2.26	−7.93	−8.61	−7.81	−8.20
F49-PB Y4-V5	−0.36	−3.29	−3.38	−3.23	−2.80
F49-V5	−0.85	−6.64	−7.91	−6.47	−6.70
F49-Y37	−0.15	−2.41	−2.47	−2.37	−2.50
F49-Y4	1.41	−3.65	−4.34	−3.56	−3.10
GA10-3.15 A–Gs	0.45	−9.43	−9.46	−9.52	−9.14
GA10-3.15 T–Cs	0.88	−5.29	−5.26	−5.24	−4.69
GC0-3.25 G–Cs	−2.04	−10.84	−10.81	−10.92	−10.80
G...C S	−2.78	−10.24	−10.43	−10.29	−8.10
G...C S1	0.25	−7.03	−7.05	−7.02	−7.70
G...C S2	−3.46	−7.86	−7.77	−7.81	−11.60
GG0-3.36 CCs036	−3.94	−4.71	−4.70	−4.70	−3.54
GG0-3.36 GGs036	3.41	−2.01	−1.90	−2.01	−1.62

Table 3.1: (continued)

Complex	PBE	Corrected PBE			Literature
		DF1	DF2	DF3	
GT10-3.15 A–Cs	2.13	−5.54	−5.40	−5.64	−5.44
GT10-3.15 T–Gs	3.44	−5.27	−5.58	−5.16	−4.96
indole benzene stack (C_1)	2.46	−5.34	−5.03	−5.50	−5.22
indole benzene t-shaped (C_1)	−2.14	−5.65	−6.70	−5.59	−5.73
methane dimer (D_{3d})	−0.08	−0.96	−0.91	−0.94	−0.53
phenol dimer (C_1)	−4.33	−7.10	−8.10	−6.97	−7.05
pyrazine dimer (C_s)	0.82	−4.56	−4.48	−4.71	−4.42
TA08-3.16 A–Ts	5.56	−4.44	−5.37	−4.55	−6.07
TG03.19 A–Cs	1.61	−4.40	−4.64	−4.34	−4.96
TG03.19 T–Gs	0.01	−5.22	−5.50	−5.06	−5.67
T...G S	1.82	−6.73	−6.66	−6.64	−6.20
uracil dimer stack (C_2)	−2.66	−9.33	−10.24	−9.12	−10.12
Root mean square error	5.92	0.81	0.93	0.82	0

The 60 complexes used for the fitting of the parameters are presented in Table 3.1. In this table the dispersion-including binding energies as obtained with ONETEP are given using the optimised parameters for the three damping functions (DF1, DF2, and DF3) and the PBE [109] functional. The table also contains the ONETEP binding energies that are obtained when no dispersion contribution is included. The binding energies are compared with the accurate *ab initio* benchmark binding energies for these complexes which are subsets of the JSCH-2005, S22 and Morgado et al. sets of complexes [106, 107]. It can be observed from the table that the inclusion of the dispersion contribution dramatically improves the binding energies, in most cases leading to an agreement with the literature results which is better than 1 kcal mol^{-1}. The optimisation of the parameters has been necessary to obtain this good agreement as, for example, for DF1 with the PBE functional the value of *Err* (defined in equation 3.2.9) was reduced by 78% in the initial parameter optimisation, and in the subsequent sulphur parameter optimisation the *Err* (for the subset of sulphur complexes) was further reduced by 32%. After optimisation DF1 (with PBE) produced binding energies with the lowest root-mean-square (RMS) difference from the literature values of 0.813 kcal mol^{-1}. DF3 had an RMS difference only slightly higher, 0.820 kcal mol^{-1}, however DF2 was noticeably worse with an RMS of 0.926 kcal mol^{-1}, and a very similar trend was observed for the standard deviations. So for the PBE functional DF1 is expected to be the most accurate and consistent. Again DF3 had a similar standard deviation to DF1 (0.823 kcal mol^{-1}) and DF2 had a higher standard deviation (0.897 kcal mol^{-1}).

The retention of the transferability of the parameters was a key concern of this approach. To check if this goal has been achieved, validation calculations, on a set of complexes that were not included in the fitting set, have been performed. These complexes are presented in Table 3.2. They are grouped into four categories: interstrand base pairs, stacked base pairs, hydrogen bonded base pairs and other hydrogen bonded complexes (from the remainder

Table 3.2: Binding energies (in kcal mol^{-1}) for complexes that were not included in the fitting of parameters. Values obtained with just the LDA and PBE exchange-correlation functionals are given as well as values calculated with PBE plus dispersion with DF1, DF2 and DF3.

	Uncorrected					
Complex	LDA	PBE	DF1	DF2	DF3	Lit
Interstrand base pairs						
AA20 3.05 ATis2005	−2.58	−1.23	−2.41	−2.68	−2.36	−2.34
GA10 3.15 A Cis	0.44	1.13	−0.08	−0.11	−0.06	−0.31
GA10 3.15 T Gis	1.02	1.20	0.56	0.56	0.56	0.58
GG0 3.36 CGis036	−3.21	−2.46	−3.94	−3.94	−3.91	−3.68
TG0319 T Cis	−1.17	−0.95	−1.39	−1.39	−1.39	−1.15
Stacked base pairs						
AAst	−6.44	0.27	−8.38	−8.17	−8.52	−8.58
CCst	−8.56	−2.65	−9.61	−9.53	−9.69	−10.02
GGst	−10.39	−2.94	−12.39	−12.30	−12.44	−12.67
UUst	−7.67	−2.31	−8.60	−8.47	−8.61	−7.46
Hydrogen bonded base pairs						
2tU 2tU	−17.40	−10.87	−12.58	−14.16	−12.40	−12.60
6tG C WC pl	−38.08	−28.66	−31.30	−33.40	−31.12	−29.50
A 4tU WC	−18.74	−12.46	−14.80	−16.07	−14.63	−13.20
adenine thymine	−15.65	−15.65	−18.21	−19.94	−18.01	−16.37
G 2tU	−21.85	−14.82	−16.65	−18.42	−16.47	−16.60
G 4tU	−23.82	−16.37	−18.46	−20.23	−18.26	−17.80
uracil dimer hb (C_{2h})	−27.09	−20.11	−21.91	−23.60	−21.75	−20.65
Other hydrogen bonded complexes						
2-pyridoxine	−23.70	−16.86	−19.20	−20.93	−19.02	−16.71
2-aminopyridine (C_1)						
ammonia dimer (C_{2h})	−5.07	−2.93	−3.35	−4.06	−3.31	−3.17
formamide dimer (C_{2h})	−22.48	−16.01	−17.23	−18.80	−17.08	−15.96
formic acid (C_{2h})	−27.28	−19.54	−20.39	−22.40	−20.18	−18.61
water dimer (C_s)	−7.83	−5.20	−5.40	−6.10	−5.31	−5.02
Root mean square error	4.49	3.59	1.07	2.05	0.97	0

of the S22 set). These systems were chosen as they represent a wide range of typical biomolecular environments and also because accurate binding energies are available for these structures in the literature [106]. For the interstrand base pairs and the stacked based pairs the dispersion interaction is the dominant interaction, the results show the same dramatic improvement in the binding energies as in the complexes of Table 3.1. Furthermore, the level of improvement in the binding energies of the stacked and interstrand base pairs is similar even though only stacked base pairs were included in the fitting set, indicating the generality of the empirical dispersion correction. For the hydrogen bonded base pairs and other hydrogen bonded complexes, where the binding is mainly due to hydrogen bonds, the inclusion of the empirical dispersion contribution is not as successful. In a few cases, such as the water dimer for example, the uncorrected ONETEP binding energy is already too large, and the dispersion correction leads to further overbinding. DF2 gave significant overbinding for every hydrogen bonded complex (RMS difference 5.777 kcal mol^{-1}), so this function is less applicable to systems with significant hydrogen bonding, which is the norm for many biological molecules. DF3 performed better than DF1 for all but two of the hydrogen bonded complexes; the RMS differences were 1.225 kcal mol^{-1} and 1.367 kcal mol^{-1} respectively. For the non-hydrogen bonded complexes all the damping functions produced binding energies of similar accuracy, the RMS differences for DF1, DF2 and DF3 were 0.444 kcal mol^{-1}, 0.452 kcal mol^{-1} and 0.430 kcal mol^{-1}, respectively, LDA binding energies have been included for comparison. As expected, these energies are too large for the hydrogen bonded complexes [92]. For the non hydrogen bonded complexes LDA (RMS 1.202 kcal mol^{-1}) produced more accurate binding energies than PBE (RMS 5.366 kcal mol^{-1}); however, they are still inferior to the corrected PBE energies.

3.3.2 Validation on structures

The effect of the dispersion contribution on the atomic forces has been investigated by examining the molecular structures obtained during geometry optimisation. Full (unconstrained) geometry optimisations were performed on four systems: a benzene dimer; a methane dimer; a methane-benzene complex and an indole-benzene complex. All the calculations were performed with DF1 and the PBE functional and a rather tight maximum absolute force convergence threshold of 0.001 E_h/a_0 was used as the forces due to dispersion are obviously very weak.

For the case of the benzene dimer, the optimisation with dispersion contributions resulted in the equilibrium structure shown in Figure 3.2 where the two benzene molecules are in a conformation with their planes parallel, at a separation of 3.7Å. This is in close agreement with the value of 3.9Å that has been obtained with CCSD(T) calculations with a near-complete basis set by Sinnokrot and Sherrill [111]. When dispersion interactions are not included, the pair of benzene molecules experience only the repulsive potential of the PBE functional and no binding is observed but the geometry optimisation is simply completed when their separation is 5.1Å, as at this distance the forces are smaller than the set threshold.

For the methane dimer when dispersion interactions are included ONETEP was able
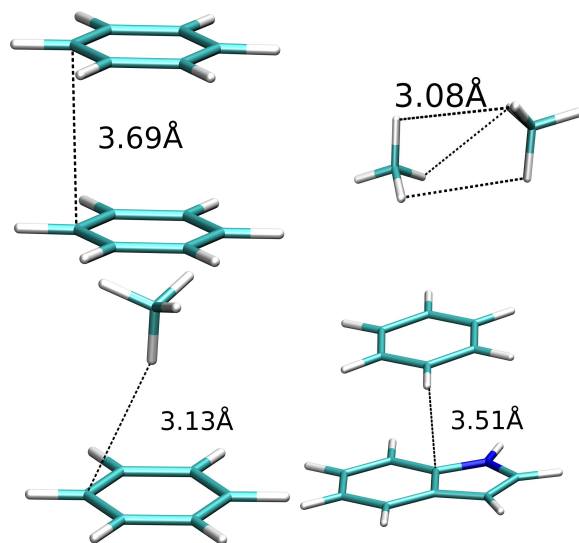
Figure 3.2: Optimised structures with ONETEP of benzene sandwich dimer (top left), methane dimer (top right), methane-benzene (bottom left), indole-benzene (bottom right).

to reproduce the geometry obtained by MP2 calculations using a large Gaussian basis set (cc_pVTZ) [112]. The final structure obtained with ONETEP is shown in Figure 3.2. The structure has the correct symmetry and the distance between hydrogen atoms of the two molecules (3.08Å) is in agreement with the MP2 value (3.07Å). When the empirical dispersion contributions are omitted from the ONETEP geometry optimisation the methane molecules end up much further from each other and their orientation is very different to that obtained using the MP2 approach.

In the case of the benzene-methane complex the benzene-methane distance obtained after optimisation when using our empirical dispersion contribution is 3.15Å (Figure 3.2) which is in close agreement with the value of 2.98Å from an accurate MP2 geometry [106]. Omitting the dispersion contribution results in a distance of 3.62Å, with this greater separation the symmetry of the complex is lost.

For the indole-benzene complex the indole-benzene distance increased from 3.40Å in an MP2-optimised geometry [106] to 3.52Å (Figure 3.2) when optimised with empirical dispersion and 3.92Å when optimised without it. Clearly in all cases the inclusion of the empirical dispersion contribution has significantly improved the geometries obtained.

Chapter 4

Quadratic-scaling Hartree-Fock exchange energy calculation in ONETEP

In this chapter the calculation of Hartree-Fock exchange is discussed. The difficulties in calculating the exchange in ONETEP are considered, including an explanation of why such a calculation is not naturally linear-scaling. A quadratically scaling implementation of the calculation of Hartree-Fock exchange will be described. Finally the performance of the method will be demonstrated by the results of some sample calculations.

4.1 Introduction

In Hartree-Fock theory the exchange term arises from the use of a Slater determinant, an antisymmetrized product of spin orbitals, to express the multi-electron wavefunction [10]. In terms of spatial orbitals the Hartree-Fock exchange energy may be expressed thus,

$$E_{\mathrm{X}}^{\mathrm{HF}} = -\sum_{ij}^{\mathrm{occ}} \int \int \frac{\psi_i^*(\mathbf{r}_1)\psi_i(\mathbf{r}_2)\psi_j^*(\mathbf{r}_2)\psi_j(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \qquad (4.1.1)$$

where $\psi_i(\mathbf{r})$ is a one electron spatial orbital. The exchange integral is only non-zero for electrons of the same spin. The form of the exchange integral is similar to that of the Hartree integral,

$$E_{\mathrm{H}} = 2\sum_{ij}^{\mathrm{occ}} \int \int \frac{\psi_i^*(\mathbf{r}_1)\psi_i(\mathbf{r}_1)\psi_j^*(\mathbf{r}_2)\psi_j(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2. \qquad (4.1.2)$$

In most conventional wavefunction approaches the Hartree and exchange integrals are evaluated in similar ways, sharing optimised code for two electron integral calculation [113]. In plane wave codes Fourier transforms are usually used to calculate the Hartree energy [114, 115].

4.1.1 Hartree-Fock exchange energy in Gaussian codes

The major difficulty in a Hartree-Fock calculation is the evaluation and manipulation of the large numbers of two-electron integrals [10]. The two electron integrals for the Hartree and exchange energies were introduced in section 1.2. The expression for the Hartree and exchange components of the matrix elements of the Fock operator (which was given earlier in equation 1.2.30) is,

$$
G_{\mu\nu} = \sum_b^{\text{occ}} \sum_{\lambda\sigma} C_{\lambda b} C_{\sigma b}^* \left(2 \int \int \phi_\mu^*(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1)r_{12}^{-1}\phi_\sigma^*(\mathbf{r}_2)\phi_\lambda(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 \\
- \int \int \phi_\mu^*(\mathbf{r}_1)\phi_\lambda(\mathbf{r}_1)r_{12}^{-1}\phi_\sigma^*(\mathbf{r}_2)\phi_\nu(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 \right).
\tag{4.1.3}
$$

Due to the use of a basis set, a large number of two electron integral evaluations are required. For example, with a set of 100 basis functions there are 12 753 775 unique two-electron integrals to be evaluated [10]. As discussed in section 1.3.2 evaluating these integrals when using Slater-type orbitals is computationally very expensive as it involves three dimensional numerical integration approaches. Nevertheless improvements to the calculation of two electron integrals with Slater-type orbitals have been developed [116–120]. Analytical formulas (based on recursion) exist to calculate the two-electron integrals in terms of Gaussian orbitals. Therefore, many electronic structure codes use Gaussian basis sets and so calculate the two-electron integrals in terms of Gaussian orbitals [113,121–123]. Due to this a lot of work has been focused on improving the efficiency of algorithms for the calculation of these integrals in terms of Gaussians [124–132].

4.1.2 Hartree Fock exchange energy in plane wave codes

In plane wave codes generally periodic systems are considered, with the supercell approximation being used to study non-periodic systems. For a periodic system the exchange energy is defined,

$$
E_{\text{X}}^{\text{HF}} = -\frac{1}{N_k} \sum_{v\mathbf{k}}^{\text{occ}} \sum_{w\mathbf{k}'}^{\text{occ}} \int \int \frac{\psi_{v\mathbf{k}}^*(\mathbf{r}_1)\psi_{w\mathbf{k}'}^*(\mathbf{r}_2)\psi_{v\mathbf{k}}(\mathbf{r}_2)\psi_{w\mathbf{k}'}(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2
\tag{4.1.4}
$$

where $\psi_{w\mathbf{k}'}(\mathbf{r}_1)$ is a single particle wave function. With plane waves it is convenient to calculate the Coulomb potential in reciprocal space [133],

$$
\int \frac{\psi_{v\mathbf{k}}^*(\mathbf{r}_1)\psi_{w\mathbf{k}'}(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 = \text{FT}\left[\frac{4\pi}{N_k V_c} \frac{Y_{v\mathbf{k},w\mathbf{k}'}(\mathbf{G})}{|\mathbf{G} - \mathbf{k} + \mathbf{k}'|^2} \right]
\tag{4.1.5}
$$

where N_k is the number of $\mathbf{k}$ points, V_c is the volume of the unit cell, and $Y_{v\mathbf{k},w\mathbf{k}'}(\mathbf{G})$ is the Fourier transforms of the co-density, $\psi_{v\mathbf{k}}^*(\mathbf{r})\psi_{w\mathbf{k}'}(\mathbf{r})$,

$$
Y_{v\mathbf{k},w\mathbf{k}'}(\mathbf{G}) = \int_{N_k V_c} e^{-i(\mathbf{G}-\mathbf{k}+\mathbf{k}')\cdot\mathbf{r}} \psi_{v\mathbf{k}}^*(\mathbf{r})\psi_{w\mathbf{k}'}(\mathbf{r})d\mathbf{r}.
\tag{4.1.6}
$$

Using the above the exchange energy in terms of a plane-wave basis is [134],

$$E_{\mathrm{X}}^{\mathrm{HF}} = -\frac{4\pi}{N_k^2 V_{\mathrm{c}}} \sum_{v\mathbf{k}}^{\mathrm{occ}} \sum_{w\mathbf{k}'}^{\mathrm{occ}} \sum_{\mathbf{G}} \frac{Y_{v\mathbf{k},w\mathbf{k}'}(\mathbf{G}) Y_{w\mathbf{k}',v\mathbf{k}}(-\mathbf{G})}{|\mathbf{G} - \mathbf{k} + \mathbf{k}'|^2}. \tag{4.1.7}$$

The Coulomb operator has a singularity at $\mathbf{G} = 0$, when $\mathbf{k} = \mathbf{k}'$ and $v = w$. The singularity is integrable only in the limit of an infinitely fine $\mathbf{k}$ point lattice. Since this is impractical other solutions have to be devised to the problem [135–138]. The usual solution is to truncate the Coulomb operator at a certain distance [114, 115, 134].

Various Hartree-Fock exchange implementations [137–140] have also been developed for the projector-augmented wave method [141], a method that is based on plane waves but which explicitly takes into account the core electrons.

4.2 Hartree-Fock exchange energy in ONETEP

Before considering the calculation of the exchange energy in ONETEP it is helpful to consider the method used to calculate the Hartree energy in ONETEP.

4.2.1 Hartree energy in ONETEP

In ONETEP the Hartree energy is calculated in two stages. Firstly the Hartree potential is calculated in momentum space using the density, which has in turn been calculated from the Non-orthogonal Generalised Wannier Functions (NGWFs) and the density kernel,

$$v_{\mathrm{H}}(\mathbf{r}) = \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \tag{4.2.1}$$

$$= \int \frac{\phi_\alpha(\mathbf{r}') K^{\alpha\beta} \phi_\beta(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \tag{4.2.2}$$

where $\phi_\alpha(\mathbf{r}')$ is an NGWF and $K^{\alpha\beta}$ is the density kernel and implicit summation of Greek indices is used. Note that the NGWFs ϕ_α and ϕ_β, connected by the density kernel elements $K^{\alpha\beta}$, overlap (otherwise the NGWF product is zero). The calculation of the Hartree potential requires a Fourier transform in the entire simulation cell (since it is calculated in momentum space) which is an $\mathcal{O}(n \log n)$ operation, however only one such operation is required for each energy evaluation [2]. The expression in momentum space is [142],

$$\tilde{v}_{\mathrm{H}}(\mathbf{G}) = \frac{4\pi \tilde{n}(\mathbf{G})}{V_{\mathrm{c}}|\mathbf{G}|^2} \tag{4.2.3}$$

where V_{c} is the cell volume. The Hartree energy can then be evaluated alongside the local pseudopotential energy,

$$E_{\mathrm{H}} + E_{\mathrm{loc}} = \int \left(v_{\mathrm{H}}(\mathbf{r}) + v_{\mathrm{loc}}(\mathbf{r}) \right) n(\mathbf{r}) d\mathbf{r}. \tag{4.2.4}$$

4.2.2   Hartree-Fock exchange in terms of NGWFs

Unfortunately the exchange energy is not as straightforward to calculate as the Hartree energy; it is written in terms of NGWFs as,

$$
E_{\mathrm{X}}^{\mathrm{HF}} = -\sum_{\alpha\beta\gamma\delta} K^{\alpha\beta} K^{\gamma\delta} \int\int \frac{\phi_\alpha^*(\mathbf{r}_1)\phi_\beta(\mathbf{r}_2)\phi_\gamma^*(\mathbf{r}_2)\phi_\delta(\mathbf{r}_1)}{|\mathbf{r}_1-\mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \qquad (4.2.5)
$$

where $K^{\alpha\beta}$ is a density kernel element and $\phi_\eta(\mathbf{r})$ is an NGWF. Note that, in contrast to situation in the Hartree potential, the NGWFs ϕ_α and ϕ_β, that are connected by the density kernel elements $K^{\alpha\beta}$, are not required to overlap. Since the indices of the density kernel correspond to different electrons it is not possible to calculate the exchange energy in the same way that the Hartree energy was calculated. This reflects the non-local nature of the exchange operator.

The evaluation of the Hartree-Fock exact exchange energy in ONETEP has been implemented as a two stage process. Initially an exchange matrix with elements,

$$
X_{\alpha\beta} = \sum_{\gamma\delta} \int \phi_\alpha^*(\mathbf{r}_1)\phi_\delta(\mathbf{r}_1) \int \frac{\phi_\beta(\mathbf{r}_2)K^{\gamma\delta}\phi_\gamma^*(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|} d\mathbf{r}_2 d\mathbf{r}_1, \qquad (4.2.6)
$$

where $\phi_\alpha(\mathbf{r}_1)$ is an NGWF and $K^{\gamma\delta}$ is a density kernel element, is calculated. From this matrix the Hartree-Fock exchange energy can be obtained by contraction with the density kernel,

$$
E_{\mathrm{X}}^{\mathrm{HF}} = -\sum_{\alpha\beta} K^{\alpha\beta} X_{\alpha\beta}. \qquad (4.2.7)
$$

Equation 4.2.6 consists of a double integral. In the inner integral the product,

$$
P_{\beta\delta}(\mathbf{r}_2) = \sum_\gamma \phi_\beta(\mathbf{r}_2)K^{\gamma\delta}\phi_\gamma^*(\mathbf{r}_2) \qquad (4.2.8)
$$

is integrated to obtain the potential,

$$
U_{\beta\delta}(\mathbf{r}_1) = \int \sum_\gamma \frac{\phi_\beta(\mathbf{r}_2)K^{\gamma\delta}\phi_\gamma^*(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|} d\mathbf{r}_2 \qquad (4.2.9)
$$

$$
= \int \frac{P_{\beta\delta}(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|} d\mathbf{r}_2. \qquad (4.2.10)
$$

So equation 4.2.6 can be written,

$$
X_{\alpha\beta} = \sum_\delta \int \phi_\alpha^*(\mathbf{r}_1)\phi_\delta(\mathbf{r}_1)U_{\beta\delta}(\mathbf{r}_1) d\mathbf{r}_1 \qquad (4.2.11)
$$

In ONETEP matrices are in general sparse and stored in atomblocked structures as described in section 2.10. It is convenient therefore, to rewrite equation 4.2.5 in terms of

atoms and NGWFs on a particular atom giving,

$$E_X^{\text{HF}} = \int \sum_A \sum_{\substack{D \\ S_{AD} \neq \mathbf{0}}} \sum_a \sum_d \phi_{A,a}^*(\mathbf{r}_1) \phi_{D,d}(\mathbf{r}_1)$$

$$\left[\int \sum_{\substack{B \\ K^{AB} \neq \mathbf{0}}} \sum_{\substack{C \\ S_{BC} \neq \mathbf{0} \\ K^{CD} \neq \mathbf{0}}} \frac{\sum_b K^{A,aB,b} \phi_{B,b}(\mathbf{r}_2) \sum_c \phi_{C,c}^*(\mathbf{r}_2) K^{C,cD,d}}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \right] d\mathbf{r}_1 \qquad (4.2.12)$$

where the indices A,B,C,D refer to atoms (atom blocks) and the indices a,b,c,d refer to the individual NGWFs centred on each atom ($\phi_{X,x}$ is the NGWF x on atom X) and S_{AD} refers to an atomblock of the (sparse) overlap matrix. Similarly the exchange matrix can be written,

$$X_{A,aB,b} = \int \sum_{\substack{D \\ S_{AD} \neq \mathbf{0}}} \sum_d \phi_{A,a}^*(\mathbf{r}_1) \phi_{D,d}(\mathbf{r}_1)$$

$$\left[\int \sum_{\substack{C \\ S_{BC} \neq \mathbf{0} \\ K^{CD} \neq \mathbf{0}}} \frac{\phi_{B,b}(\mathbf{r}_2) \sum_c \phi_{C,c}^*(\mathbf{r}_2) K^{C,cD,d}}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \right] d\mathbf{r}_1. \qquad (4.2.13)$$

The contraction in equation 4.2.7 rewritten in atomblock notation is,

$$E_X^{\text{HF}} = \sum_A \sum_{\substack{B \\ K^{AB} \neq \mathbf{0}}} \sum_a \sum_b K^{A,aB,b} X_{A,aB,b}. \qquad (4.2.14)$$

The computational requirements of equation 4.2.13 can be better understood by looking at figure 4.1, which shows the NGWFs involved in the calculation of the AB atomblock of the exchange matrix.

Inside the square brackets for each atom B, the procedure illustrated in figure 4.2 is required. Firstly the NGWF product is calculated,

$$P_{B,bD,d}(\mathbf{r}_2) = \sum_{\substack{C \\ S_{BC} \neq \mathbf{0} \\ K^{CD} \neq \mathbf{0}}} \phi_{B,b}(\mathbf{r}_2) \sum_c \phi_{C,c}^*(\mathbf{r}_2) K^{C,cD,d}. \qquad (4.2.15)$$

Since the interpolation and multiplication can be carried out inside an FFT box the computational cost of these operations is, in principle, independent of the system size. The
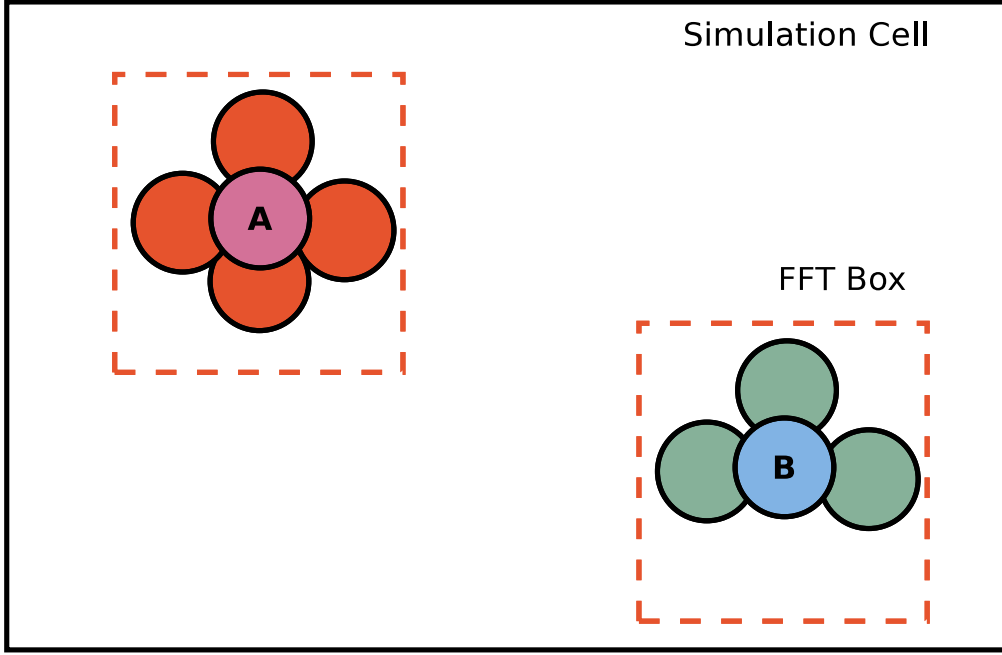
Figure 4.1: The NGWFs involved in the calculation of the AB atomblock of the exchange matrix in equation 4.2.13. The NGWFs on and surrounding atom A are the $\phi_{D,d}s$, the NGWFs on and surrounding atom B are the $\phi_{C,c}s$. It is clear that this calculation is not local to a single FFT box.

inner integral in equation 4.2.13,

$$U_{B,bD,d}(\mathbf{r}_1) = \int \sum_{\substack{C \\ S_{BC} \neq \mathbf{0} \\ K^{CD} \neq \mathbf{0}}} \frac{\phi_{B,b}(\mathbf{r}_2) \sum_c \phi^*_{C,c}(\mathbf{r}_2) K^{C,cD,d}}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \tag{4.2.16}$$

$$= \int \frac{P_{B,bD,d}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \tag{4.2.17}$$

can be evaluated in momentum space with linear-scaling cost. However since a pair of Fourier transforms of the whole simulation cell are required this is an expensive operation, especially in terms of communications bandwidth. The problem is that this operation has to be performed for each A, a and D, d an $\mathcal{O}(N)$ number of times. Therefore if the energy were to be calculated using this scheme then the computational cost would be $\mathcal{O}(N^2)$; there would also be a large prefactor for the reasons outlined above. Since it is clear that this approach would not be suited to large systems due to the high computational cost and quadratic-scaling. The implementation was carried out only as an accuracy benchmark with which to validate the linear-scaling approaches that are developed later in chapter 5. In this implementation the FFT box size was set to be the same as the cell size for simplicity. As FFT boxes are not distributed across cores the many Fourier transforms required can be performed independently on each core. The full implementation in ONETEP is shown in figure 4.3.

In order to achieve linear-scaling it is clear that an alternative approach is required,
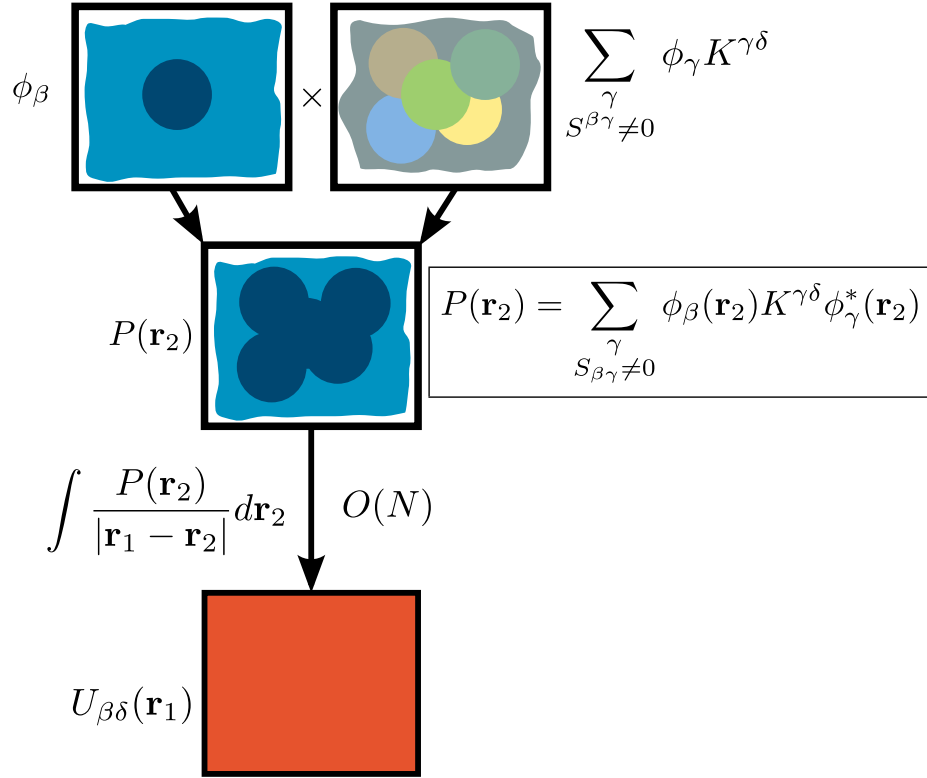
Figure 4.2: Calculation of the product of NGWFs in the inner integral (equation 4.2.9) in the evaluation of the exchange energy in equation 4.2.6. Interpolation is required to prevent aliasing during the multiplication. A similar procedure is carried out by ONETEP in the calculation of the density [2]. The product $P(\mathbf{r}_2)$ is then used to generate the potential $U_{\beta\delta}(\mathbf{r}_1)$ by Fourier transform.

such approaches are developed in chapter 5. The above method could be used in a linear-scaling approach to evaluate the integrals only in case where all the NGWFs involved lie within a normal-sized FFT box. The cost of the Fourier transforms required in this case will be independent of system size. However a different method to deal with the long range exchange interactions would be required.

4.2.3 Exchange matrix sparsity

The Hartree-Fock exchange matrix requires a different sparsity pattern to that used by the standard GGA (or LDA) Hamiltonian. This need arises because Hartree-Fock exchange is a non-local operator and the calculation of the exchange matrix and exchange energy uses all the non-zero elements of the density kernel matrix. This can be clearly seen from equation 4.2.14 where there is a sum over all the atom pairs A and B with non-zero AB density kernel atomblocks in the calculation of the exchange energy. As discussed in section 2.10.1 the non-zero elements in the density kernel are determined by a distance-based cutoff. Since the only restriction on A and B in the calculation of the exchange energy is that corresponding atomblock in the density kernel is non-zero, the exchange matrix will have the same sparsity pattern as the density kernel. Any non-zero atomblock in the exchange
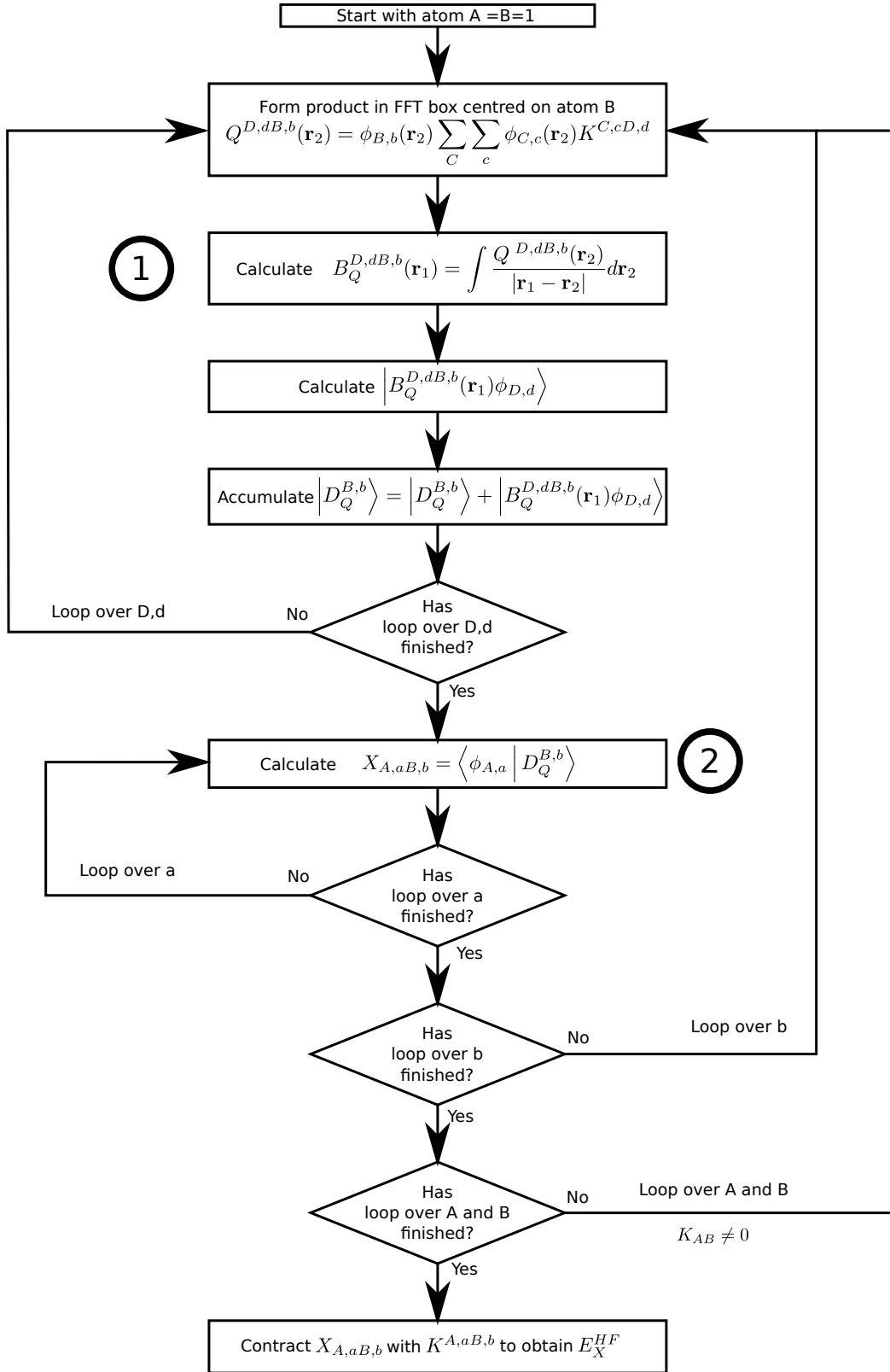
Figure 4.3: Flowchart of the implementation of the calculation of Hartree-Fock exchange energy in ONETEP. In the implementation described in this chapter the step labelled 1 is done using FFTs. The step labelled two is modified for the calculation of the NGWF gradient.

matrix that does not have a corresponding non-zero atomblock in the density kernel would not contribute to the energy (see equation 4.2.7) so is not relevant to the calculation. Also if a density kernel atomblock AB is non-zero then the exchange matrix atomblock AB will also be non-zero, since there will be a non-zero contribution from the case where $D = A$ and $C = B$ (so $K^{CD} = K^{AB}$; S^{AA} and S^{BB} are clearly non-zero). For the standard Hamiltonian, only those density kernel elements that are non-zero in a sparsity pattern determined by the overlap between NGWFs and those pairs of NGWFs that both overlap with a common projector (in the pseudopotential), are relevant (see section 2.10.1). So the Hamiltonian matrix that includes Hartree-Fock exchange will be in general less sparse than the standard Hamiltonian: its sparsity will be the same as the sparsity of the density kernel and dictated by the kernel cutoff.

4.2.4 Cutoff Coulomb operator

In momentum space the exchange potential can be evaluated in the same way as the Hartree potential [142],

$$\tilde{U}_{\beta\delta}(\mathbf{G}) = \frac{4\pi \tilde{P}_{\beta\delta}(\mathbf{G})}{V|\mathbf{G}|^2}, \tag{4.2.18}$$

where V is the volume of the FFT box. However, equation 4.2.18 has a singularity at $\mathbf{G} = \mathbf{0}$, due to a zero denominator.

To avoid this singularity at $\mathbf{G} = \mathbf{0}$, whilst still using Fourier transforms, there are two options introducing a uniform background charge or using a (real space) truncated Coulomb operator [114, 115]. With a uniform background charge the $\mathbf{G} = \mathbf{0}$ term is set to zero. ONETEP uses this method to deal with the singularity for the calculation of the Hartree and Ewald energies. In these cases the background charges that are introduced are opposite in sign and in neutral systems these cancel each other out. It is not desirable to introduce a large unphysical background charge into the system, so the truncated Coulomb operator approach was chosen. In real space the form of this operator is,

$$f(\mathbf{r}) = \begin{cases} \frac{1}{|\mathbf{r}|} & |\mathbf{r}| < R \\ 0 & |\mathbf{r}| \geq R \end{cases} \tag{4.2.19}$$

and in reciprocal space,

$$V(\mathbf{G}) = \frac{4\pi}{V} \frac{1 - \cos(|\mathbf{G}|R)}{|\mathbf{G}|^2}. \tag{4.2.20}$$

So the expression for the exchange potential in reciprocal space becomes,

$$\tilde{U}_{\beta\delta}(\mathbf{G}) = \frac{4\pi \tilde{P}_{\beta\delta}(\mathbf{G})[1 - \cos(|\mathbf{G}|R)]}{V|\mathbf{G}|^2}, \tag{4.2.21}$$

where V is the volume of the FFT box, and $\tilde{P}_{\beta\delta}(\mathbf{G})$ is the Fourier transform of the product

defined in equation 4.2.8. At $\mathbf{G} = \mathbf{0}$ equation 4.2.21 has the value,

$$\tilde{U}_{\beta\delta}(\mathbf{0}) = \frac{2\pi \tilde{P}_{\beta\delta}(\mathbf{0}) R^2}{V}. \tag{4.2.22}$$

(Obtained by taking the limit and applying l'Hôpital's rule.) In ONETEP the cutoff used is equivalent to half the width of the FFT box. This minimum image convention (see section 1.4.4.1) also ensures that only nearest periodic images interact with the exchange term. Recent functionals use an attenuated Coulomb operator, which is a cutoff Coulomb operator with a gradual decay to zero (rather than a sharp cutoff) [143]. It would not be too hard to extend the present method to include this behaviour.

4.3   Optimisation of the density kernel

To calculate the Hartree-Fock exchange energy self-consistently, the gradient of the exchange energy with respect to variation of the density kernel elements and the gradient of the exchange energy with respect to variation of the expansion coefficients of the NGWFs are required. The gradient of the exchange energy with respect to the auxiliary density kernel (rather than simply the gradient with respect to the density kernel) is required as the density kernel optimisation is based on the Li-Nunes-Vanderbilt (LNV) method described in section 2.6.3. The derivative of the exchange energy with respect to the density kernel can be derived,

$$\frac{\partial E_X^{\mathrm{HF}}}{\partial L^{\eta\pi}} = \frac{\partial E_X^{\mathrm{HF}}}{\partial K^{\varepsilon\theta}} \frac{\partial K^{\varepsilon\theta}}{\partial L^{\eta\pi}} \tag{4.3.1}$$

$$\frac{\partial E_X^{\mathrm{HF}}}{\partial K^{\varepsilon\theta}} = 2 \sum_{\gamma\delta} \int \phi_\theta^*(\mathbf{r}_1) \phi_\delta(\mathbf{r}_1) \int \frac{\phi_\varepsilon(\mathbf{r}_2) K^{\gamma\delta} \phi_\gamma^*(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 d\mathbf{r}_1 \tag{4.3.2}$$

$$= 2 X_{\varepsilon\theta}. \tag{4.3.3}$$

An expression for $\frac{\partial K^{\varepsilon\theta}}{\partial L^{\eta\pi}}$ was given in equation 2.8.11. Currently the derivative of the total energy with respect to the auxiliary density kernel is calculated in ONETEP by,

$$\frac{\partial E_{\mathrm{tot}}}{\partial L^{\eta\pi}} = 2 H_{\varepsilon\theta} \frac{\partial K^{\varepsilon\theta}}{\partial L^{\eta\pi}}. \tag{4.3.4}$$

Therefore the exchange matrix (in terms of the density kernel) simply needs to be added to the rest of Hamiltonian matrix in the LNV optimisation.

A notable difficulty in early tests of the exchange calculation was the lack of convergence in geometry optimisations as there were (energy) convergence problems after several geometry steps. The cause of this problem is that the Hartree-Fock exchange matrix gradually lost its Hermiticity during the calculation. Hermiticity is important because the eigenvalues of a Hermitian matrix are stable (small changes in the coefficients give small changes in the eigenvalues) whereas the eigenvalues of large non-Hermitian matrices are in general not [5]. This loss of Hermiticity was due to numerical errors (because of the finite precision available) in the calculation of the Hartree-Fock exchange matrix. These errors were then

transferred to the density kernel (as the Hartree-Fock matrix is a component of the density kernel gradient) which also became slightly non-Hermitian. In the next iteration the now not quite Hermitian, density kernel is used in the calculation of the exchange matrix; the result of this is that the exchange matrix becomes slightly less Hermitian. The level of Hermiticity remains acceptable during single-point energy calculations; however, during a geometry optimisation the increased number of density kernel optimisations eventually results in significant loss of Hermiticity in the density kernel and other matrices. When Hermiticity is lost in this way the calculation fails to converge. The solution that was implemented is that the Hartree-Fock matrix is made symmetric (by averaging with the transpose); this has resolved the problems with geometry optimisations.

4.4 Optimisation of NGWFs

Since the Hartree-Fock exchange energy also depends on the NGWFs for self-consistency the expression for the derivative of the exchange energy with respect to the NGWFs is also required,

$$\frac{\delta E_{\mathrm{X}}^{\mathrm{HF}}}{\delta \phi_{\varepsilon}^*(\mathbf{r})} = K^{\alpha\beta} \frac{\delta X_{\alpha\beta}}{\delta \phi_{\varepsilon}^*(\mathbf{r})} + 2 \frac{\delta K^{\alpha\beta}}{\delta \phi_{\varepsilon}^*(\mathbf{r})} X_{\alpha\beta} \qquad (4.4.1)$$

taking each term separately,

$$K^{\alpha\beta} \frac{\delta X_{\alpha\beta}}{\delta \phi_{\varepsilon}^*(\mathbf{r})} = 2K^{\alpha\beta} \frac{\delta X_{\alpha\beta}}{\delta \phi_{\alpha}^*(\mathbf{r})} \frac{\partial \phi_{\alpha}^*(\mathbf{r})}{\partial \phi_{\varepsilon}^*(\mathbf{r})} \qquad (4.4.2)$$

$$= 2K^{\alpha\beta} \phi_{\delta}(\mathbf{r}) \int \frac{\phi_{\beta}(\mathbf{r}_2) K^{\gamma\delta} \phi_{\gamma}^*(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \delta_{\varepsilon}^{\alpha} \qquad (4.4.3)$$

$$= 2K^{\varepsilon\beta} \phi_{\delta}(\mathbf{r}) \int \frac{\phi_{\beta}(\mathbf{r}_2) K^{\gamma\delta} \phi_{\gamma}^*(\mathbf{r}_2)}{|\mathbf{r} - \mathbf{r}_2|} d\mathbf{r}_2. \qquad (4.4.4)$$

The calculation of this term has been implemented in the code by modifying the last operation in the code (this operation is labelled 2 in figure 4.3) which calculates the exchange matrix to instead return the NGWF gradient. The second term is,

$$\frac{\delta K^{\alpha\beta}}{\delta \phi_{\varepsilon}^*(\mathbf{r})} X_{\alpha\beta} = (3L^{\alpha\varepsilon} \phi_{\pi}(\mathbf{r}) L^{\pi\beta} - 2L^{\alpha\varepsilon} \phi_{\pi}(\mathbf{r}) L^{\pi\nu} S_{\nu\mu} L_{\mu\beta}$$

$$+ 2L^{\alpha\nu} S_{\nu\mu} L^{\mu\varepsilon} \phi_{\pi}(\mathbf{r}) L^{\pi\beta}) X_{\alpha\beta} \qquad (4.4.5)$$

$$= \phi_{\pi}(\mathbf{r})(3\mathbf{LXL} - 2\mathbf{LSLXL} - 2\mathbf{LXLSL})^{\pi\epsilon}. \qquad (4.4.6)$$

This term is included by adding the exchange matrix to the Hamiltonian matrix, in a similar manner to the density kernel optimisation, therefore it is included in the **Q** defined

in equation 2.9.9 [25]. So equation 4.4.1 becomes,

$$\frac{\delta E_{\mathrm{X}}^{\mathrm{HF}}}{\delta \phi_{\varepsilon}^{*}(\mathbf{r})} = 2K^{\varepsilon\beta}\phi_{\delta}(\mathbf{r}) \int \frac{\phi_{\beta}(\mathbf{r}_2)K^{\gamma\delta}\phi_{\gamma}^{*}(\mathbf{r}_2)}{|\mathbf{r} - \mathbf{r}_2|} d\mathbf{r}_2$$
$$+ 2\phi_{\pi}(\mathbf{r})(3\mathbf{LXL} - 2\mathbf{LSLXL} - 2\mathbf{LXLSL})^{\pi\epsilon}. \quad\quad (4.4.7)$$

The NGWF gradient due to the Hartree-Fock exchange contribution needs to be pre-conditioned in reciprocal space, to match the preconditioning of the rest of total energy's NGWF gradient, see section 2.9.1.1 [65].

4.5   Coarse grid approach

Many of the steps in the calculation of the Hartree-Fock exchange are done in the fine grid to avoid aliasing problems. However there is a large computational cost for working in the fine grid. A fine grid FFT box contains eight times as many points as a coarse grid FFT box and so operations on the former will take at least eight times more effort and memory than computation on the latter. Also the use of the fine grid requires expensive interpolation and filtering operations which imposes a significant additional burden. Since the cost of using the fine grid is so high the possibility of using the coarse grid to calculate the Hartree-Fock exchange energy was investigated. As long as the errors introduced by applying this approximation are small the reduction in the cost makes it worthwhile. Others have used similar reasons to allow aliasing in the calculation of the Hartree potential [114].

As discussed in section 2.2.2 if the coarse grid psincs can be expanded in terms of plane waves with $\mathbf{G}$ vectors up to a certain $|\mathbf{G}_{\mathrm{max}}|$ then the fine grid psincs may be expanded in plane waves with $\mathbf{G}$ vectors up to $2|\mathbf{G}_{\mathrm{max}}|$. By using the coarse grid psincs instead of the fine grid psincs information expressed by the plane waves with the higher cutoff frequencies will be lost. This loss of information is known as aliasing and is a problem when products of psincs are considered as shown in equation 2.2.28. However in reciprocal space both the Coulomb and cutoff Coulomb operators contain a $|\mathbf{G}|^{-2}$ factor. This factor will damp the higher frequency components, so their contribution to the potential and energy will therefore be less significant. It can be expected, therefore that the omission of these high frequency contributions will not significantly change the convergence pattern or energies obtained in the calculation. The effect of this approximation was tested by running tests using both the coarse and fine grid methods and comparing the outcome. These tests were run as Hartree-Fock calculations where the full exchange contribution is included. When a hybrid functional is used the exchange contribution is scaled, so any inaccuracy in the exchange contribution will be similarly reduced. The timings of the two methods will also be compared to assess the value of the trade-off.

4.6   Parallel algorithms

In order to perform tests on large systems in reasonable amounts of time, a parallel imple-mentation is required. In the current implementation each process calculates elements of
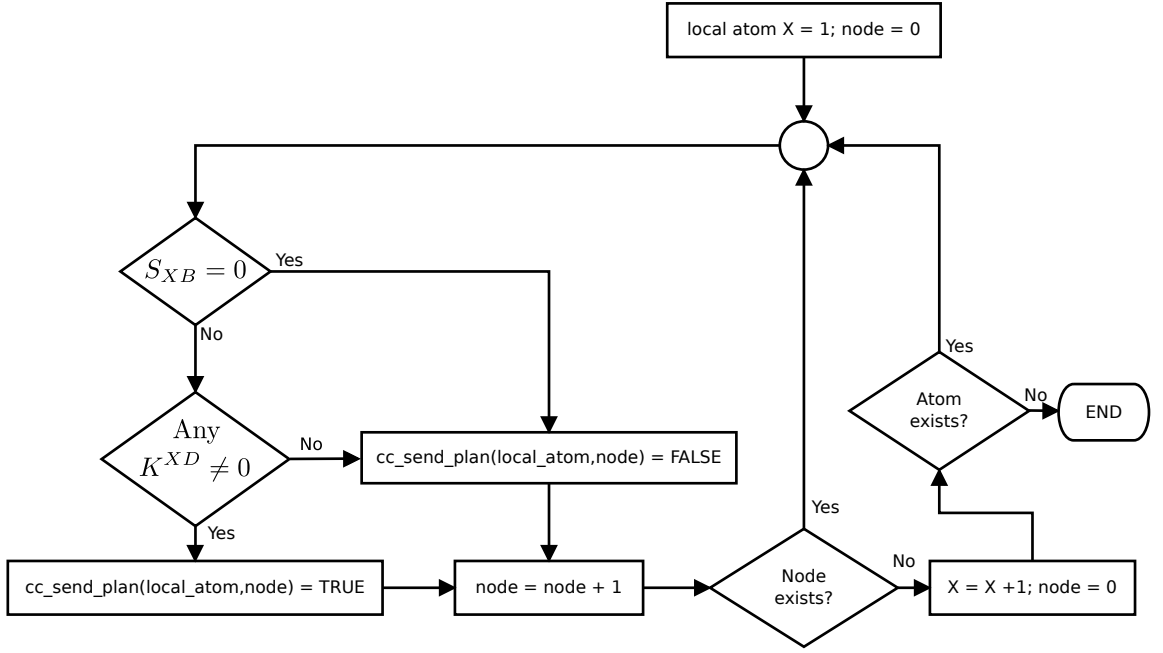
Figure 4.4: Flow chart illustrating the construction of the communications plan for the C, c NGWFs required for the calculation of the Hartree-Fock exchange matrix in equation 4.2.13.

the atom blocks in the exchange matrix that are local to that process. These atom blocks are in the block columns of the exchange matrix corresponding to atoms local to that process. The column index corresponds to the global index of atom A. The NGWFs on the B, C and D atoms (see equation 4.2.12 are sent to the core that holds atom A. Potentially the $K_{C,cD,d}$ matrix elements will also need to be sent to the core holding A. The operations are performed in batches of D, d NGWFs, the default batch size is 10. Using batches improves the FFT efficiency as a pair of real FFT boxes can be Fourier transformed or interpolated in a single operation. At the start of each batch each process communicates the contents of the batch with the other cores (the index of atom B and the indices of the D, d NGWFs). From this a plan for the necessary send and receive operations is constructed. The construction of this plan for the C, c NGWFs is illustrated in figure 4.4; a similar procedure is used to to construct a plan for the D, d NGWFs. This enables all the necessary NGWFs to be sent as soon as possible which minimises waiting times for the receiving processes. This scheme can be quite inefficient if there is not a similar number of NGWFs local to each process; this is likely to be the case where the the number of processes is high relative to the number of atoms. Table 4.1 shows the performance of the parallel implementation on an Si_{16} unit cell by comparing the time taken by a single NGWF iteration on varying numbers of cores. The time taken for the calculation to complete decreased as the number of cores used increased. However the total CPU time used also tended to increase, but this does not have such a consistent pattern; for example, using four cores was more efficient than using three cores. There was a large drop in efficiency (a large increase in the total CPU time used) between the serial calculations and the parallel calculations; however the efficiency did not drop too dramatically further as the number of cores used increased. Since the

Table 4.1: Timings for Si_{16} ONETEP Hartree-Fock calculations running on a various numbers of cores. The timings were the minimum obtained from six sets of single NGWF iteration calculations.

	Time/s per core		Total time/s		Speed up	
Cores	Fine	Coarse	Fine	Coarse	Fine	Coarse
1	9403	2257	9403	2257	1.00	1.00
2	8681	1806	17362	3611	1.08	1.25
3	6274	1251	18822	3752	1.50	1.80
4	4652	876	18609	3506	2.02	2.58
6	3668	687	22011	4122	2.56	3.28
8	2684	516	21473	4132	3.50	4.37
16	1403	268	22440	4290	6.70	8.42
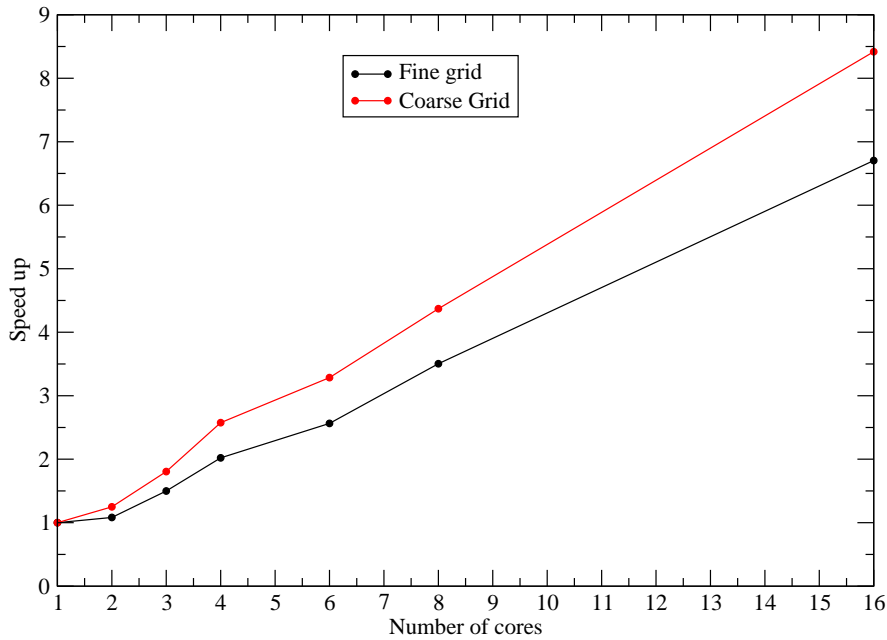
Figure 4.5: Speed up with number of cores for a ONETEP Hartree-Fock calculation on an Si_{16} unit cell. The speed up is defined in equation 4.6.1.

system is a dense solid there are many overlapping NGWFs and so a lot of communication is required, this impedes the parallel performance. Some time (over 10% for the coarse grid on two cores) is also wasted in the parallel calculations as some cores wait for the others to catch up. Figure 4.5 illustrates the parallel performance of the implementation on the Si_{16} unit cell. The speed up is defined as,

$$S_n = \frac{t_{1\text{core}}}{t_{\text{ncores}}}. \tag{4.6.1}$$

4.7 Results and discussion

Various tests of the implementation have been undertaken to ensure it performed correctly. Initially finite differences were used to check that the calculated NGWF and density kernel gradients were correct. For a water molecule, an Si_{16} unit cell, a water and formaldehyde
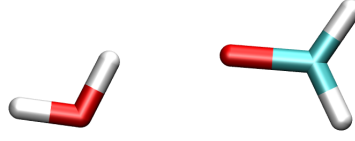
Figure 4.6: Water and formaldehyde complex

Table 4.2: Comparison of energy components in E_h obtained with ONETEP and CASTEP for a water and formaldehyde complex (see figure 4.6) using the PBE exchange-correlation functional. The same kinetic energy cutoff and pseudopotentials were used with both codes.

	ONETEP	CASTEP
Kinetic	26.53	27.10
Pseudopotential (local)	−143.58	−144.44
Pseudopotential (non-local)	5.29	5.19
Hartree	57.51	57.77
Exchange-correlation	−9.74	−9.79
Ewald	24.23	24.23
Total	−39.75	−39.93

Table 4.3: Comparison of energy components in E_h obtained from Hartree-Fock calculations with ONETEP and CASTEP for a water and formaldehyde complex (see figure 4.6). ONETEP calculations were performed on both the fine and coarse grids. The same kinetic energy cutoff and pseudopotentials were used with both codes.

	ONETEP		CASTEP
Grid	Fine	Coarse	Fine
Kinetic	26.09	26.12	26.64
Pseudopotential (local)	−143.27	−143.31	−144.16
Pseudopotential (non-local)	5.26	5.25	5.18
Hartree	57.71	57.72	58.00
Hartree-Fock exchange	−8.98	−8.99	−9.03
Ewald	24.23	24.23	24.23
Total	−38.95	−38.96	−39.13

As a further validation of the implementation, a comparison was made with the results from CASTEP [144], a plane wave code. CASTEP is able to give energies comparable to those obtained with ONETEP when the same pseudopotentials and kinetic energy cutoff are used with both codes. Table 4.2 shows the agreement between CASTEP and ONETEP energies for the PBE functional for a water and formaldehyde complex. The energies agree reasonably well, the differences are due to the differences in the definition of the plane wave basis set in the two codes (as a sphere of **g**-vectors in CASTEP, in ONETEP as a cube) [145].

Similar agreement would be expected from Hartree-Fock calculations; the results of these calculations are shown in table 4.3. The results show that the errors introduced by the use of the coarse grid did not have a significant effect on the energy in this case. The discrepancies between the ONETEP and CASTEP were very similar to those seen when using PBE (table 4.2).

A geometry optimisation has been performed on ethene, the results of this are presented in table 4.4. NWChem [146], an Gaussian based code was used as to provide a comparison. Two basis sets were used for the NWChem calculations aug-cc-pVDZ and the much larger aug-cc-pV5Z set. It can be seen from the table that the parameters obtained with the two basis sets did not differ dramatically. Total energies are not comparable between ONETEP and NWChem because NWChem performs all-electron calculations while ONETEP uses pseudopotentials. However the structures resulting from a geometry optimisation should be comparable. There is reasonable agreement between the NWChem and ONETEP parameters. There was no significant difference between the structures obtained using the fine and coarse grids with ONETEP.

Table 4.4: Comparison of parameters of ethene structures obtained from Hartree-Fock geometry optimisations using NWChem and ONETEP. The same parameters were obtained from both the fine and coarse grid ONETEP calculations.

Parameter	NWChem		ONETEP
	aug-cc-pVDZ	aug-cc-pV5Z	
C-H bond length / Å	1.08	1.07	1.07
C-C bond length / Å	1.32	1.31	1.29
H-C-C bond angle	121.60	121.65	122.29

Further testing and validation has been done to verify the suitability of the coarse grid approach. Table 4.5 compares the energies of several structures for the coarse and fine grids. A mixture of systems have been chosen with some small molecules, a bulk solid and and organometallic system, these are pictured in figure 4.7. The computational demands of the fine grid approach prevent the study of larger systems. The energies obtained agree well between calculations done on the coarse and fine grids. This suggests that the error introduced by this approximation is not significant and so the coarse grid can be used for further calculations with the great benefit of the significantly reduced computational effort.

Table 4.6 compares the performance of the coarse and fine grid Hartree-Fock implementations. In all cases the coarse grid calculation are an order of magnitude faster per iteration than when the fine grid is used. However the computational cost of coarse grid Hartree-Fock calculations far exceeds that of calculations using the PBE functional. The vastly increased computational demands restrict the practicality of performing Hartree-Fock calculations with ONETEP to small systems. The PBE calculations were run with the FFT box the size of the simulation cell, to match the Hartree-Fock calculations. The much larger simulation cell in the PBE water calculation caused it to take far longer than the Si_{16} calculation despite fewer atoms being involved. The size of the FFT box also has some

Table 4.5: Comparison of energies in E_h from Hartree-Fock exchange calculations using ONETEP for a selection of structures (see figure 4.7) using the coarse and fine grid. PBE energies are provided for comparison.

	Total energy / E_h		
	Coarse	Fine	PBE
ammonia	−11.471	−11.473	−11.740
benzene	−36.880	−36.883	−37.987
ethene	−13.400	−13.405	−13.794
$Fe(NH_2)_4$	−60.635	−60.635	−62.777
hydrogen cyanide	−15.843	−15.841	−16.233
Si_{16}	−60.732	−60.732	−63.108
water	−16.921	−16.921	−17.214
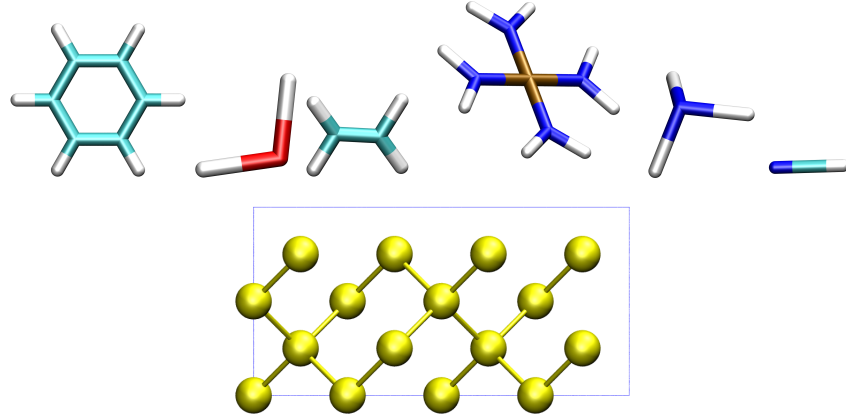
Figure 4.7: The systems that were tested. Top row (left to right): benzene, water, ethene, $Fe(NH_2)_4$, ammonia, hydrogen cyanide. Bottom row: Si_{16} unit cell.

effect on the Hartree-Fock calculations but this is outweighed by the increased number of operations that comes with the greater number of NGWFs involved.

Finally it is necessary to ensure the internal consistency of the implementation. This means that kernel and NGWF gradients have to be consistent with the energy. In addition to the finite difference testing above, tests have been performed to show that convergence is correct.

Figure 4.8 compares the NGWF convergence behaviour of the fine and coarse grid implementations with that of PBE. In the initial iterations the convergence behaviour of both methods matches that of PBE. There is, however, a difference in the last few iterations between the three. For ethene the fine grid Hartree-Fock calculation converges to the RMS gradient threshold (2×10^{-6}) fastest. In the ethene calculation the coarse grid method struggles slightly to converge but there is not a significant difference, and convergence is achieved only two iterations after the fine grid method. For water the convergence of the fine and coarse grid methods are very similar and both converge slightly quicker than PBE (but all three pass the threshold on iteration 12).
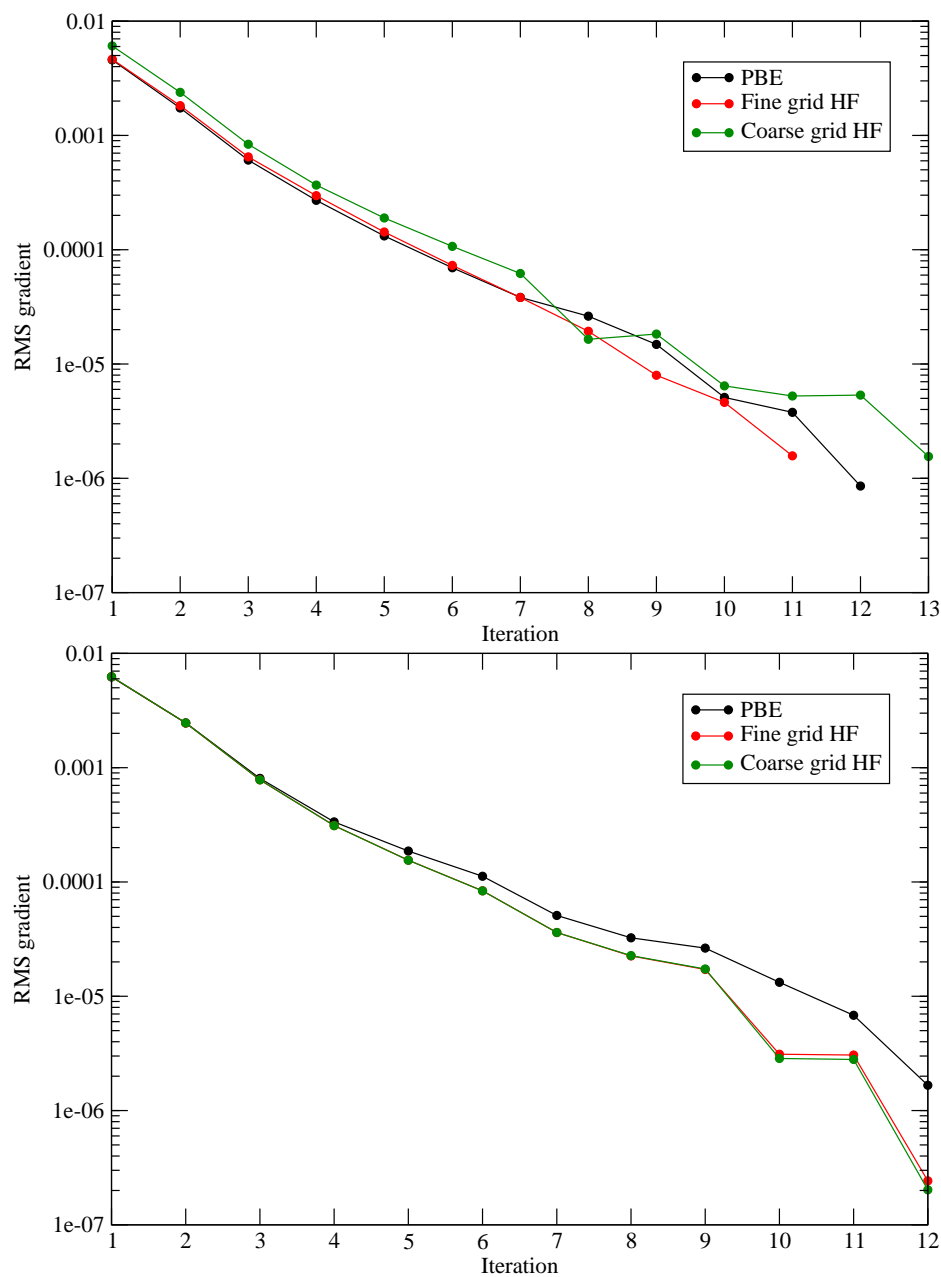
Figure 4.8: NGWF convergence of ONETEP Hartree-Fock single-point energy calculations for ethene (top) and water (bottom) with the coarse and fine grid methods. Convergence for PBE is included for comparison.
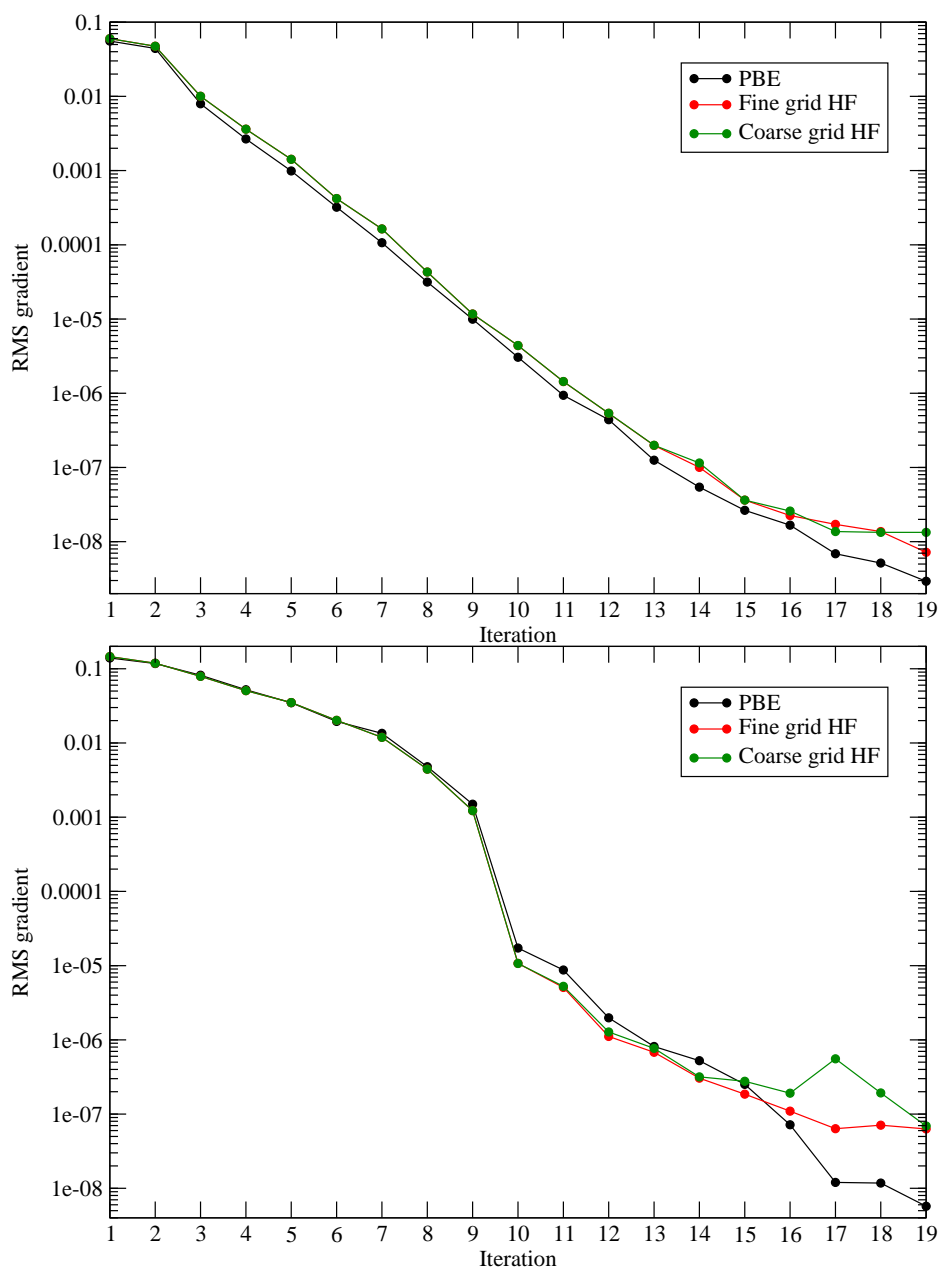
Figure 4.9: LNV convergence of ONETEP Hartree-Fock single-point energy calculations for ethene (top) and water (bottom) with the coarse and fine grid methods. Convergence for PBE is included for comparison.

Table 4.6: Comparison of times per iteration for Hartree-Fock exchange calculations using ONETEP for a selection of structures (see figure 4.7) using the coarse and fine grid. PBE timings are provided for comparison.

	Time per iteration / s			
	Coarse	Fine	PBE	Cores
ammonia	102	1040	50	2
benzene	2573	33478	59	6
ethene	145	1740	19	2
$Fe(NH_2)_4$	1292	14950	69	4
H_2O-H_2CO	165	2287	19	6
hydrogen cyanide	119	1194	46	2
Si_{16}	876	4652	4	4
water	60	618	44	2

Figure 4.9 compares the LNV convergence behaviour of the fine and coarse grid implementations with that of PBE. The coarse and fine grid methods both show convergence behaviour similar to that of PBE. For ethene there are minor differences even in the last few iterations. There are more noticeable differences in the final iterations of the water calculations. The coarse grid method appeared to have convergence problems in iteration 17 but recovers to match the fine grid method by iteration 19.

4.8   Summary

An implementation of the calculation of Hartree-Fock exchange in ONETEP has been presented. It is rather slow and scales quadratically, which restricts the systems that may be studied. The coarse grid can be used instead of the fine grid to provide a large speed up with minimal lose of accuracy. However a faster method with more favourable scaling is required for practical calculations. This will be developed in the next chapter. The method described in this chapter will provide a suitable benchmark for such a method, because it is based on the most straightforward and exact, albeit slow, application of the theory.

Chapter 5

Linear-scaling Hartree-Fock exchange energy calculation in ONETEP

In this chapter a linear-scaling approach to the calculation of Hartree-Fock exchange will be presented. First we will look at previous work towards the linear-scaling calculation of Hartree-Fock exchange. We will then consider two approaches to calculating the exchange in ONETEP. The focus will be on the second approach and the many implementation choices and challenges this presents.

5.1   Previous work on implementations of linear-scaling Hartree-Fock exchange

There are several existing approaches that allow the Hartree-Fock exchange energy to be calculated in linear-scaling manner. The expression for the Hartree and exchange components of the matrix elements of the Fock operator (which was given earlier in equation 1.2.30) is,

$$
\begin{aligned}
G_{\mu\nu} = \sum_{b}^{N/2} \sum_{\lambda\sigma} C_{\lambda b} C_{\sigma b}^{*} \Bigg(&2 \int \int \phi_{\mu}^{*}(\mathbf{r}_1)\phi_{\nu}(\mathbf{r}_1) r_{12}^{-1} \phi_{\sigma}^{*}(\mathbf{r}_2)\phi_{\lambda}(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \\
&- \int \int \phi_{\mu}^{*}(\mathbf{r}_1)\phi_{\lambda}(\mathbf{r}_1) r_{12}^{-1} \phi_{\sigma}^{*}(\mathbf{r}_2)\phi_{\nu}(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \Bigg).
\end{aligned}
\tag{5.1.1}
$$

The key to many of these methods is a prescreening step where those integrals that are expected to make a small contribution to the exchange matrix (and hence energy) are excluded [147]. This will generally result in a non-variational method as the effect on the energy of excluding these contributions is uncertain [148]. A basic screening criterion is a threshold for the magnitude of the bare two-electron repulsion integrals. An upper bound

107

for the value of the integral can be obtained using Schwartz's inequality [147, 149],

$$(\mu\lambda|\sigma\nu) = \int\int \phi_\mu^*(\mathbf{r}_1)\phi_\lambda(\mathbf{r}_1)r_{12}^{-1}\phi_\sigma^*(\mathbf{r}_2)\phi_\nu(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 \tag{5.1.2}$$

$$\leq \sqrt{(\mu\lambda|\mu\lambda)}\sqrt{(\sigma\nu|\sigma\nu)}. \tag{5.1.3}$$

The two integrals in this upper bound may be precomputed, so the upper bound can be calculated very cheaply. However since the expansion coefficients are ignored this is not a reliable guess of the magnitude of the contribution. More sophisticated methods use thresholding with density weighted two-electron repulsion integrals, this accounts for the size of the expansion coefficients so provides a more reliable estimate of the significance of the contribution [147, 150]. Any existing optimised integral routines present in the code are still used for the actual computation of the integral. Often this prescreening step is not linear-scaling but since this step is fast this was not a problem for the system sizes practical at the time. Also while after prescreening the number of integrals evaluated should (asymptotically) scale linearly with system size, the onset of linear-scaling will be delayed if the prescreening step does not avoid the calculation of a large number of integrals [150]. Later algorithms were developed that performed the prescreening with linear-scaling effort [151, 152].

ONX_A (originally ONX) [150, 153] is a method using three thresholding criteria to eliminate integrals. The normal prescreening based on an estimate of the exchange matrix element is supplemented by two further criteria which eliminate some of the integrals that contribute to each matrix element. The additional criteria filter out insignificant distributions, a distribution (or co-density) is defined $\rho_{ac}(\mathbf{r}) = \phi_a(\mathbf{r})\phi_c(\mathbf{r})$, and some of the insignificant integrals that result from combinations of these distributions. Ordered lists are used to minimise the overhead of the comparisons by allowing early exit from the inner loops. A later development was ONX [151]. This method uses a different method to pick significant combinations of distributions, ordered lists are also used at the thresholding stage to ensure linear-scaling. ONX can be used with periodic boundary conditions by introducing the minimum image convention (see section 1.4.4.1) [30].

LinK [152] uses two criteria to select integrals and also constructs ordered lists during thresholding. Additionally before the thresholding step there is preselection of distributions using overlap criteria and preordering based on integral estimates. The LinK algorithm, unlike ONX, makes use of the permutational symmetry of the integrals which reduces the prefactor. In the chain of spheres exchange (COSX) algorithm [154] the inner exchange integral is done analytically and the outer integral is done numerically. This, as the authors note, destroys the Hermitian character of the two electron integrals. Thresholding is based on the overlap of basis functions and the contribution of pairs of functions to the density (the matrix elements of a density matrix).

An alternative to prescreening with a threshold is to use a radial cutoff. Some methods use a screened Coulomb potential [155, 156] and ignore the long range Hartree-Fock exchange interactions. Other distance based cutoff methods use lower accuracy method to evaluate the integrals between distant centres, an example of this is the near-field exchange

method [157] which uses a fast multipole method to evaluate the integrals between distant centres.

5.1.1 Resolution of the identity

A further type of approach uses Resolution of the Identity (RI) techniques to calculate exchange [148, 158]. It involves the expansion of products of basis functions (i.e. co-densities) in terms of auxiliary basis functions,

$$\phi_\mu^*(\mathbf{r})\phi_\nu(\mathbf{r}) = \sum_i c_{\mu\nu i} f_i(\mathbf{r}) \tag{5.1.4}$$

where $f_i(\mathbf{r})$ is an auxiliary basis function and $c_{\mu\nu i}$ is an expansion coefficient. Methods of fitting basis function products to an auxiliary basis set are discussed later in section 5.3. The exchange integral in equation 5.1.1 can then be expressed in terms of the auxiliary basis set,

$$\int\int \phi_\mu^*(\mathbf{r}_1)\phi_\lambda(\mathbf{r}_1)r_{12}^{-1}\phi_\sigma^*(\mathbf{r}_2)\phi_\nu(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 = \sum_{ij} c_{\mu\lambda i}c_{\sigma\nu j}\int\int f_i(\mathbf{r}_1)r_{12}^{-1}f_j(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2.$$

$$\tag{5.1.5}$$

A similar approach can be used to calculate the Hartree energy by expanding the density in an auxiliary basis set [159–161].

However, the use of an auxiliary basis set on its own is only sufficient to reduce the integral evaluation to cubic-scaling for exchange [162]. To achieve linear-scaling locality needs to be enforced in the fitting [163]. Without this the cost of fitting each product will grow linearly with system size (as the number of fitting functions would increase). The Atomic RI method [163] considers only auxiliary basis functions in the region of the centres of the basis functions in the co-density. However in that method there are still other quadratic-scaling operations so the algorithm is not linear-scaling. Another approach [164] uses localised orbitals which results in localised co-densities. A localised co-density fitting is applied which causes the number of exchange integral evaluations to scale only asymptomatically linearly with system size (other steps in the calculation, such as Fock matrix diagonalisation, are still cubic-scaling).

5.2 Numerical pointwise approach

As shown in chapter 4, approaches to the calculation of Hartree-Fock exchange which rely on Fourier transforms result in quadratic-scaling. For ONETEP a linear-scaling algorithm is required. The expensive Fourier transform was used to evaluate the integral (from equation

4.2.16),

$$U_{B,bD,d}(\mathbf{r}_1) = \int \sum_{\substack{C \\ S_{BC} \neq \mathbf{0} \\ K^{CD} \neq \mathbf{0}}} \frac{\phi_{B,b}(\mathbf{r}_2) \sum_c \phi_{C,c}^*(\mathbf{r}_2) K^{C,cD,d}}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \qquad (5.2.1)$$

$$= \int \frac{P_{B,bD,d}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2. \qquad (5.2.2)$$

A linear-scaling method will have to provide an alternative method of evaluating this integral. One approach is to evaluate the integral numerically as a pointwise sum over the grid points that are used to express $P_{B,bD,d}(\mathbf{r}_2)$,

$$U_{\beta\delta}(\mathbf{r}_1) = \frac{\omega_{\text{cell}}}{N} \sum_{\mathbf{r}_2} \frac{P_{\beta\delta}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}, \qquad (5.2.3)$$

where ω_{cell} is the cell volume and N is the number of grid points. This effectively treats every point in the co-density as a point charge when the Coulomb operator is applied. The cost of this calculation is independent of system size. To include periodicity the minimum image convention (described in section 1.4.4.1) is used.

This approach does present a difficulty when $\mathbf{r}_1 = \mathbf{r}_2$, since the distance between the two points is zero, so is the denominator, which introduces a singularity. This is solved by calculating the potential due to the psinc function localised on the grid point. The psinc function is approximated by a truncated sinc function. This is equivalent to an $l = 0$ spherical wave (these are introduced later in section 5.4.3) with the grid point spacing, d (chosen to be the smallest spacing if this is different for each dimension), as the cutoff radius,

$$f(\mathbf{r}) = \begin{cases} \frac{\sin(qr)}{qr} & r < d \\ 0 & r \geq d \end{cases}. \qquad (5.2.4)$$

Note that the spherical wave has been multiplied by a factor of $\sqrt{4\pi}$ to remove the unwanted effect of the spherical harmonic. The first Bessel zero is taken so $q = \frac{\pi}{d}$. An equation for the potential (within the cutoff radius) due to an $l = 0$ spherical wave is given in equation 5.6.16 as,

$$U_{\text{SW}}^{l=0}(\mathbf{r}_1) = 4\pi Z_{00}(\theta_1, \phi_1) \left(\frac{\sin(qr_1)}{q^3 r_1} - \frac{\cos(qd)}{q^2} \right). \qquad (5.2.5)$$

Evaluating this at zero and multiplying by the $\sqrt{4\pi}$ factor (which cancels out $Z_{00}(\theta_1, \phi_1)$) gives,

$$U_{\text{SW}}'^{l=0}(\mathbf{0}) = \frac{8d^2}{\pi}. \qquad (5.2.6)$$

So this is used instead of the Coulomb operator for these points.

Using the fine grid the approach of pointwise numerical integration is not feasible, due to the immense computational demands, for all but the very simplest systems. At each grid

point, $\mathbf{r}_1$, in the FFT box for which the Hartree-Fock exchange potential is desired a sum over all the points, $\mathbf{r}_2$, in another FFT box, multiplied by the distance between the two points as in equation 5.2.3 So to calculate the potential in an single FFT box the number of operations required is the square of the number of points in an FFT box.

Using the coarse grid the computational demands become much lower. The use of the coarse grid, as opposed to the fine grid, saves a factor of 64 because there are 2^3 fewer points in each FFT box. Also because there is no need for interpolation, FFT boxes are no longer required and tightboxes can be used instead. A tightbox is, in general, 3 times smaller in each dimension than an FFT box so saving a factor of $(3^3)^2 = 729$. In total there is a 46656 fold reduction in the computational effort required which brings this method into the realms of practicality. Some simple optimisation is possible by grouping points separated by the same distance to reduce the number of times that reciprocal separation distance factor has to be recalculated and to make use of array operations. Dr Jacek Dziedzic, a post-doctoral fellow in the research group, further optimised the code in order to assess the feasibility of using this method on large systems. In spite of these improvements the methods detailed below were faster for larger systems.

5.2.1 Convergence

As shown in figure 5.1 the NGWF convergence of this method is very similar to that seen with the coarse grid FFT method. The LNV convergence shown in figure 5.2 is also similar for both methods, with the numerical pointwise approach slightly better towards the end. This is as expected because the two methods should be equivalent (apart from the contributions from points where $\mathbf{r}_1 = \mathbf{r}_2$). This verifies that the implementation is correct.

5.3 Expanding the NGWF product in a basis set

Since the numerical integration approach was not a satisfactory solution to the problem and alternative was sought. An obvious alternative is to apply a Resolution of the Identity (RI) method to ONETEP. This involves expanding the NGWF product, defined in equation 4.2.15, in terms of an auxiliary basis set,

$$P_{B,bD,d}(\mathbf{r}_2) \approx \sum_i^{N_{\text{fit}}} f_i(\mathbf{r}_2)c_i \tag{5.3.1}$$

that allows the potential integral,

$$U_{B,bD,d}(\mathbf{r}_1) \approx \sum_i^{N_{\text{fit}}} c_i \int \frac{f_i(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \tag{5.3.2}$$

to be obtained analytically. The expansion coefficients for the NGWF product in the set of basis functions are found by solving the set of linear equations expressed,

$$\mathbf{Vc} = \mathbf{b} \tag{5.3.3}$$
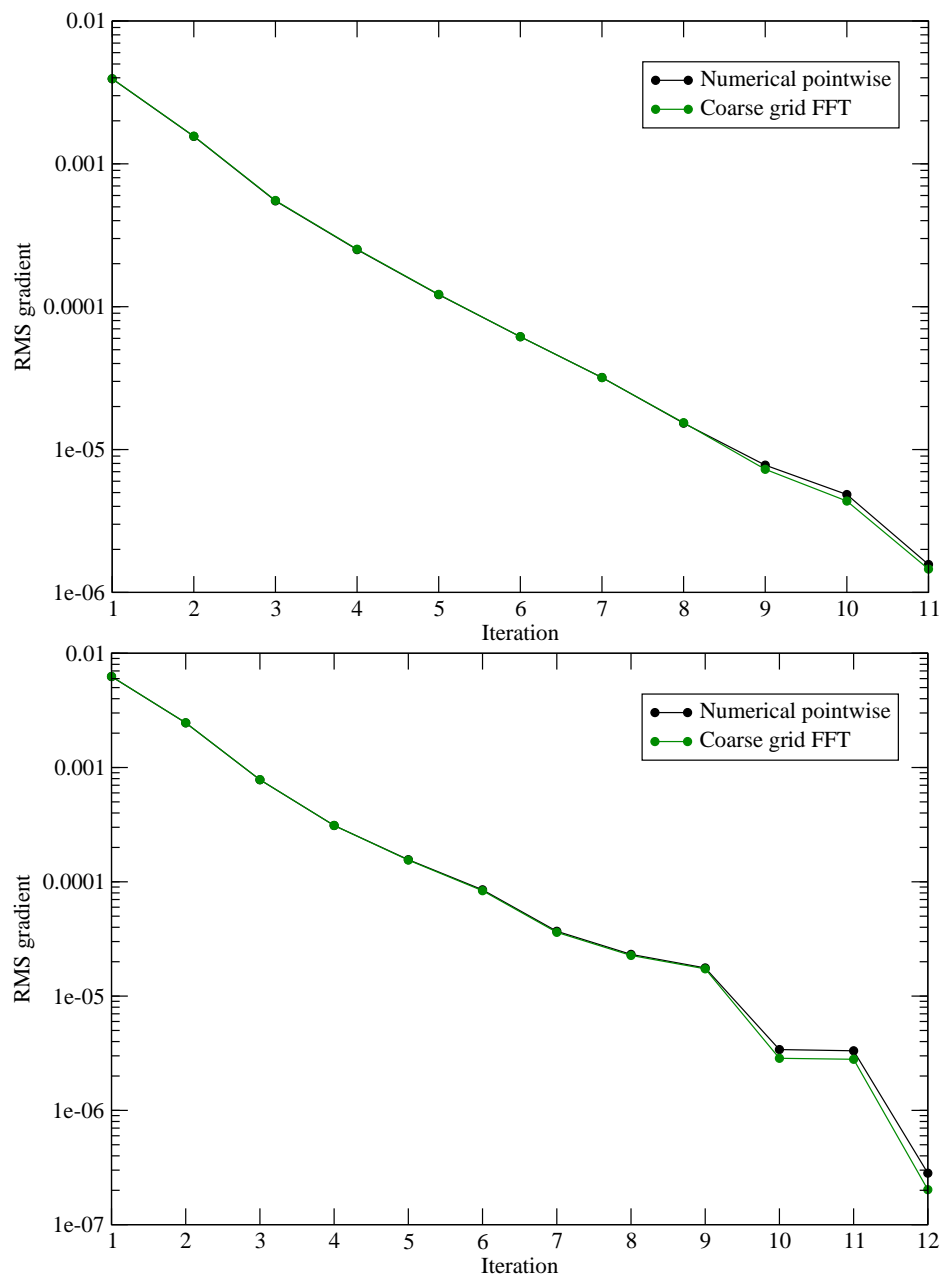
Figure 5.1: NGWF convergence of ONETEP Hartree-Fock single-point energy calculations for ethene (top) and water (bottom) using the numerical pointwise and the coarse grid FFT methods.
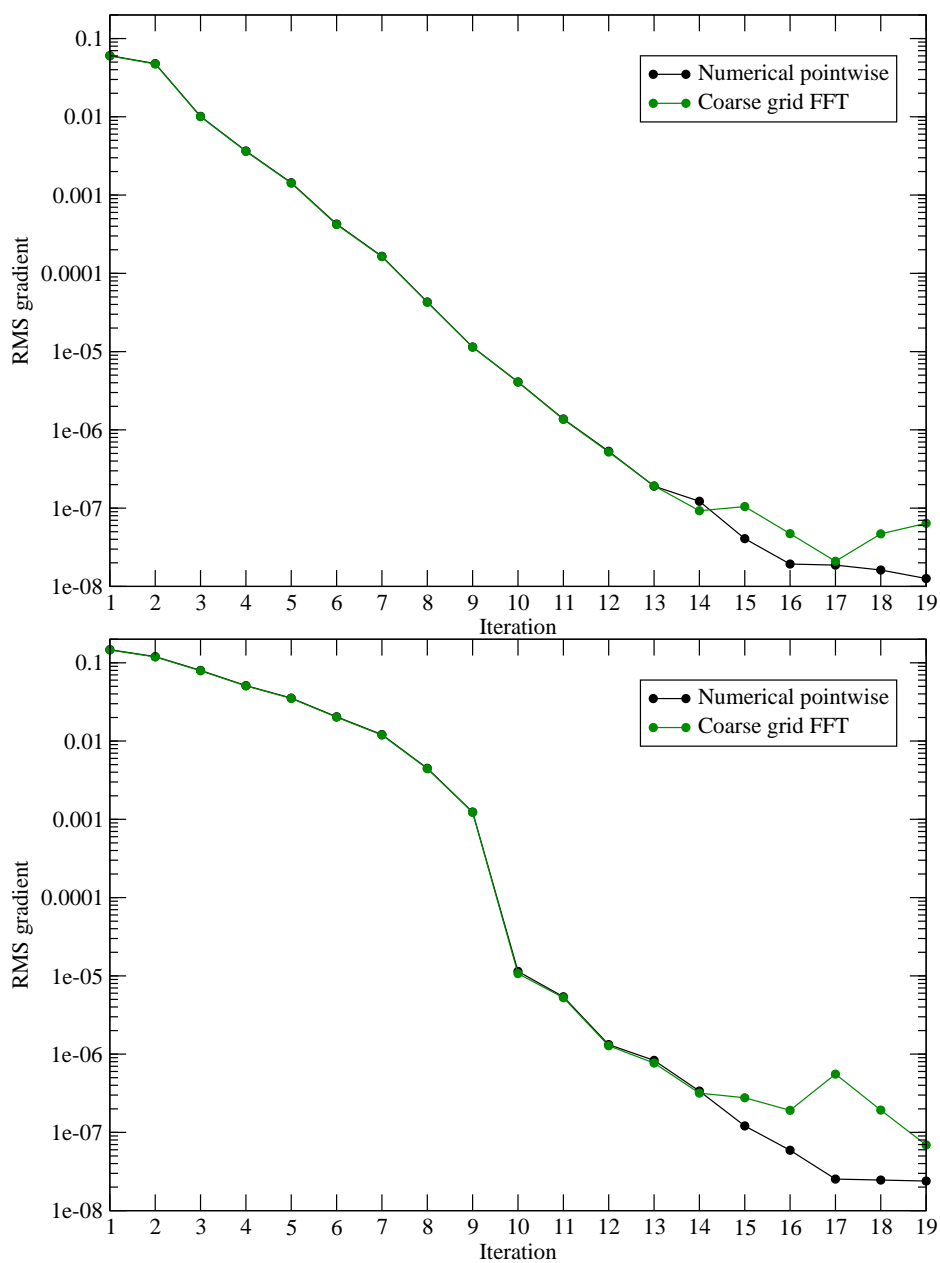
Figure 5.2: LNV convergence of ONETEP Hartree-Fock single-point energy calculations for ethene (top) and water (bottom) using the numerical pointwise and the coarse grid FFT methods.

or equivalently,

$$\mathbf{c} = \mathbf{V}^{-1}\mathbf{b} \tag{5.3.4}$$

where $\mathbf{V}$ is the basis function-basis function metric matrix, $\mathbf{c}$ is a vector of expansion coefficients and $\mathbf{b}$ is the NGWF product-basis function inner product vector.

5.3.1   Overlap metric

In the overlap metric space the inner product (for real functions) is defined as,

$$\langle f_i \,|\, f_j \rangle = \int f_i(\mathbf{r}) f_j(\mathbf{r}) d\mathbf{r} \tag{5.3.5}$$

Since only real functions are used no distinction will be made between a function and its complex conjugate. This inner product satisfies the usual properties of an inner product [165],

$$\langle f_i \,|\, f_j \rangle = \int f_i(\mathbf{r}) f_j(\mathbf{r}) d\mathbf{r}$$
$$= \int f_j(\mathbf{r}) f_i(\mathbf{r}) d\mathbf{r}$$
$$= \langle f_j \,|\, f_i \rangle \tag{5.3.6}$$
$$\langle f_i \,|\, \lambda f_j + \mu f_k \rangle = \int f_i(\mathbf{r}) \left[\lambda f_j(\mathbf{r}) + \mu f_k(\mathbf{r}) \right] d\mathbf{r}$$
$$= \lambda \int f_i(\mathbf{r}) f_j(\mathbf{r}) d\mathbf{r} + \mu \int f_i(\mathbf{r}) f_k(\mathbf{r}) d\mathbf{r} \tag{5.3.7}$$
$$\langle f_i \,|\, f_i \rangle = \int f_i(\mathbf{r}) f_i(\mathbf{r}) d\mathbf{r}$$
$$\geq 0 \tag{5.3.8}$$
$$\langle f_i \,|\, f_i \rangle = 0 \Leftrightarrow f_i(\mathbf{r}) = 0 \qquad \forall \, \mathbf{r} \in \mathbb{R}^3. \tag{5.3.9}$$

The NGWF product-basis function overlap vector ($\mathbf{b}$) has elements,

$$b_i = \int f_i(\mathbf{r}) P_{\beta\delta}(\mathbf{r}) d\mathbf{r}. \tag{5.3.10}$$

The matrix elements of the overlap metric matrix ($\mathbf{V}$) are,

$$V_{ij} = \langle f_i \,|\, f_j \rangle \tag{5.3.11}$$
$$= V_{ji}. \tag{5.3.12}$$

Only basis functions centred on atom B have been used for the fitting. Using basis functions centred on the atoms overlapping with atom B (C atoms) does not make sense as the NGWF product is localised within the localisation sphere of B. On the coarse grid this localisation is exact; on the fine grid interpolation of the NGWFs means that the product is not strictly localised. The use of an increased number of basis functions also increases

the computational effort required to perform the calculation.

5.3.2 Electrostatic metric

In the electrostatic metric space the inner product (for real functions) is defined as,

$$(f_i|f_j) = \int \int \frac{f_i(\mathbf{r}_1)f_j(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}d\mathbf{r}_1 d\mathbf{r}_2. \tag{5.3.13}$$

This inner product satisfies the usual properties of an inner product [165],

$$
\begin{aligned}
(f_i|f_j) &= \int \int \frac{f_i(\mathbf{r}_1)f_j(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}d\mathbf{r}_1 d\mathbf{r}_2 \\
&= \int \int \frac{f_j(\mathbf{r}_1)f_i(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}d\mathbf{r}_1 d\mathbf{r}_2 \\
&= (f_j|f_i) \tag{5.3.14}
\end{aligned}
$$

$$
\begin{aligned}
(f_i|\lambda f_j + \mu f_k) &= \int \int \frac{f_i(\mathbf{r}_1)\left[\lambda f_j(\mathbf{r}_2) + \mu f_k(\mathbf{r}_2)\right]}{|\mathbf{r}_1 - \mathbf{r}_2|}d\mathbf{r}_1 d\mathbf{r}_2 \qquad \lambda, \mu \in \mathbb{R} \\
&= \lambda \int \int \frac{f_i(\mathbf{r}_1)f_j(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}d\mathbf{r}_1 d\mathbf{r}_2 + \mu \int \int \frac{f_i(\mathbf{r}_1)f_k(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}d\mathbf{r}_1 d\mathbf{r}_2 \\
&= \lambda(f_i|f_j) + \mu(f_i|f_k) \tag{5.3.15}
\end{aligned}
$$

$$
\begin{aligned}
(f_i|f_i) &= \int \int \frac{f_i(\mathbf{r}_1)f_i(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}d\mathbf{r}_1 d\mathbf{r}_2 \\
&\geq 0 \tag{5.3.16}
\end{aligned}
$$

$$(f_i|f_i) = 0 \Leftrightarrow f_i(\mathbf{r}) = 0 \qquad \forall \mathbf{r} \in \mathbb{R}^3. \tag{5.3.17}$$

The elements of the NGWF product-basis function inner product vector ($\mathbf{b}$) are,

$$b_i = (f_i|P_{\beta\delta}). \tag{5.3.18}$$

The elements of the electrostatic metric matrix ($\mathbf{V}$) are,

$$
\begin{aligned}
V_{ij} &= (f_i|f_j) \\
&= V_{ji}. \tag{5.3.19}
\end{aligned}
$$

Fitting functions that don't overlap with the NGWF product need to be included, so, in contrast to the overlap metric, basis functions centred on both atoms A and B are used for the fitting. This means that the fitting basis set used to expand a product centred on atom A is the same as that used to expand a product on atom B (i.e. the entire set of fitting functions on centres A and B). The use of a consistent fitting basis set is a useful property with the electrostatic metric and will be exploited below. With the electrostatic metric expanding one of the NGWF products in the exchange matrix expression is equivalent to expanding both products [161]. This can be shown by defining the expanded NGWF

products (indicated by a double tilde to prevent confusion with Fourier transforms) as,

$$P_{\beta\delta}(\mathbf{r}_2) = \phi_\beta(\mathbf{r}_2) K^{\gamma\delta} \phi_\gamma^*(\mathbf{r}_2) \tag{5.3.20}$$

$$\tilde{\tilde{P}}_{\beta\delta}(\mathbf{r}_2) = \sum_{ij} f_i(\mathbf{r}_2) [V^{-1}]^{ij} (f_j|P_{\beta\delta}) \tag{5.3.21}$$

$$Q_{\alpha\delta}(\mathbf{r}_1) = \phi_\alpha^*(\mathbf{r}_1) \phi_\delta(\mathbf{r}_1) \tag{5.3.22}$$

$$\tilde{\tilde{Q}}_{\alpha\delta}(\mathbf{r}_1) = \sum_{kl} f_k(\mathbf{r}_1) [V^{-1}]^{kl} (Q_{\alpha\delta}|f_l) \tag{5.3.23}$$

and then using both expanded products in the expression for the exchange matrix,

$$X_{\alpha\beta} = \sum_{\substack{\delta \\ S_{\alpha\delta}\neq 0}} (Q_{\alpha\delta}|P_{\beta\delta}) \tag{5.3.24}$$

$$\approx \sum_{\substack{\delta \\ S_{\alpha\delta}\neq 0}} \left(\tilde{\tilde{Q}}_{\alpha\delta}\Big|\tilde{\tilde{P}}_{\beta\delta}\right) \tag{5.3.25}$$

$$\left(\tilde{\tilde{Q}}_{\alpha\delta}\Big|\tilde{\tilde{P}}_{\beta\delta}\right) = \left(\sum_{kl} f_k(\mathbf{r}_1)[V^{-1}]^{kl}(Q_{\alpha\delta}|f_l)\Bigg|\sum_{ij} f_i(\mathbf{r}_2)[V^{-1}]^{ij}(f_j|P_{\beta\delta})\right) \tag{5.3.26}$$

rearranging this expression using the definition and symmetry of **V** (equation 5.3.19) gives,

$$= \sum_{ijkl}(Q_{\alpha\delta}|f_l)[V^{-1}]^{lk}(f_k|f_i)[V^{-1}]^{ij}(f_j|P_{\beta\delta})$$

$$= \sum_{ijkl}(Q_{\alpha\delta}|f_l)[V^{-1}]^{lk}V_{ki}[V^{-1}]^{ij}(f_j|P_{\beta\delta})$$

$$= \sum_{jl}(Q_{\alpha\delta}|f_l)[V^{-1}]^{lj}(f_j|P_{\beta\delta})$$

$$= \left(Q_{\alpha\delta}\Bigg|\sum_{jl} f_l(\mathbf{r}_2)[V^{-1}]^{lj}(f_j|P_{\beta\delta})\right)$$

$$= \left(Q_{\alpha\delta}\Big|\tilde{\tilde{P}}_{\beta\delta}\right) \tag{5.3.27}$$

in which only one of the products has been expanded. Hence,

$$\left(\tilde{\tilde{Q}}_{\alpha\delta}\Big|\tilde{\tilde{P}}_{\beta\delta}\right) = \left(Q_{\alpha\delta}\Big|\tilde{\tilde{P}}_{\beta\delta}\right). \tag{5.3.28}$$

This is only possible because of how the metric is defined, as an electrostatic integral (equation 5.3.19), this is not a property of other metrics. Without this result the NGWF gradient will not be strictly consistent if only $P_{\beta\delta}$ is expanded. So expanding $Q_{\alpha\delta}$ would also be required for a consistent NGWF gradient. Expanding both products would complicate the calculation of the exchange matrix since two fittings would be required. Expanding only $Q_{\alpha\delta}$ completely negates the advantage of fitting the NGWF product as the inner exchange integral (equation 4.2.16) would have to be calculated without fitting. As an approximation in the overlap metric only a single product is expanded so the NGWF gradient can't be

Table 5.1: Cartesian form of the first few real spherical harmonics. ($\mathbf{r} = (x, y, z)$; $r = |\mathbf{r}|$.)

l	m	$Z_{lm}(\mathbf{r})$
0	0	$\dfrac{1}{\sqrt{4\pi}}$
	-1	$\sqrt{\dfrac{3}{4\pi}}\dfrac{y}{r}$
1	0	$\sqrt{\dfrac{3}{4\pi}}\dfrac{z}{r}$
	1	$\sqrt{\dfrac{3}{4\pi}}\dfrac{x}{r}$
	-2	$\sqrt{\dfrac{15}{4\pi}}\dfrac{xy}{r^2}$
	-1	$\sqrt{\dfrac{15}{4\pi}}\dfrac{yz}{r^2}$
2	0	$\sqrt{\dfrac{5}{16\pi}}\dfrac{3z^2 - 1}{r^2}$
	1	$\sqrt{\dfrac{15}{4\pi}}\dfrac{xz}{r^2}$
	2	$\sqrt{\dfrac{15}{16\pi}}\dfrac{x^2 - y^2}{r^2}$

expected to be fully consistent. This is not a problem for the kernel gradient since making the exchange matrix symmetric fixes this inconsistency. Variational fitting is important and is only provided by the electrostatic metric [166]. So only the electrostatic metric will be expected to be variational with respect to the auxiliary basis set.

5.4 Basis sets for NGWF product expansion

Two options for auxiliary basis sets are considered as part of this work, spherical Gaussians and spherical waves. Both spherical Gaussians and spherical waves involve spherical harmonics so it is helpful to first consider some of their properties.

5.4.1 Spherical harmonics

Real spherical harmonics have been used in this work, the first few are illustrated in table 5.1. The real spherical harmonics (Z_{lm}) are related to the complex spherical harmonics (Y_l^m) by the expressions [167],

$$Z_{l0} = Y_l^0 \tag{5.4.1}$$

$$\left.\begin{aligned} Z_{lm+} &= \tfrac{1}{\sqrt{2}} \left[(-1)^m Y_l^m + Y_l^{-m} \right], \\ Z_{lm-} &= \tfrac{1}{i\sqrt{2}} \left[(-1)^m Y_l^m - Y_l^{-m} \right], \end{aligned}\right\} \qquad m > 0. \tag{5.4.2}$$

Since concentric spherical harmonics are orthogonal the integral of a product of two spherical harmonics (in spherical coordinates) is [168],

$$\int_0^{2\pi} \int_0^{\pi} Z_{lm}(\theta_2, \phi_2) Z_{l'm'}(\theta_2, \phi_2) \sin(\theta_2) d\theta_2 d\phi_2 = \delta_{ll'} \delta_{mm'}. \qquad (5.4.3)$$

5.4.1.1   Legendre polynomial addition theorem

In the calculation of exchange integrals it is helpful to use the Legendre polynomial addition theorem [168],

$$\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} = \sum_{l=0}^{\infty} \frac{r_<^l}{r_>^{l+1}} P_l(\cos \gamma) \qquad (5.4.4)$$

$$P_l(\cos \gamma) = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y_l^m(\theta_1, \phi_1) Y_l^{m*}(\theta_2, \phi_2). \qquad (5.4.5)$$

The complex spherical harmonics in equation 5.4.5 can be substituted for real spherical harmonics since,

$$
\begin{aligned}
Z_{lm}(1)Z_{lm}(2) &+ Z_{l(-m)}(1)Z_{l(-m)}(2) \\
&= \tfrac{1}{2} \left(Y_l^m(1)Y_l^m(2) + (-1)^m Y_l^m(1)Y_l^{-m}(2) \right. \\
&\quad \left. + (-1)^m Y_l^{-m}(1)Y_l^m(2) + Y_l^{-m}(1)Y_l^{-m}(2) \right) \\
&\quad - \tfrac{1}{2} \left(Y_l^m(1)Y_l^m(2) - (-1)^m Y_l^m(1)Y_l^{-m}(2) \right. \\
&\quad \left. - (-1)^m Y_l^{-m}(1)Y_l^m(2) + Y_l^{-m}(1)Y_l^{-m}(2) \right) \\
&= (-1)^m Y_l^m(1)Y_l^{-m}(2) + (-1)^m Y_l^{-m}(1)Y_l^m(2)
\end{aligned}
$$

(5.4.6)

(5.4.7)

using $Y_l^{m*} = (-1)^m Y_l^{-m}$,

$$= Y_l^m(1)Y_l^{m*}(2) + Y_l^{-m}(1)Y_l^{-m*}(2) \qquad (5.4.8)$$

which are the pairs of terms appearing in the sum alongside $Y_l^0(1)Y_l^0(2)$ $(= Z_{l0}(1)Z_{l0}(2))$.

5.4.2   Spherical Gaussians

Gaussians are the obvious choice for the task of fitting a product of NGWFs. It is common in atomic-orbital codes (which usually use a Gaussian basis set) to fit densities and pseudo-densities using specially designed Gaussian basis sets to reduce the cost of two-electron integral evaluation [169]. The use of a Gaussian basis set in this role allows the existing efficient two-electron integral engines in these codes to continue to be used. In calculations employing Gaussian basis sets the use of an Gaussian auxiliary basis set is also natural because the product of two Gaussians functions is also a Gaussian function. There are two

forms of spherical Gaussians that appear in the literature,

$$G_{lm}(\mathbf{r}) = N_l r^l e^{-\alpha r^2} Z_{lm}(\hat{\mathbf{r}}) \tag{5.4.9}$$

$$G_{lmn}(\mathbf{r}) = N_l r^n e^{-\alpha r^2} Z_{lm}(\hat{\mathbf{r}}) \tag{5.4.10}$$

where n is the principal quantum number. The latter form is more general since if $l = n$ the former form is obtained. The former form will be used in subsequent expressions.

5.4.3 Spherical waves

The NGWF product that is to be expanded can be written in terms of fine grid psinc functions which can be expressed as a finite sum of plane waves. Plane waves ($e^{i\mathbf{g}\cdot\mathbf{r}}$) are solutions to the Helmholtz equation (the time-independent free-electron Schrödinger equation),

$$(\nabla^2 + q^2)\psi = 0 \tag{5.4.11}$$

(in a box) with periodic boundary conditions ($E = \frac{1}{2}q^2$). Truncated spherical waves are solutions to the Helmholtz equation with boundary conditions enforcing spherical localisation (particle in a sphere of radius a),

$$\psi(\mathbf{r}) = \begin{cases} j_l(qr)Z_{lm}(\hat{\mathbf{r}}) & r < a \\ 0 & r \geq a \end{cases} \tag{5.4.12}$$

where $j_l(qr)$ is a spherical Bessel function and $Z_{lm}(\hat{\mathbf{r}})$ is a real spherical harmonic. The value of q is chosen so that $j_l(qa) = 0$, ensuring that the truncation of the Bessel function does not introduce a discontinuity. The spherical waves are eigenfunctions of the kinetic energy operator within the localisation region with eigenvalue $E = \frac{1}{2}q^2$. Therefore the same kinetic energy cutoff that determines the plane wave basis (and in ONETEP the psinc spacing) can be used to restrict the values of q and l in the spherical wave basis. Since plane waves and spherical waves are solutions to the same equation with different boundary conditions, a set of spherical waves can be expected to be a suitable basis set to expand a quantity expressed in plane waves, already localised in spherical regions [170].

5.4.3.1 Spherical Bessel functions

The Bessel functions are so named because they are solutions of Bessel's equation,

$$x^2 \frac{d^2 y(kx)}{dx^2} + x \frac{dy(kx)}{dx} + \left(k^2 x^2 - n^2\right) y(kx) = 0 \qquad n \in \mathbb{Z}, k \in \mathbb{R}. \tag{5.4.13}$$

There are various types of Bessel functions satisfying different formulations of Bessel's equation but in this work only the spherical Bessel functions of the first kind are used. These are solutions to the spherical Bessel equation,

$$r^2 \frac{d^2 y(kr)}{dr^2} + r \frac{dy(kr)}{dr} + \left(k^2 r^2 - \left(n + \frac{1}{2}\right)^2\right) y(kr) = 0 \qquad n \in \mathbb{Z}, k \in \mathbb{R} \tag{5.4.14}$$
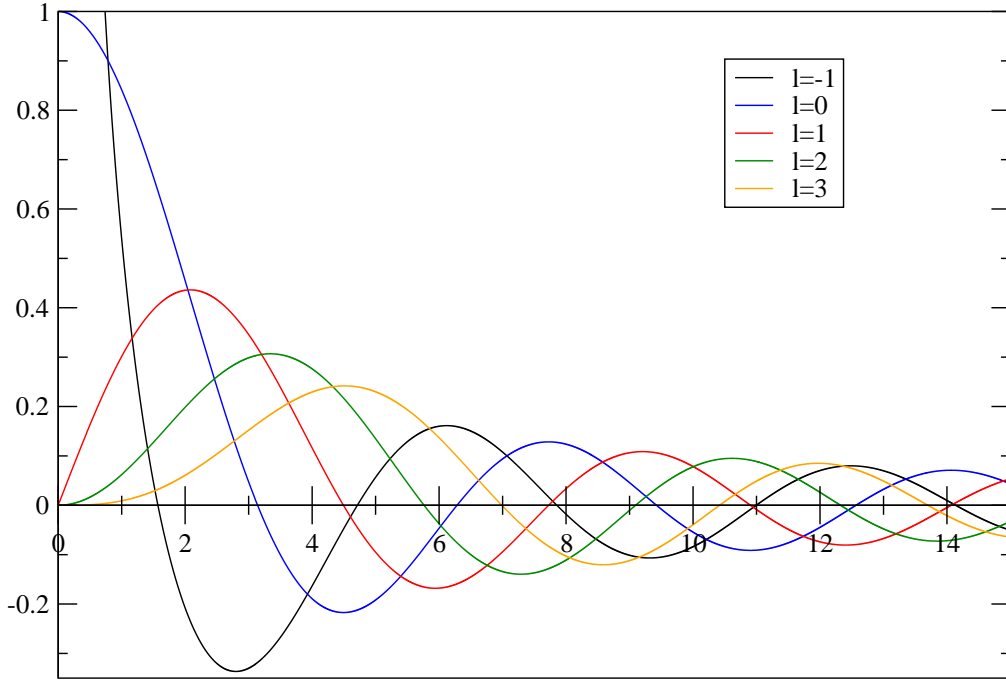
Figure 5.3: Graph of several spherical Bessel functions ($j_l(x)$).

which arises (as the radial part) when the Helmholtz equation (equation 5.4.11) is separated in spherical polar coordinates [168] .

The expressions for the first few spherical Bessel functions are,

$$j_{-1}(x) = \frac{\cos x}{x} \tag{5.4.15}$$

$$j_0(x) = \frac{\sin x}{x} \tag{5.4.16}$$

$$j_1(x) = \frac{\sin x}{x^2} - \frac{\cos x}{x} \tag{5.4.17}$$

$$j_2(x) = \left(\frac{3}{x^3} - \frac{1}{x}\right) \sin x - \frac{3\cos x}{x^2} \tag{5.4.18}$$

$$j_3(x) = \left(\frac{15}{x^4} - \frac{6}{x^2}\right) \sin x - \left(\frac{15}{x^3} - \frac{1}{x}\right) \cos x \tag{5.4.19}$$

these functions are also illustrated in figure 5.3. The spherical Bessel functions can be generated using the following expression (one of the Rayleigh formulas),

$$j_n(x) = (-x)^n \left(\frac{1}{x}\frac{d}{dx}\right)^n \left(\frac{\sin(x)}{x}\right). \tag{5.4.20}$$

The spherical Bessel functions obey the following orthogonality relation,

$$\int_0^a j_n\left(\alpha_{np}\frac{r}{a}\right) j_n\left(\alpha_{nq}\frac{r}{a}\right) r^2 dr = \frac{r^3}{2} \left(j_{n+1}(\alpha_{np})\right)^2 \delta_{pq} \tag{5.4.21}$$

where α_{np} and α_{nq} zeroes or roots of j_n, i.e. $j_n(\alpha_{np}) = j_n(\alpha_{nq}) = 0$ [168].

5.4.3.2 Fourier transform of a spherical wave

The spherical waves may be constructed in reciprocal space and Fourier transformed back to real space. The Fourier transform of a function is defined as,

$$\tilde{f}(\mathbf{g}) = \int_{\mathrm{AS}} e^{-i\mathbf{g}\cdot\mathbf{r}} f(\mathbf{r}) d\mathbf{r} \tag{5.4.22}$$

where AS denotes that the integral is over all space. Applying this expression to a spherical wave χ_{nlm} gives,

$$\tilde{\chi}_{nlm}(\mathbf{g}) = \int_0^a r^2 j_l(q_{nl}r) \int e^{-i\mathbf{g}\cdot\mathbf{r}} Z_{lm}(\Omega_{\mathbf{r}}) d\Omega dr \tag{5.4.23}$$

where a is the cutoff radius of the spherical Bessel function (see equation 5.4.12). An expansion for $e^{-i\mathbf{g}\cdot\mathbf{r}}$ in terms of spherical harmonics and Bessel functions exists,

$$e^{-i\mathbf{g}\cdot\mathbf{r}} = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} i^{-l} j_l(gr) Z_{lm}(\Omega_{\mathbf{g}}) Z_{lm}(\Omega_{\mathbf{r}}). \tag{5.4.24}$$

Inserting this into equation 5.4.23 and evaluating the angular integral using equation 5.4.3 gives,

$$\tilde{\chi}_{nlm}(\mathbf{g}) = 4\pi i^{-l} Z_{lm}(\Omega_{\mathbf{g}}) \int_0^a r^2 j_l(q_{nl}r) j_l(gr) dr \tag{5.4.25}$$

the radial integral can now be evaluated by considering two cases [170],

$$= 4\pi i^{-l} Z_{lm}(\Omega_{\mathbf{g}}) \begin{cases} \frac{q_{nl}a^2}{g^2 - q_{nl}^2} j_l(ga) j_{l-1}(q_{nl}a) & g \neq q_{nl} \\ \frac{q_{nl}a^3}{g + q_{nl}} j_{l-1}^2(q_{nl}a) & g = q_{nl}. \end{cases} \tag{5.4.26}$$

5.5 Calculating exchange using a Gaussian fitting basis set

A Gaussian fitting basis set was used to expand the NGWF product. Firstly an analytic expression for the integral in equation 5.3.2 for Gaussians is needed.

5.5.1 Potential integrals of Gaussian functions

Each Gaussian function in the expansion of the NGWF product in equation 5.3.1 will give rise to an integral that needs to be evaluated. Since a large number of integral evaluations will be required it is essential that the integration can be performed efficiently. An outline of the evaluation of the integrals follows, starting from,

$$U_{\mathrm{G}}(\mathbf{r}_1) = \int \frac{G_{l'}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \tag{5.5.1}$$

$$= \int \frac{N_{l'} r_2^{l'} e^{-\alpha r_2^2} Z_{l'm'}(\hat{\mathbf{r}}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \tag{5.5.2}$$

now substituting equations 5.4.4 and 5.4.5 into equation 5.5.2 and using spherical polar coordinates,

$$
= \int_0^\infty N_{l'} r_2^{l'+2} e^{-\alpha r_2^2} \sum_{l=0}^\infty \frac{r_<^l}{r_>^{l+1}} \frac{4\pi}{2l+1} \sum_{m=-l}^l Z_{lm}(\theta_1, \phi_1) dr_2
$$
$$
\int_0^{2\pi} \int_0^\pi Z_{lm}(\theta_2, \phi_2) Z_{l'm'}(\theta_2, \phi_2) \sin(\theta_2) d\theta_2 d\phi_2
\tag{5.5.3}
$$

where $r_<$ is the lower of r_1 and r_2 and $r_>$ is the higher. Equation 5.4.3 is used to obtain,

$$
U_G(\mathbf{r}_1) = \left(\int_0^\infty N_{l'} r_2^{l'+2} e^{-\alpha r_2^2} \frac{r_<^{l'}}{r_>^{l'+1}} dr_2 \right) \left(\frac{4\pi}{2l'+1} Z_{l'm'}(\theta_1, \phi_1) \right).
\tag{5.5.4}
$$

This integral can be evaluated by splitting the range of integration into two intervals (and dropping the primes),

$$
= N_l \frac{4\pi}{2l+1} Z_{lm}(\theta_1, \phi_1) \left(r_1^{-l-1} \int_0^{r_1} r_2^{2l+2} e^{-\alpha r_2^2} dr_2 + r_1^l \int_{r_1}^\infty r_2 e^{-\alpha r_2^2} dr_2 \right).
\tag{5.5.5}
$$

The integral in the interval $[0, r_1]$ can be solved by repeated integration by parts. For example for $l = 1$,

$$
\int_0^{r_1} r_2^4 e^{-\alpha r_2^2} dr_2 = -\frac{r_1^3 e^{-\alpha r_1^2}}{2\alpha} + \frac{3}{2\alpha} \int_0^{r_1} r_2^2 e^{-\alpha r_2^2} dr_2
\tag{5.5.6}
$$
$$
= -\frac{r_1^3 e^{-\alpha r_1^2}}{2\alpha} - \frac{3r_1 e^{-\alpha r_1^2}}{4\alpha^2} + \frac{3}{4\alpha^2} \int_0^{r_1} e^{-\alpha r_2^2} dr_2
\tag{5.5.7}
$$
$$
= -\frac{r_1^3 e^{-\alpha r_1^2}}{2\alpha} - \frac{3r_1 e^{-\alpha r_1^2}}{4\alpha^2} + \frac{3\sqrt{\pi}}{8\alpha^2\sqrt{\alpha}} \operatorname{erf}(\sqrt{\alpha} r_1).
\tag{5.5.8}
$$

where $\operatorname{erf}(r)$ is the error function. Note that the integral on the right hand side of equation 5.5.6 is the same integral that appears in the case where $l = 0$. This could be used to allow efficient evaluation of the integrals. The integral over the interval $[r_1, \infty]$ can be solved analytically,

$$
\int_{r_1}^\infty r_2 e^{-\alpha r_2^2} dr_2 = \frac{e^{-\alpha r_1^2}}{2\alpha}.
\tag{5.5.9}
$$

The integral in this equation is much more complicated to solve when the spherical Gaussians are of the form given in equation 5.4.10.

5.5.2   NGWF product expansion with Gaussians

It was found that the a set of Gaussian functions poorly represented the NGWF product with the overlap metric. This was caused by problems with the representation of the Gaussians on a grid. Suitable Gaussian exponents must satisfy two conflicting requirements. If the Gaussian exponent is too large then the resulting function will be too sharp to represent on the grid, so small exponents are desirable. However if the Gaussian exponent
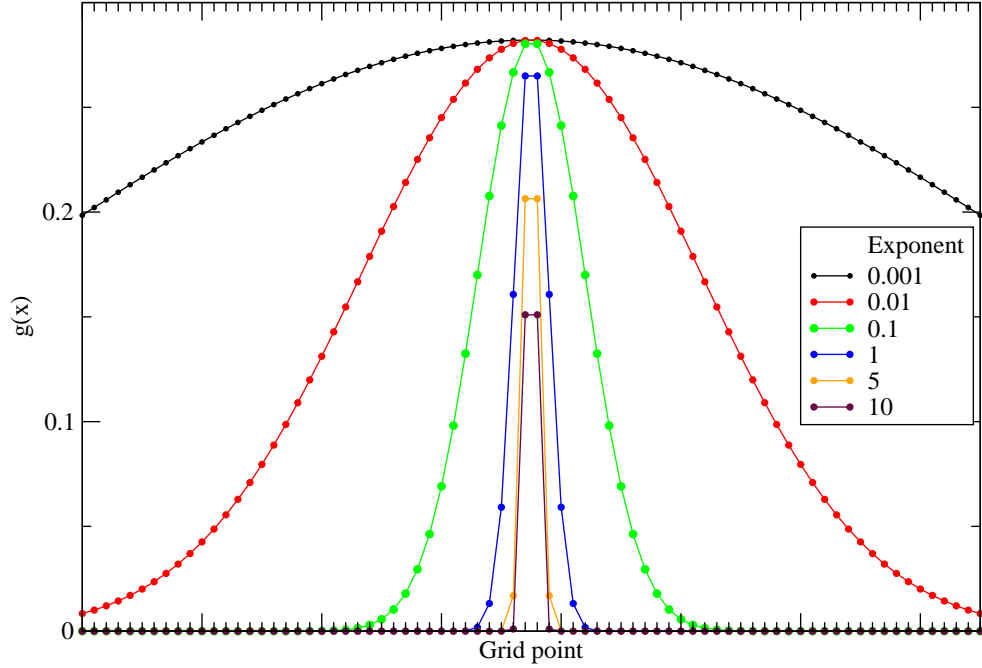
Figure 5.4: Representation of several Gaussian functions on a grid. It is obvious that not all these functions are adequately represented on the grid. In the code only those exponents that are well represented are retained.

is too small then the resulting function will be too diffuse and truncation of the function at the tightbox boundary will introduce errors. To quantify the suitability of an exponent the normalised overlap integral was calculated by integrating numerically at every point in the tightbox. Inspection of the diagonal elements of this matrix (which should be unity) gave a measure of the suitability of each exponent (by choice of an acceptable error bound). Following this a lower bound on the exponent based on the size of the tightbox was applied, this eliminates exponents which lead to excessively diffuse functions. The upper bound ensures that the half-width of the resulting Gaussian is appropriate for the grid spacing (which in turn depends on the cutoff energy). If the exponent is too large then the Gaussian peak will be too sharp to reproduce on the grid, this is illustrated in figure 5.4 where it is clear that the exponents larger than 1 are not represented accurately by the grid used in the graph. Figure 5.4 also shows that for exponents which are too small the Gaussian function still has a significant value at the tightbox boundary. The Gaussian overlap matrix elements calculated analytically involving such a Gaussian will therefore not be accurate, since there is a significant contribution to the overlap from outside of the tightbox and the code is only using the portion of the Gaussian inside the tightbox. This will have detrimental effect on the accuracy of the expansion coefficients of the Gaussians (which are calculated by solving the series of linear equations defined by the Gaussian overlap matrix). If the Gaussian overlap metric matrix were calculated on the grid of the NGWF product, the poor representation of Gaussians on the grid would also cause this to be inaccurate. Also since, on the coarse grid, the NGWF product is strictly localised inside the NGWF sphere, when this is overlapped with a Gaussian only the portion of the Gaussian within the
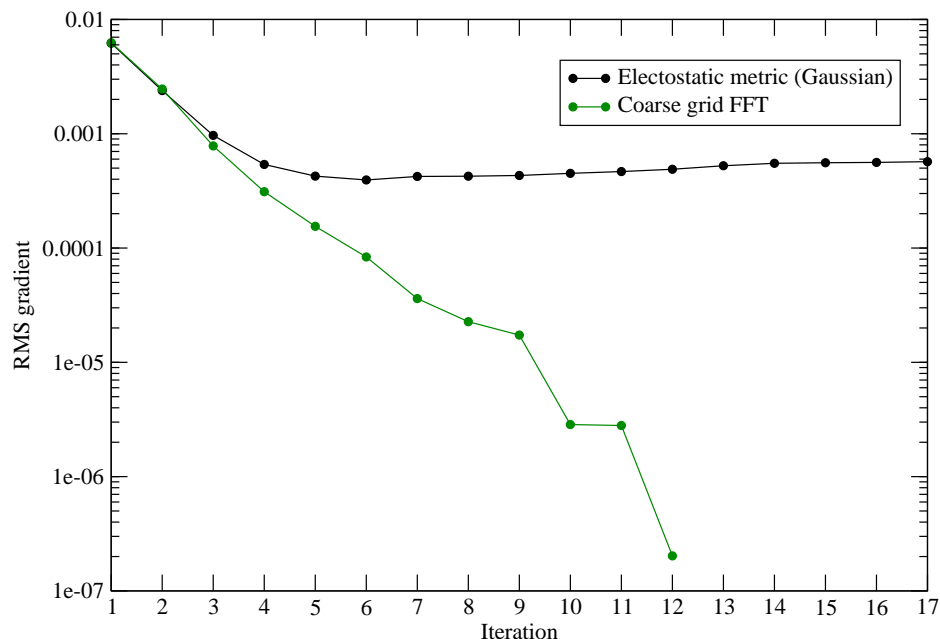
Figure 5.5: NGWF convergence of ONETEP Hartree-Fock single-point energy calculations for water using the Gaussian electrostatic metric and the coarse grid FFT methods.

NGWF sphere is considered the other part is ignored. These problems prevent an accurate fitting of the product to a set of Gaussians, which subsequently leads to poor energies and gradients.

The available Gaussian basis sets are usually designed to either represent atomic orbitals or products of Gaussians in the case of density-fitting basis sets. These sets are not designed to represent localised NGWFs products. It is hard to construct an auxiliary basis set that is suitable for this purpose. Linear dependence is a problem encountered by other approaches [160, 161]. Linear independence of the basis functions is hard to achieve with such a restricted set of admissible Gaussian exponents. The constructed basis sets must not be linearly dependent (or nearly linearly dependent). It would be useful to be able to adjust the size of the constructed fitting set variationally to control the accuracy.

The electrostatic metric avoids some of these problems. An atom centred spherical grid has been used to calculated the metric matrix elements [171] which avoids the need to represent Gaussians on the same Cartesian grid as the NGWF product. However the problem of constructing a suitable basis set still remains.

5.5.3   Results

The Ahlrichs carbon Coulomb fitting basis set [160] has been used for the calculations in this section. The Gaussian overlap metric method did not converge due to the problems discussed above. Figure 5.5 shows the NGWF convergence of the Gaussian electrostatic method for water. The graph shows that convergence is very poor for this method. NGWF convergence completely broke down for ethene. Figure 5.6 presents the LNV convergence of this method. Convergence for both ethene and water was poor. In view of the problems

with the use of a Gaussian auxiliary basis set an alternative basis set will used instead.

5.6 Calculating exchange using a spherical waves fitting basis set

Spherical waves were subsequently employed as an alternative fitting basis set. As discussed in section 5.4.3 spherical waves are expected to be a suitable basis set due to their link with plane waves.

5.6.1 Choice of basis set parameters

There are a few parameters that needed to be selected to determine the extent of the spherical wave basis set. One of these parameters is the maximum value of l. Larger values of l_{max} can be expected to give a more accurate fit to the NGWF product as more degrees of freedom are provided. However increasing l_{max} significantly increases the number of spherical waves in the calculation and hence the computational cost. Also the increased complexity of the spherical waves with large l values may not be well represented on the grid. The number of Bessel zeroes included can also be adjusted. Only the first few Bessel zeroes give rise to spherical waves that can be represented on a grid. There is also a computational cost to using a large number of Bessel zeros. In section 5.4.3 it was suggested that in theory the parameters can be determined by the kinetic energy cutoff, in practice a stricter cutoff may be needed. However over-restricting the number of zeroes leads to an inaccurate fitting of the NGWF product. An inaccurate fit to the NGWF product affects the accuracy of the Hartree-Fock energy calculation. It is important that a sufficient number of spherical waves are included to give an accurate energy. To select suitable values for these parameters the convergence and energies obtained from a series of calculations were compared. Calculations were performed on ammonia, benzene, ethene and water for $l_{max} = 2, 3, 4$ and the number of zeros restricted to values between 5 and 14. These parameters were chosen independently for the two metrics and these will be presented in sections 5.6.3.1 and 5.6.4.1.

5.6.2 Potential integrals of spherical waves

The general form of the potential integrals of a spherical wave on a single centre (located at $\mathbf{r} = \mathbf{0}$) is,

$$U_{\text{SW}}(\mathbf{r_1}) = \int \frac{j_l(q_{nl}r_2)Z_{lm}(\hat{\mathbf{r}}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r_2}. \tag{5.6.1}$$
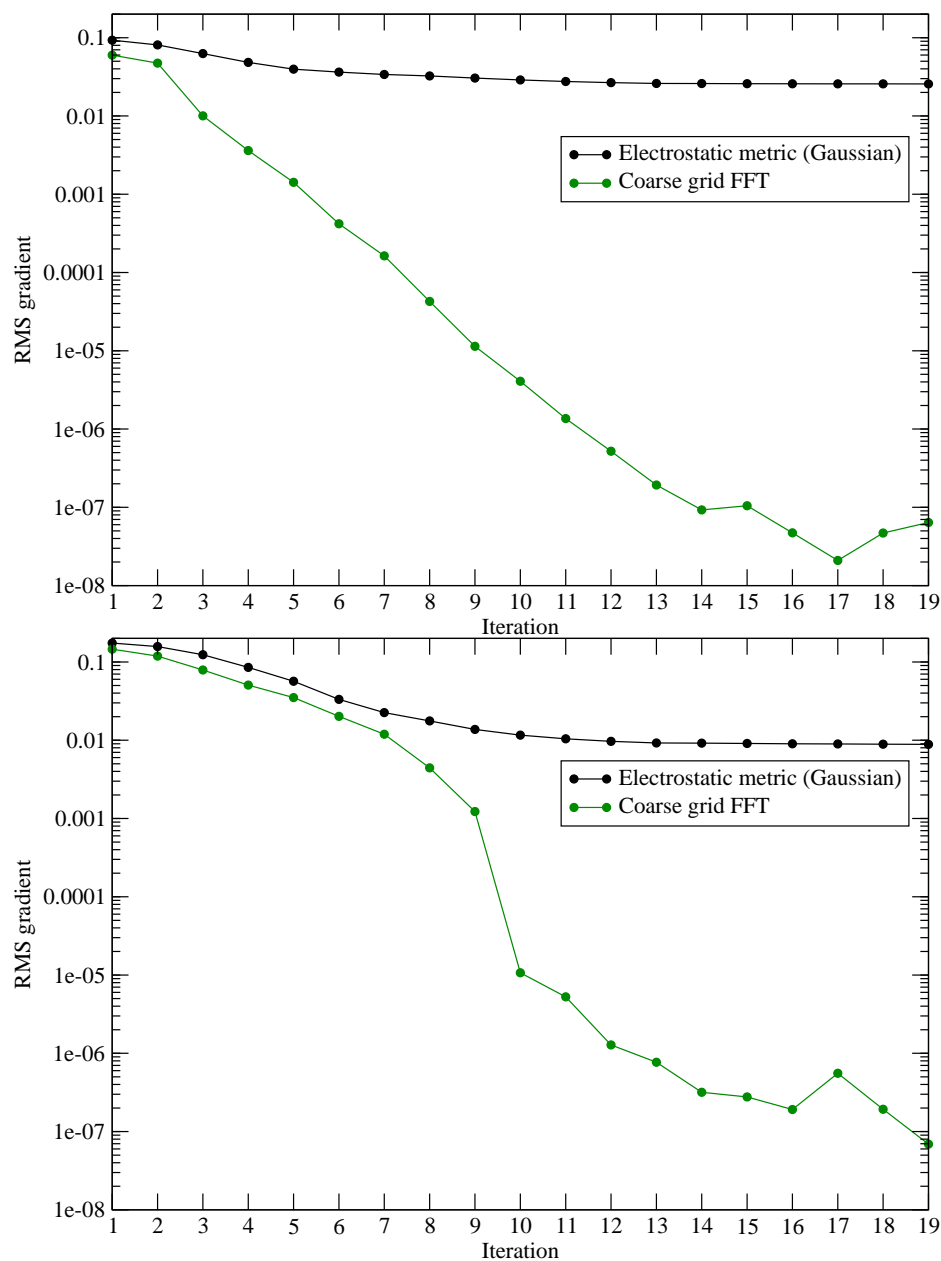
Figure 5.6: LNV convergence of ONETEP Hartree-Fock single-point energy calculations for ethene (top) and water (bottom) using the Gaussian electrostatic metric and the coarse grid FFT methods.

The integral can be evaluated by substituting equations 5.4.4 and 5.4.5 into equation 5.6.1 and using spherical polar coordinates giving,

$$U_{\text{SW}}(\mathbf{r}_1) = \int_0^a j_l(q_{nl}r_2)r_2^2 \sum_{l'=0}^{\infty} \frac{r_<^{l'}}{r_>^{l'+1}} \frac{4\pi}{2l'+1} \sum_{m'=-l'}^{l'} Z_{l'm'}(\theta_1, \phi_1) dr_2$$

$$\int_0^{2\pi} \int_0^{\pi} Z_{lm}(\theta_2, \phi_2) Z_{l'm'}(\theta_2, \phi_2) \sin(\theta_2) d\theta_2 d\phi_2 \tag{5.6.2}$$

where $r_<$ is the lower of r_1 and r_2 and $r_>$ is the higher (a is the cut off radius of the spherical wave as defined in equation 5.4.12); using the orthogonality relation in equation 5.4.3 gives,

$$U_{\text{SW}}(\mathbf{r}_1) = \left(\int_0^a r_2^2 j_l(q_{nl}r_2) \frac{r_<^l}{r_>^{l+1}} dr_2 \right) \left(\frac{4\pi}{2l+1} Z_{lm}(\theta_1, \phi_1) \right). \tag{5.6.3}$$

This integral can be evaluated by splitting the range of integration into two intervals if necessary (depending on the value of r_1),

$$U_{\text{SW}} = \frac{4\pi}{2l+1} Z_{lm}(\theta_1, \phi_1) \left(\begin{cases} r_1^{-l-1} \int_0^{r_1} r_2^{2+l} j_l(qr_2) dr_2 + r_1^l \int_{r_1}^a r_2^{1-l} j_l(qr_2) dr_2 & r_1 < a \\ r_1^{-l-1} \int_0^a r_2^{2+l} j_l(qr_2) dr_2 & r_1 \geq a \end{cases} \right). \tag{5.6.4}$$

These integrals can be calculated analytically, for example for $l = 1$, when $r_1 < a$,

$$U_{\text{SW}}^{l=1}(\mathbf{r}_1) = \frac{4\pi}{3} Z_{1m}(\theta_1, \phi_1) \left(r_1^{-2} \int_0^{r_1} r_2^3 j_1(qr_2) dr_2 + r_1^1 \int_{r_1}^a r_2^0 j_1(qr_2) dr_2 \right) \tag{5.6.5}$$

$$= \frac{4\pi}{3} Z_{1m}(\theta_1, \phi_1) \left(r_1^{-2} \int_0^{r_1} r_2^3 \left(\frac{\sin(qr_2)}{(qr_2)^2} - \frac{\cos(qr_2)}{qr_2} \right) dr_2 \right.$$
$$\left. + r_1^1 \int_{r_1}^a r_2^0 \left(\frac{\sin(qr_2)}{(qr_2)^2} - \frac{\cos(qr_2)}{qr_2} \right) dr_2 \right) \tag{5.6.6}$$

$$= \frac{4\pi}{3} Z_{1m}(\theta_1, \phi_1) \left(\int_0^{r_1} \left(\frac{r_2 \sin(qr_2)}{(qr_1)^2} - \frac{r_2^2 \cos(qr_2)}{qr_1^2} \right) dr_2 \right.$$
$$\left. + \int_{r_1}^a \left(\frac{r_1 \sin(qr_2)}{(qr_2)^2} - \frac{r_1 \cos(qr_2)}{qr_2} \right) dr_2 \right) \tag{5.6.7}$$

$$= \frac{4\pi}{3} Z_{1m}(\theta_1, \phi_1) \left(\left[-\frac{3r_2 \cos(qr_2)}{q^3 r_1^2} - \frac{(q^2 r_2^2 - 3) \sin(qr_2)}{q^4 r_1^2} \right]_0^{r_1} \right.$$
$$\left. + \left[-\frac{r_1 \sin(qr_2)}{q^2 r_2} \right]_{r_1}^a \right) \tag{5.6.8}$$

$$= \frac{4\pi}{3} Z_{1m}(\theta_1, \phi_1) \left(-\frac{3r_1 \cos(qr_1)}{q^3 r_1^2} - \frac{(q^2 r_1^2 - 3) \sin(qr_1)}{q^4 r_1^2} - \frac{r_1 \sin(qa)}{q^2 a} \right.$$
$$\left. + \frac{r_1 \sin(qr_2)}{q^2 r_1} \right) \tag{5.6.9}$$

$$= \frac{4\pi}{3} Z_{1m}(\theta_1, \phi_1) \left(-\frac{3 \cos(qr_1)}{q^3 r_1} + \frac{3 \sin(qr_1)}{q^4 r_1^2} - \frac{r_1 \sin(qa)}{q^2 a} \right). \tag{5.6.10}$$

When $r_1 \geq a$ the resulting expression is simpler,

$$U_{\mathrm{SW}}^{l=1}(\mathbf{r}_1) = \frac{4\pi}{3} Z_{1m}(\theta_1, \phi_1) r_1^{-2} \int_0^a r_2^3 j_1(qr_2) dr_2 \tag{5.6.11}$$

$$= \frac{4\pi}{3} Z_{1m}(\theta_1, \phi_1) r_1^{-2} \int_0^a r_2^3 \left(\frac{\sin(qr_2)}{(qr_2)^2} - \frac{\cos(qr_2)}{qr_2} \right) dr_2 \tag{5.6.12}$$

$$= \frac{4\pi}{3} Z_{1m}(\theta_1, \phi_1) \left[-\frac{3r_2 \cos(qr_2)}{q^3 r_1^2} - \frac{(q^2 r_2^2 - 3)\sin(qr_2)}{q^4 r_1^2} \right]_0^a \tag{5.6.13}$$

$$= \frac{4\pi}{3} Z_{1m}(\theta_1, \phi_1) \left(-\frac{3a \cos(qa)}{q^3 r_1^2} - \frac{(q^2 a^2 - 3)\sin(qa)}{q^4 r_1^2} \right) \tag{5.6.14}$$

$$= \frac{4\pi}{3} Z_{1m}(\theta_1, \phi_1) r_1^{-2} \left(-\frac{3a \cos(qa)}{q^3} - \frac{(q^2 a^2 - 3)\sin(qa)}{q^4} \right). \tag{5.6.15}$$

The formulae for other values of l are, for $r_1 < a$,

$$U_{\mathrm{SW}}^{l=0}(\mathbf{r}_1) = 4\pi Z_{00}(\theta_1, \phi_1) \left(\frac{\sin(qr_1)}{q^3 r_1} - \frac{\cos(qa)}{q^2} \right) \tag{5.6.16}$$

$$U_{\mathrm{SW}}^{l=2}(\mathbf{r}_1) = \frac{4\pi}{3} Z_{2m}(\theta_1, \phi_1) \left(r_1^2 \frac{qa \cos(qa) - \sin(qa)}{q^3 a^3} \right.$$
$$\left. - \frac{15 \cos(qr_1)}{q^4 r_1^2} - \frac{5(q^2 r^2 - 3)\sin(qr_1)}{q^5 r_1^3} \right) \tag{5.6.17}$$

$$U_{\mathrm{SW}}^{l=3}(\mathbf{r}_1) = \frac{4\pi}{5} Z_{3m}(\theta_1, \phi_1) \left(\frac{(-105 + 7q^2 r_1^2) \cos(qr_1)}{q^5 r_1^3} \right.$$
$$\left. + \frac{(105 - 42q^2 r_1^2)\sin(qr_1)}{q^6 r_1^4} + r_1^3 \frac{3qa \cos(qa) + (q^2 a^2 - 3)\sin(qa)}{q^4 a^5} \right) \tag{5.6.18}$$

$$U_{\mathrm{SW}}^{l=4}(\mathbf{r}_1) = \frac{4\pi}{7} Z_{4m}(\theta_1, \phi_1) \left(\frac{(90q^2 r_1^2 - 945) \cos(qr_1)}{q^6 r_1^4} \right.$$
$$+ \frac{(9q^4 r_1^4 - 405q^2 r_1^2 + 945)\sin(qr_1)}{q^7 r_1^5}$$
$$\left. + r_1^4 \frac{qa(15 - q^2 a^2) \cos(qa) + (6q^2 a^2 - 15)\sin(qa)}{q^5 a^7} \right) \tag{5.6.19}$$

and for $r_1 \geq a$,

$$U_{\mathrm{SW}}^{l=0}(\mathbf{r}_1) = 4\pi Z_{00}(\theta_1, \phi_1) \frac{\sin(qa) - qa \cos(qa)}{q^3} r_1^{-1} \tag{5.6.20}$$

$$U_{\mathrm{SW}}^{l=2}(\mathbf{r}_1) = \frac{4\pi}{3} Z_{2m}(\theta_1, \phi_1) \frac{qa(q^2 a^2 - 15) \cos(qa) - (6q^2 a^2 - 15)\sin(qa)}{q^5} r_1^{-3} \tag{5.6.21}$$

$$U_{\mathrm{SW}}^{l=3}(\mathbf{r}_1) = \frac{4\pi}{5} Z_{3m}(\theta_1, \phi_1) \frac{qa(10q^2 a^2 - 105) \cos(qa) - (q^4 a^4 - 45q^2 a^2 + 105)\sin(qa)}{q^6} r_1^{-4} \tag{5.6.22}$$

$$U_{\mathrm{SW}}^{l=4}(\mathbf{r}_1) = \frac{4\pi}{7} Z_{4m}(\theta_1, \phi_1) \left(\frac{-a(q^4 a^4 - 105q^2 a^2 + 945) \cos(qa)}{q^6} \right.$$
$$\left. + \frac{(15q^4 a^4 - 420q^2 a^2 + 945)\sin(qa)}{q^7} \right) r_1^{-5}. \tag{5.6.23}$$

For all values of l when $r_1 \geq a$ the potential has the form $(CZ_{lm}r^{-l-1})$, which is relatively cheap to compute. This compares favourably to the expressions (involving exponential and error functions) arising from the evaluation of the potential from a set of Gaussians.

5.6.3 Overlap metric for spherical waves

The simplest metric to implement with spherical waves is the overlap metric. As discussed in section 5.3.1, only spherical waves centred on centre B, functions centred on centre C are not included. The use of spherical waves centred on atom A would also not make sense because these would only contribute when atoms A and B overlap. So the set of spherical waves used to expand the NGWF product would vary depending on the position of A. Obviously the drawbacks discussed above associated with using spherical waves on the C atoms would also apply. Due to the orthogonality conditions in equations 5.4.3 and 5.4.21, and the use of spherical waves centred only on atom B, $\mathbf{V}$ will be diagonal,

$$V_{ij} = \int f_i(\mathbf{r})f_j(\mathbf{r})d\mathbf{r} \tag{5.6.24}$$

$$= \delta_{ij} \int f_i(\mathbf{r})f_i(\mathbf{r})d\mathbf{r}. \tag{5.6.25}$$

So the expression for the expansion coefficients simplifies to,

$$c_i = \frac{\int f_i(\mathbf{r})P_{\beta\delta}(\mathbf{r})d\mathbf{r}}{\int f_i(\mathbf{r})f_i(\mathbf{r})d\mathbf{r}}. \tag{5.6.26}$$

5.6.3.1 Choice of basis set parameters

Since the spherical overlap metric is unable to converge to the desired RMS NGWF gradient threshold (2×10^{-6}) the convergence with different basis set parameters was important to assess. These convergence tests as described in section 5.6.1 are presented in table 5.2. When comparing the energies in table 5.3, it can be seen that this method is not variational with respect to the basis set size. A compromise has been made with $l_{\mathrm{max}} = 4$ and 10 Bessel zeros the chosen parameters.

5.6.3.2 Convergence

Convergence problems were found with the use of the overlap metric. These can clearly be seen in figure 5.7. The convergence of the overlap metric matches that of the FFT method for the first few iterations. However subsequently convergence stalls without reaching the threshold. The LNV convergence behaviour for ethene and water is shown in figure 5.8. For both molecules the LNV convergence was poor and stalled at a high RMS gradient. This poor convergence is partially caused by the inherent deficiencies of the overlap metric described in section 5.3.

Table 5.2: Convergence (RMS NGWF gradient in $E_h/(a^{3/2}10^{-6})$) with different basis set parameters using the spherical wave electrostatic metric method for: (a) ammonia; (b) benzene; (c) ethene; (d) water.

(a)	l_{max}			(b)	l_{max}		
zeros	2	3	4	zeros	2	3	4
5	86.70	53.40	35.46	5	133.77	117.08	81.90
6	49.67	5.83	8.28	6	4.96	7.17	70.10
7	80.85	50.13	35.48	7	82.10	85.91	54.27
8	65.71	26.45	4.38	8	38.20	63.27	34.97
9	66.28	40.03	22.38	9	48.76	66.85	42.01
10	64.42	36.34	15.97	10	46.47	64.84	37.15
11	63.95	38.51	17.04	11	52.80	73.93	45.94
12	63.86	35.72	15.94	12	44.25	62.79	36.86
13	65.00	38.91	17.80	13	51.06	69.62	44.24
14	62.51	34.72	13.71	14	47.17	64.68	40.17
(c)	l_{max}			(d)	l_{max}		
zeros	2	3	4	zeros	2	3	4
5	76.87	84.45	64.52	5	92.55	45.31	22.64
6	31.47	11.04	14.48	6	60.90	10.96	3.30
7	244.51	61.40	44.60	7	91.92	57.54	35.97
8	45.53	38.70	21.74	8	79.79	36.82	5.63
9	42.34	42.54	35.89	9	88.41	50.61	18.66
10	37.27	40.61	21.64	10	83.56	42.44	18.57
11	40.98	49.17	34.02	11	87.99	49.96	19.81
12	36.35	41.11	20.80	12	81.23	41.03	13.71
13	40.77	47.87	30.57	13	86.96	45.83	18.38
14	37.15	42.34	23.95	14	83.08	42.05	15.95

Table 5.3: Total energy in E_h with different basis set parameters for ONETEP Hartree-Fock calculations using the spherical wave overlap metric method for: (a) ammonia; (b) benzene; (c) ethene; (d) water.

(a)	l_{max}			(b)	l_{max}		
zeros	2	3	4	zeros	2	3	4
5	-11.421	-11.430	-11.431	5	-36.788	-36.834	-36.843
6	-11.379	-11.391	-11.394	6	-36.706	-36.767	-36.783
7	-11.443	-11.454	-11.457	7	-36.776	-36.836	-36.855
8	-11.437	-11.449	-11.452	8	-36.773	-36.830	-36.847
9	-11.447	-11.458	-11.461	9	-36.782	-36.841	-36.859
10	-11.454	-11.466	-11.469	10	-36.792	-36.852	-36.871
11	-11.449	-11.460	-11.464	11	-36.780	-36.839	-36.857
12	-11.459	-11.471	-11.474	12	-36.798	-36.859	-36.877
13	-11.451	-11.463	-11.466	13	-36.784	-36.844	-36.861
14	-11.459	-11.471	-11.474	14	-36.796	-36.857	-36.875

(c)	l_{max}			(d)	l_{max}		
zeros	2	3	4	zeros	2	3	4
5	-13.397	-13.403	-13.404	5	-16.813	-16.818	-16.819
6	-13.340	-13.357	-13.362	6	-16.766	-16.773	-16.775
7	-13.301	-13.397	-13.402	7	-16.877	-16.884	-16.885
8	-13.374	-13.388	-13.392	8	-16.869	-16.876	-16.877
9	-13.383	-13.397	-13.401	9	-16.895	-16.901	-16.903
10	-13.380	-13.395	-13.400	10	-16.905	-16.912	-16.913
11	-13.383	-13.396	-13.401	11	-16.902	-16.909	-16.910
12	-13.383	-13.398	-13.404	12	-16.917	-16.925	-16.926
13	-13.382	-13.396	-13.401	13	-16.907	-16.915	-16.916
14	-13.384	-13.399	-13.404	14	-16.919	-16.926	-16.928
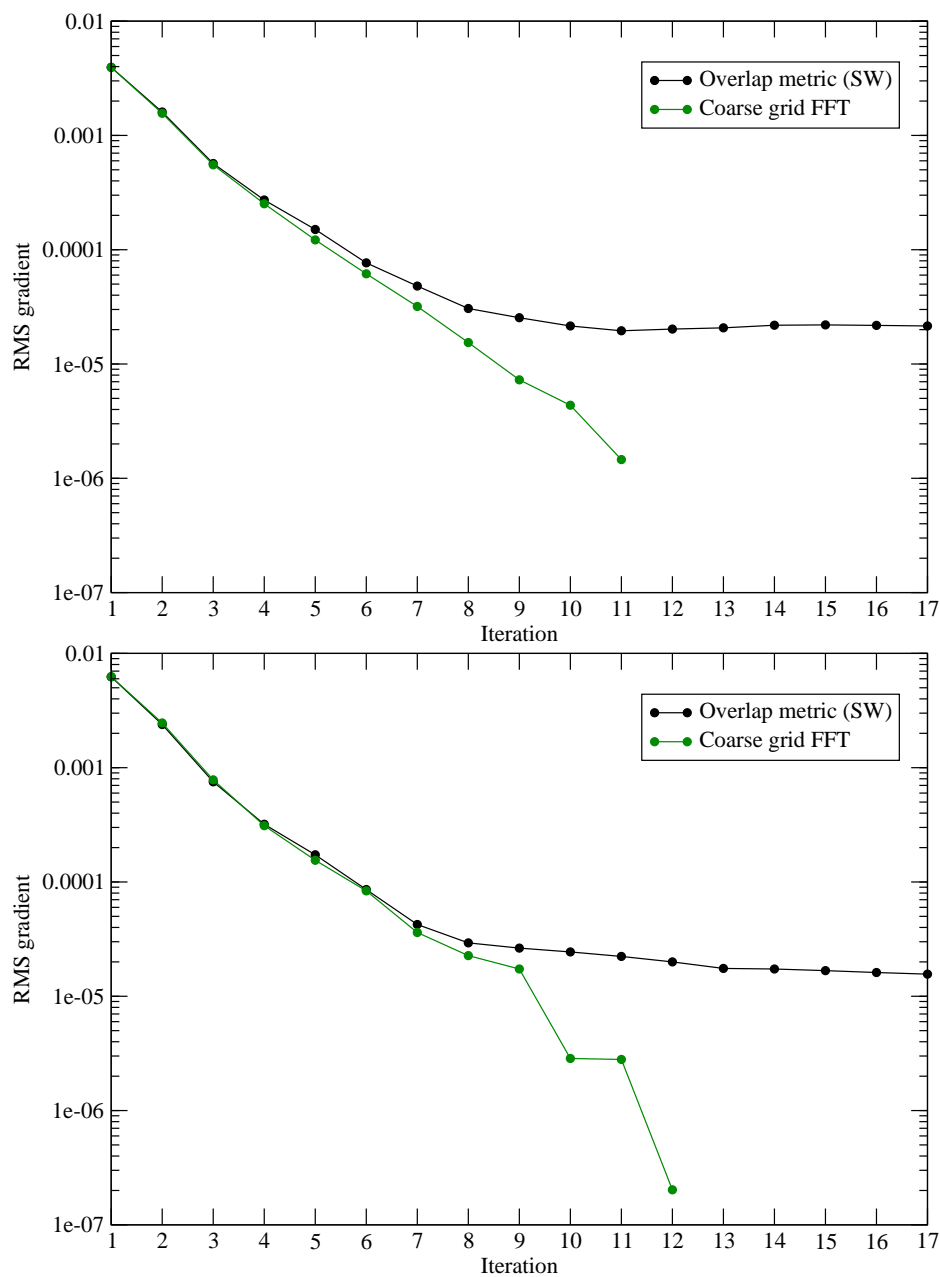
Figure 5.7: NGWF convergence of ONETEP Hartree-Fock single-point energy calculations for ethene (top) and water (bottom) using the spherical wave overlap metric and the coarse grid FFT methods.
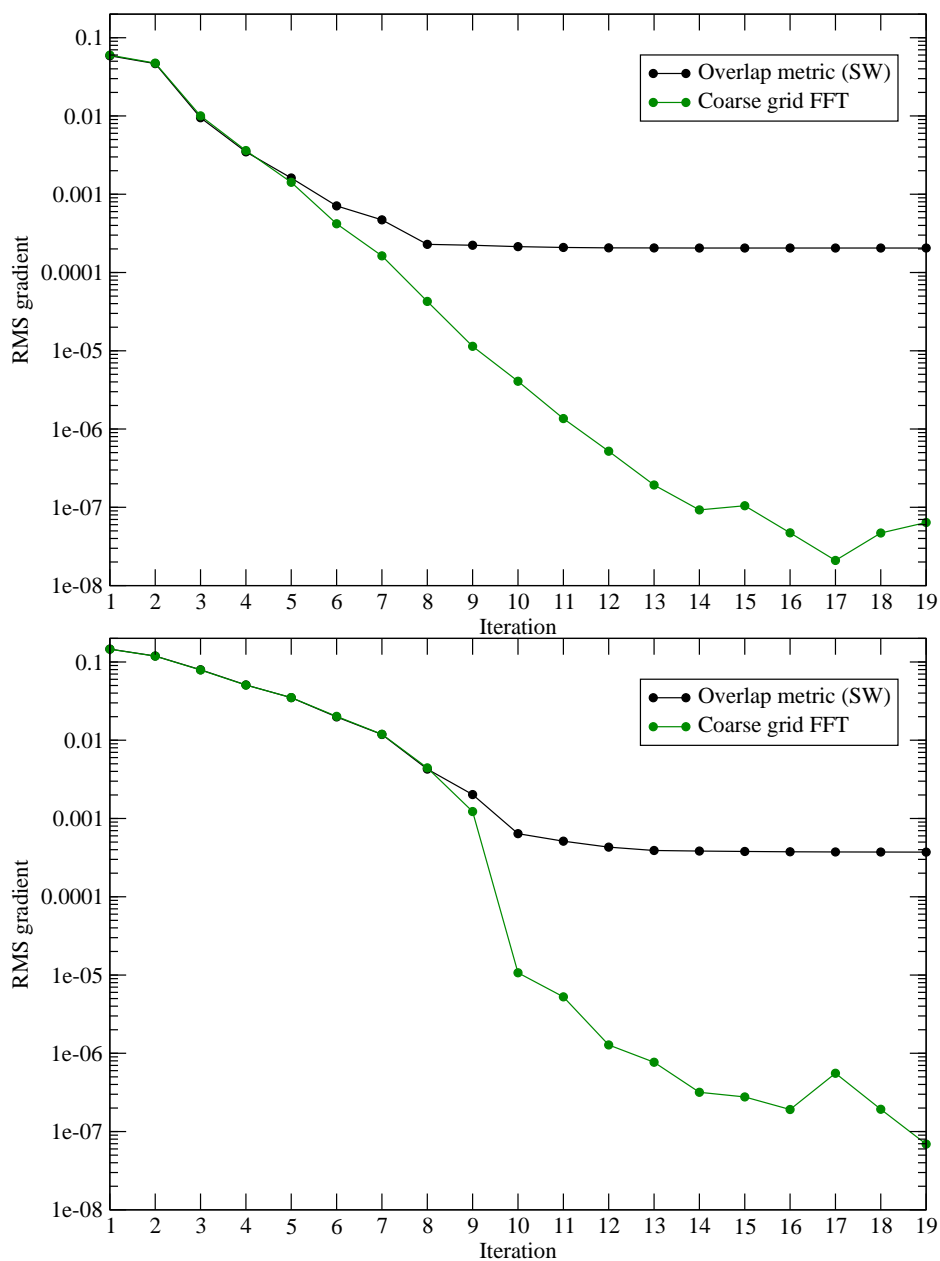
Figure 5.8: LNV convergence of ONETEP Hartree-Fock single-point energy calculations for ethene (top) and water (bottom) using the spherical wave overlap metric and the coarse grid FFT methods.

5.6.4   Electrostatic metric for spherical waves

In view of the practical difficulties encountered in the implementation of the overlap metric matrix and its theoretical limitations it was replaced by the electrostatic metric. The electrostatic matrix avoids the need to accurately represent the NGWF product in terms of spherical waves instead the most relevant property, the electrostatic potential, is central to the fitting. Implementing the electrostatic metric is more complicated, particularly the calculation of the metric matrix. The challenges encountered will be discussed in the next section.

5.6.4.1   Choice of basis set parameters

The spherical wave electrostatic metric was unable to reliably converge to the desired RMS NGWF gradient threshold (2×10^{-6}) for all systems. Therefore convergence with different basis set parameters was important to assess. The convergence tests described in section 5.6.1 are presented in table 5.4 for the electrostatic metric. Table 5.5 shows that, in contrast to the overlap metric (see section 5.6.3.1), the electrostatic metric is variational with respect to the basis set size. A compromise has been made with $l_{\max} = 4$ and 10 Bessel zeros the chosen parameters, the same choice of parameters as for the overlap metric.

5.7   Calculating the electrostatic metric matrix for spherical waves

The form of the electrostatic metric matrix is very important, for the fit to be successful the matrix must be accurate and internally consistent (i.e. elements that should be equal must be equal (to numerical precision)). If the metric matrix is not calculated accurately then convergence suffers. The metric matrix should be symmetric and each atomblock will have a structure whereby the individual elements are either symmetric or antisymmetric to each other. The metric matrix elements depend only on the atomic positions and NGWF radii so only need to be calculated once per single point energy calculation. In this implementation the elements are calculated once at the beginning of a single point energy calculation and stored in a sparse matrix structure. This is to reduce the time spent calculating this matrix; without this storage of the elements, each element would be computed multiple times during the calculation. The atomblocked scheme of this matrix is the same as the exchange matrix but instead of atomblocks of NGWFs there are atomblocks of spherical waves.

The elements in the diagonal atomblocks may be efficiently calculated analytically using

Table 5.4: Convergence (RMS NGWF gradient in $E_h/(a^{3/2}10^{-6})$) with different basis set parameters using the spherical wave electrostatic metric method for: (a) ammonia; (b) benzene; (c) ethene; (d) water.

(a)	l_{max}			(b)	l_{max}		
zeros	2	3	4	zeros	2	3	4
5	16.72	25.78	44.66	5	121.30	83.75	52.38
6	17.47	1.98	18.78	6	124.08	82.15	43.91
7	19.12	11.08	4.10	7	125.65	82.58	40.47
8	19.95	8.63	1.87	8	126.43	82.91	38.33
9	20.19	9.51	5.08	9	126.78	83.05	39.05
10	20.29	15.01	1.71	10	126.96	83.14	38.84
11	20.35	14.48	4.11	11	127.05	83.20	38.73
12	20.39	14.31	4.96	12	127.10	83.25	38.69
13	20.41	14.28	4.67	13	127.14	83.27	38.63
14	20.43	14.72	4.06	14	127.16	83.07	38.43
(c)	l_{max}			(d)	l_{max}		
zeros	2	3	4	zeros	2	3	4
5	35.80	29.16	20.16	5	20.95	47.57	67.64
6	41.49	9.74	17.40	6	14.56	8.30	35.82
7	44.74	11.79	26.05	7	14.77	3.52	4.62
8	46.27	15.05	21.26	8	15.18	4.64	2.76
9	46.96	14.69	19.65	9	15.22	4.67	3.07
10	47.28	14.89	18.83	10	15.38	7.00	4.19
11	47.44	15.01	18.29	11	15.49	7.43	2.04
12	47.52	15.07	18.31	12	15.54	7.74	3.47
13	47.57	15.08	18.03	13	15.60	7.57	3.51
14	47.60	15.36	17.79	14	15.58	7.52	3.46

Table 5.5: Total energy in E_h with different basis set parameters for ONETEP Hartree-Fock calculations using the spherical wave electrostatic metric method for: (a) ammonia; (b) benzene; (c) ethene; (d) water.

(c)	l_{max}			(a)	l_{max}		
zeros	2	3	4	zeros	2	3	4
5	−11.353	−11.370	−11.376	5	−36.599	−36.673	−36.715
6	−11.402	−11.416	−11.421	6	−36.685	−36.749	−36.784
7	−11.426	−11.439	−11.443	7	−36.718	−36.779	−36.811
8	−11.438	−11.451	−11.455	8	−36.734	−36.794	−36.825
9	−11.445	−11.457	−11.461	9	−36.743	−36.803	−36.833
10	−11.449	−11.461	−11.465	10	−36.749	−36.809	−36.839
11	−11.451	−11.464	−11.467	11	−36.752	−36.812	−36.842
12	−11.453	−11.465	−11.469	12	−36.753	−36.814	−36.843
13	−11.454	−11.466	−11.470	13	−36.754	−36.815	−36.844
14	−11.454	−11.467	−11.471	14	−36.755	−36.815	−36.845
(d)	l_{max}			(b)	l_{max}		
zeros	2	3	4	zeros	2	3	4
5	−13.311	−13.336	−13.349	5	−16.722	−16.734	−16.739
6	−13.348	−13.367	−13.376	6	−16.802	−16.813	−16.817
7	−13.362	−13.379	−13.387	7	−16.847	−16.857	−16.860
8	−13.368	−13.385	−13.393	8	−16.872	−16.881	−16.884
9	−13.372	−13.389	−13.396	9	−16.887	−16.896	−16.899
10	−13.374	−13.391	−13.398	10	−16.895	−16.905	−16.908
11	−13.375	−13.392	−13.399	11	−16.901	−16.911	−16.914
12	−13.375	−13.392	−13.399	12	−16.904	−16.915	−16.917
13	−13.376	−13.393	−13.400	13	−16.907	−16.917	−16.920
14	−13.376	−13.393	−13.400	14	−16.907	−16.917	−16.920
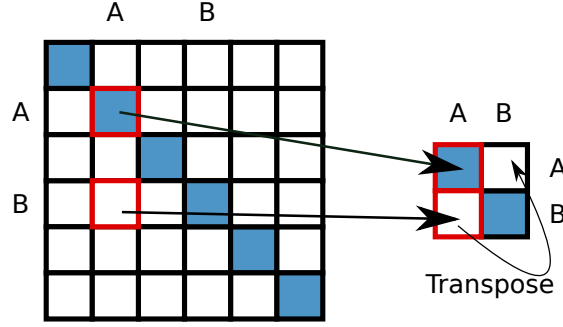
Figure 5.9: Diagram of the procedure for calculating the metric matrix. The blue blocks can be calculated analytically, the other blocks numerically. The blocks are precalculated at the beginning of the calculation and then retrieved and assembled to create the necessary metric matrix. This approach avoids communication since the blocks in the column of atom B are not in general held on the same core as those in the column of A.

the formula,

$$X_{ij} = \int Z_{l_i m_i}(\hat{\mathbf{r}}_1) j_{l_i}(q_i r_1) V_X^{q_j n_j l_j}(\mathbf{r}_1) d\mathbf{r}_1 \tag{5.7.1}$$

$$= \frac{4\pi}{2l_j + 1} \int\int Z_{l_i m_i}(\theta_1, \phi_1) Z_{l_j m_j}(\theta_1, \phi_1) \sin\theta d\theta d\phi$$
$$\int_0^a j_{l_i}(q_i r) \left(r_1^{-l_j - 1} \int_0^{r_1} r_2^{2+l_j} j_{l_j}(q_j r_2) dr_2 + r_1^{l_j} \int_{r_1}^a r_2^{1-l_j} j_{l_j}(q_j r_2) dr_2 \right) dr_1 \tag{5.7.2}$$

using the orthogonality relation in equation 5.4.3 gives,

$$= \frac{4\pi}{2l_i + 1} \delta_{l_i l_j} \delta_{m_i m_j}$$
$$\int_0^a j_{l_i}(q_i r) \left(r_1^{-l_j - 1} \int_0^{r_1} r_2^{2+l} j_{l_j}(q_j r_2) dr_2 + r_1^{l_j} \int_{r_1}^a r_2^{1-l_j} j_{l_j}(q_j r_2) dr_2 \right) dr_1. \tag{5.7.3}$$

These integrals can be evaluated analytically on a case by case basis.

Figure 5.9 shows the procedure to construct the metric matrix, $\mathbf{V}$, in equation 5.3.3. The AA and BB atomblocks are calculated analytically, the precalculated BA atomblock is retrieved and then the transpose is taken to obtain the AB atomblock. This scheme prevents the need for communication at this point in the calculation, since in general the AA and BA, and BB and AB atomblocks are on different cores. If all the atomblocks are local to the core then the whole metric matrix can be copied from the precalculated values. The calculation of the off diagonal elements cannot be done analytically so a numerical integration scheme is required. With numerical integration there is a trade off between speed and accuracy. Several possible integration schemes were investigated and these are described below.

5.7.1 Reciprocal space grid

Using equation 5.4.26 the spherical waves can be generated in reciprocal space on a Cartesian grid. In this implementation this is done in a tightbox since generation of the spherical

Table 5.6: Time to calculate the metric matrix for various levels of grid fineness.

	Time /s	
Grid fineness	Ethene	Water
1	70	17
2	541	134
3	1955	485
4	4361	1076
6	15889	3869

waves in an FFT box is too costly. The spherical waves generated by this method are delocalised over the tightbox. This smearing of the spherical wave is beneficial because overall a more accurate representation is obtained. It was found that enforcing the strict localisation on the generated spherical waves resulted in inferior matrix elements. In the current implementation a batch of spherical waves are generated and then multiplied by the potential due to a spherical wave which is calculated on the points of the tightbox. This process is repeated until all the spherical waves have been included.

5.7.2 Cartesian real space grid

The spherical waves can also be generated on a Cartesian real space grid. The resulting spherical waves are then strictly localised in real space. This method is cheaper than calculating the spherical waves in reciprocal space since there is no need to perform an expensive Fourier transform. In real space it is easy to use an adjustably finer grid to calculate the metric matrix. In theory a grid of arbitrary fineness could be used. This helps to some extent with the poor description of the spherical waves in real space compared to that in reciprocal space. In testing it was found that the improvement obtained by using finer grids was not sufficient to justify the vastly increased computational costs.

5.7.2.1 Convergence and finer grids

The graphs in figures 5.10 and 5.11 show that the use of finer grids does not significantly improve convergence. The convergence behaviour was very similar with all levels of grid fineness for ethene and water. Table 5.6 shows the significant cost that the use of finer grids imposes in the calculation of the metric matrix. It is clear that the use of finer Cartesian real space grids does not deliver the required improvement in convergence.

5.7.3 Spherical real space grid

Using a spherical real space grid independent of the Cartesian grid normally used by ONETEP was the next attempt to improve the matrix elements obtained. In theory, since the parameters of the grid can be adjusted independently from the rest of the calculation, the matrix elements are able to be obtained with arbitrary precision; in practice, however the finite precision of the computer limits this. A uniform radial grid with an 11-point
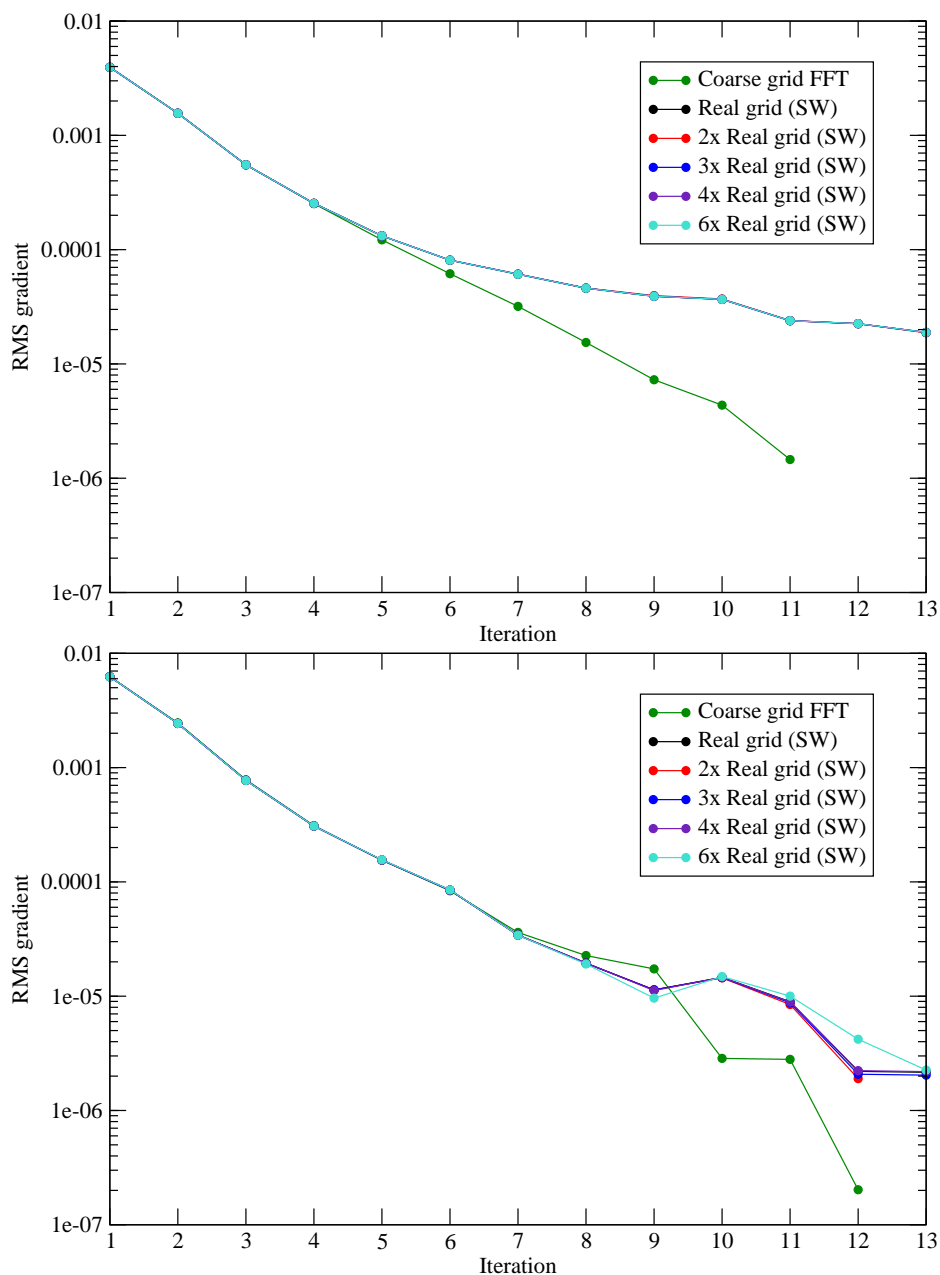
Figure 5.10: NGWF convergence of ONETEP Hartree-Fock single-point energy calculations for ethene (top) and water (bottom) using Cartesian real space grids of varying fineness to calculate the spherical wave electrostatic metric matrix. Convergence for the coarse grid FFT method is included for comparison.
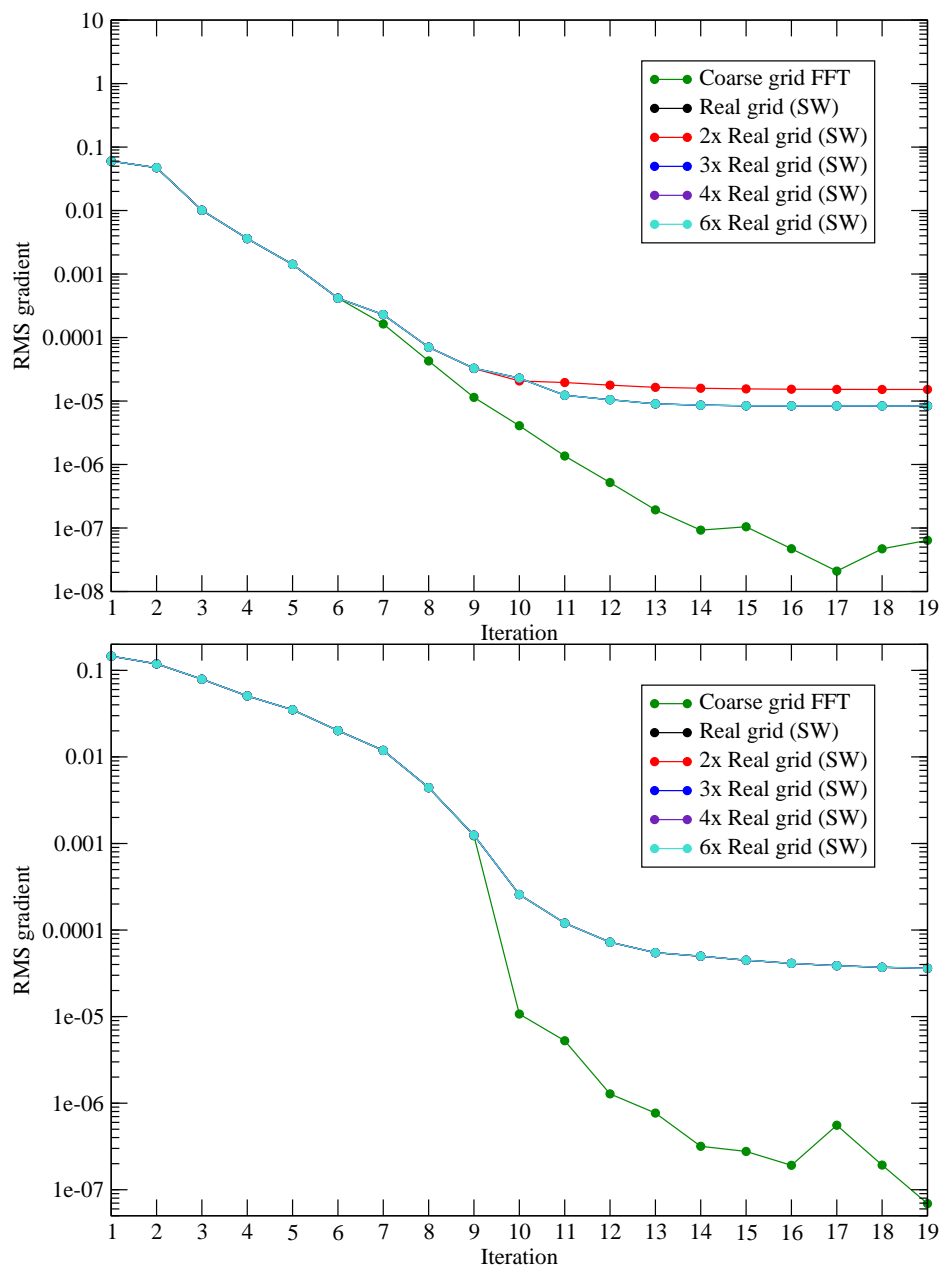
Figure 5.11: LNV convergence of ONETEP Hartree-Fock single-point energy calculations for ethene (top) and water (bottom) using Cartesian real space grids of varying fineness to calculate the spherical wave electrostatic metric matrix. Convergence for the coarse grid FFT method is included for comparison.

Newton-Cotes formula [172] was used to evaluate the radial integral,

$$
\begin{aligned}
\int_{x_0}^{x_{10}} f(x)dx =& \frac{5h}{299736} \left[16076(f(x_0) + f(x_{10})) + 106300(f(x_1) + f(x_9)) \right. \\
&+ 48525(f(x_2) + f(x_8)) + 272400(f(x_3) + f(x_7)) \\
&\left. + 206550(f(x_4) + f(x_6)) + 427368f(x_5) \right] - \frac{1346350}{326918592} f^{(12)}(\zeta)h^{13}
\end{aligned} \tag{5.7.4}
$$

where $h = x_1 - x_0$ and $x_0 < \zeta < x_1$. The 11-point formula was chosen as this is convenient for integrating with the number of radial points divisible by 10. Such a method was chosen because of the oscillatory nature of the spherical Bessel functions. For the angular integration, the Euler-Maclaurin formula [172] has been used,

$$
\begin{aligned}
\int_{x_0}^{x_n} f(x)dx =& h \left[\frac{f(x_0)}{2} + f(x_1) + f(x_2) + \cdots + f(x_{n-1}) + \frac{f(x_n)}{2} \right] \\
&- \frac{B_2 h^2}{2!}(f'(x_n) - f'(x_0)) - \cdots - \frac{B_{2k}h^{2k}}{(2k)!}(f^{(2k-1)}(x_n) - f^{(2k-1)}(x_0)) \\
&+ R_{2k}
\end{aligned} \tag{5.7.5}
$$

where R_{2k} is a remainder term and B_{2k} is a Bernoulli number. If the value of the low derivatives of the function at the endpoints of the integration can be made equal to each other (or zero), then the convergence should be rapid [171]. The angular integrals were done on a uniform grid for the ϕ integral and non-uniform quadrature grid for the θ integral [171]. The non-uniform grid involves a transformation from $\theta \in [0, \pi]$ to $q_\theta \in [0, \pi]$ by,

$$
\theta = q_\theta^2 \frac{3\pi - 2q_\theta}{\pi^2}. \tag{5.7.6}
$$

Only the θ integral needs such treatment because in general for the ϕ integral the values and derivatives at the end points should be the same. Similar transformations exist for their radial integration but this is more suited to integrals involving Gaussians which don't oscillate and are not strictly localised [171].

5.7.3.1 Parameters for the grid

The number of points to use for the integrals in the metric matrix needed to be optimised to ensure a balance between accuracy and computational effort. Twice the number of points were used for the ϕ integral as the θ integral, the number of points used for the radial integral can be adjusted similarly. The use of large numbers of points should lead to a more accurate metric matrix and therefore better convergence in the calculation. However the computational demands of using large number of points are high so a compromise is needed. To determine how many points to use, calculations were performed for an array of different numbers of angular and radial points with ammonia, benzene, ethene and water. The convergence of these calculations was measured to assess whether the number of points used was sufficient. Table 5.7 shows the convergence with differing numbers of

radial grid points. The tests show the sensitivity of convergence to the accuracy of the metric matrix. One reason that convergence is easier for the smaller systems is that the same-centre contributions are more significant in a small system like water, than in a larger system like benzene. The same-centre contributions use metric matrix elements that have been calculated analytically. Further increasing the number of radial and angular points above 100 did not result in a significant improvement in convergence. From the tests the chosen parameters were 50 radial points and 60 angular points.

5.7.4　Convergence behaviour

Figure 5.12 compares the NGWF convergence behaviour obtained using each of the three grids for ethene and water calculations. The convergence behaviour was very similar for the three methods of calculating the metric matrix. Also the converged energies were very similar differing between the three grids by less then an millihartree for both ethene and water.

Figure 5.13 shows the LNV convergence behaviour when the three different grids are used to calculate the metric matrix. For ethene and water the convergence pattern is the same for the three methods. There were some differences for benzene from the thirteenth iteration onwards. The Cartesian real space grid performed best, the spherical real space grid method performed worst. The convergence of the reciprocal space grid method after a jump at iteration 13, stalls. Attempts to improve convergence with the Cartesian real and reciprocal grid methods by imposing the correct symmetry structure within each atomblock failed. Doing this prevented LNV convergence for ethene as the calculated density kernel gradient is incorrect (-3.0 instead of -0.51 on the first LNV iteration with the real grid).

The timings for the three grids are given in table 5.8. The spherical real space grid is by far the most computationally expensive method. It does not provide any clear benefit to convergence and so is not a suitable choice. The Cartesian real and reciprocal space grid methods behave similarly in convergence and take a similar amount of time to calculate the metric matrix. Since the reciprocal space grid is also used to calculate spherical waves in the overlap metric method (and elsewhere in ONETEP), it has be chosen over the Cartesian real space grid method.

5.8　Parallel algorithms

The parallel implementation with spherical waves is much the same as that described in section 4.6 for the FFT based version. Some information about the spherical waves needs to be communicated at the start of each iteration, but the overhead of this is negligible. Batches are used extensively in the calculation to operate on multiple NGWF products, potentials or spherical waves at once thus speeding up the calculation.

The option exists to change the method so that NGWF product is calculated on the core holding atom B and then the expansion coefficients are communicated to the core holding atom A. This would reduce the amount of communication required since the vector of coefficients is smaller than an array holding an NGWF. This has not yet been

Table 5.7: Convergence (RMS NGWF gradient in $E_h/(a^{3/2}10^{-6})$) with different spherical grid parameters for the calculation of the spherical wave electrostatic metric for: (a) ammonia; (b) benzene; (c) ethene; (d) water.

(a)		Radial points						
		30	40	50	60	70	80	90
Angular points	30	4.99	4.30	9.17	4.58	4.60	4.58	4.58
	40	5.13	4.42	3.99	4.60	4.46	4.59	4.58
	50	5.04	4.85	4.96	4.61	4.53	4.59	4.59
	60	4.96	0.25	4.72	4.63	4.51	4.58	4.58
	70	4.89	11.58	4.15	4.62	4.53	4.60	4.60
	80	4.87	6.24	3.28	4.62	4.53	4.59	4.59
	90	4.85	7.49	4.96	4.66	4.53	4.59	4.59

(b)		Radial points						
		30	40	50	60	70	80	90
Angular points	30	41.90	39.06	38.64	38.73	38.69	38.70	38.75
	40	42.12	40.35	38.70	38.62	38.69	38.69	38.69
	50	41.61	38.94	38.69	38.66	38.67	38.68	38.67
	60	41.67	38.81	38.73	38.64	38.65	38.65	38.65
	70	41.70	38.79	38.66	38.65	38.66	38.66	38.66
	80	41.72	39.00	38.70	38.65	38.67	38.67	38.67
	90	41.72	39.48	38.71	38.65	38.67	38.67	38.67

(c)		Radial points						
		30	40	50	60	70	80	90
Angular points	30	14.12	19.33	19.94	19.88	19.91	19.88	19.88
	40	15.91	19.27	24.10	19.87	19.80	19.82	19.82
	50	15.82	17.97	20.10	19.89	19.87	19.87	19.87
	60	15.78	19.42	19.94	19.88	19.86	19.86	19.86
	70	15.77	18.08	19.39	19.89	19.89	19.89	19.89
	80	15.76	17.94	19.94	19.89	19.88	19.88	19.88
	90	15.76	17.86	19.88	19.89	19.87	19.87	19.87

(d)		Radial points						
		30	40	50	60	70	80	90
Angular points	30	6.00	2.12	1.75	1.25	1.27	1.00	1.27
	40	6.01	9.65	3.02	1.22	9.01	1.31	1.31
	50	6.11	9.73	3.13	1.20	9.51	1.31	1.31
	60	6.21	10.11	1.91	1.18	1.99	1.31	1.31
	70	6.21	4.24	9.82	1.20	5.77	1.31	1.30
	80	6.21	4.24	0.81	1.20	5.22	1.31	1.31
	90	6.20	1.97	1.19	1.20	0.83	1.31	1.31

Figure 5.12: NGWF convergence of ONETEP Hartree-Fock single-point energy calculations for ethene (top) and water (bottom) using three different methods to calculate the spherical wave electrostatic metric matrix. Convergence for the coarse grid FFT method is included for comparison.

Table 5.8: Time to calculate the metric matrix using the Cartesian real, reciprocal and spherical real grids on a single core.

	Time /s		
Grid	Ethene	Water	Benzene
Real	130	26	554
Reciprocal	133	27	538
Spherical	637	127	2763

Figure 5.13: LNV convergence of ONETEP Hartree-Fock single-point energy calculations for ethene (top), water (middle) and benzene (bottom) using three different methods to calculate the spherical wave electrostatic metric matrix. Convergence for the coarse grid FFT method is included for comparison.

implemented.

5.9   Results and discussion

5.9.1   Accuracy and convergence

As discussed in the previous chapter it is possible to calculate the diagonal atomblocks of the Hartree-Fock exchange matrix using FFTs with linear scaling cost. Figures 5.14 and 5.15 show the effect this method has on convergence. It can be seen that convergence is generally improved with this method. The NGWF convergence for water however was slightly worse.

  Table 5.9 shows that the energies obtained with the methods agree reasonably well with

Table 5.9: Comparison of energies in E_h from Hartree-Fock exchange calculations using ONETEP for a selection of structures (see figure 4.7) using the methods described in this chapter. Coarse grid FFT Hartree-Fock and PBE energies are provided for comparison. EM indicates the electrostatic metric; NPA indicates the numerical pointwise approach; OM represents the overlap metric; SW stands for spherical wave.

| | Total energy / E_h | | | | | |
	FFT	NPA	SW OM	SW EM	SW EM with FFT	PBE
ammonia	-11.471	-11.459	-11.469	-11.465	-11.472	-11.740
benzene	-36.880	-36.814	-36.871	-36.839	-36.880	-37.987
ethene	-13.400	-13.376	-13.400	-13.398	-13.409	-13.794
H_2O-H_2CO	-38.802	-38.730	-38.753	-38.770	-38.800	-39.590
HCN	-15.843	-15.816	-15.832	-15.831	-15.841	-16.233
Si_{16}	-60.732	-58.239	-61.058	-61.077	-60.732	-63.108
water	-16.921	-16.890	-16.898	-16.898	-16.921	-17.214

each other and those obtained from the coarse grid FFT method. NGWF radii are all $7.0a_0$, except for in the Si_{16} unit cell where $5.1a_0$ was used. This is required due to the small unit cell (the cell has a length of $10.3a_0$ in two dimensions) and the requirement that the NGWFs do not overlap with periodic images of themselves. Some of the discrepancy is caused by the use in the coarse grid FFT method of a different sized FFT box (because of the requirement for it to coincide with the simulation cell). The numerical pointwise approach had problems with Si_{16} convergence. This system is different to the others because a small NGWF radius is required. Using the FFT method to calculate the diagonal elements of the Hartree-Fock exchange matrix leads to energies that are closer to the coarse grid FFT method energies compared to the other methods.

  The LNV convergence of the overlap metric was inferior (figure 5.8) to that obtained when the electrostatic metric is used (figure 5.13). However the NGWF convergence of the two methods was similar (figures 5.7 and 5.12).

5.9.2   Timings

Table 5.10 shows the time taken for a single NGWF iteration for the various methods in this chapter. It can be seen that in general the FFT method was the fastest for this set of
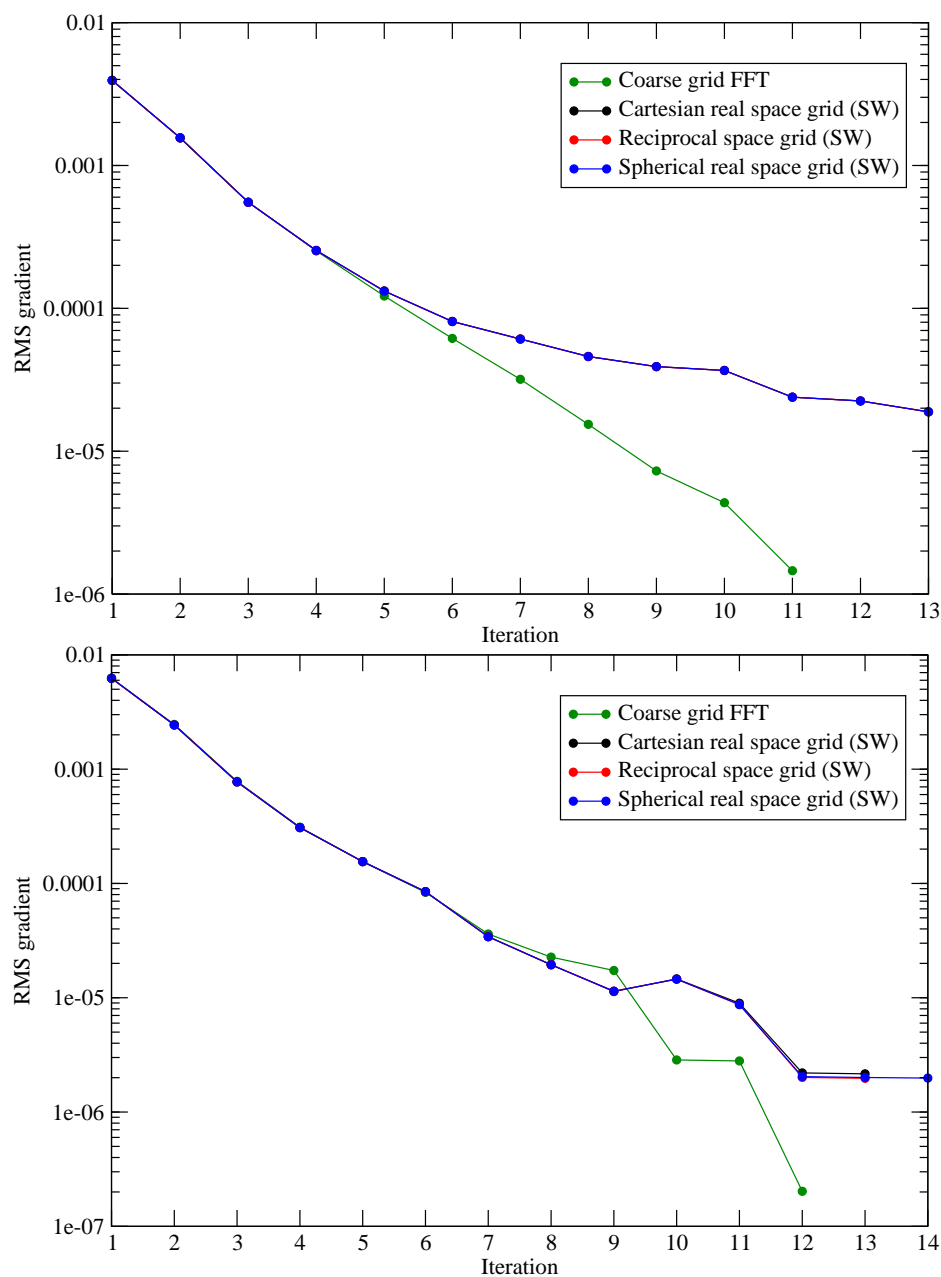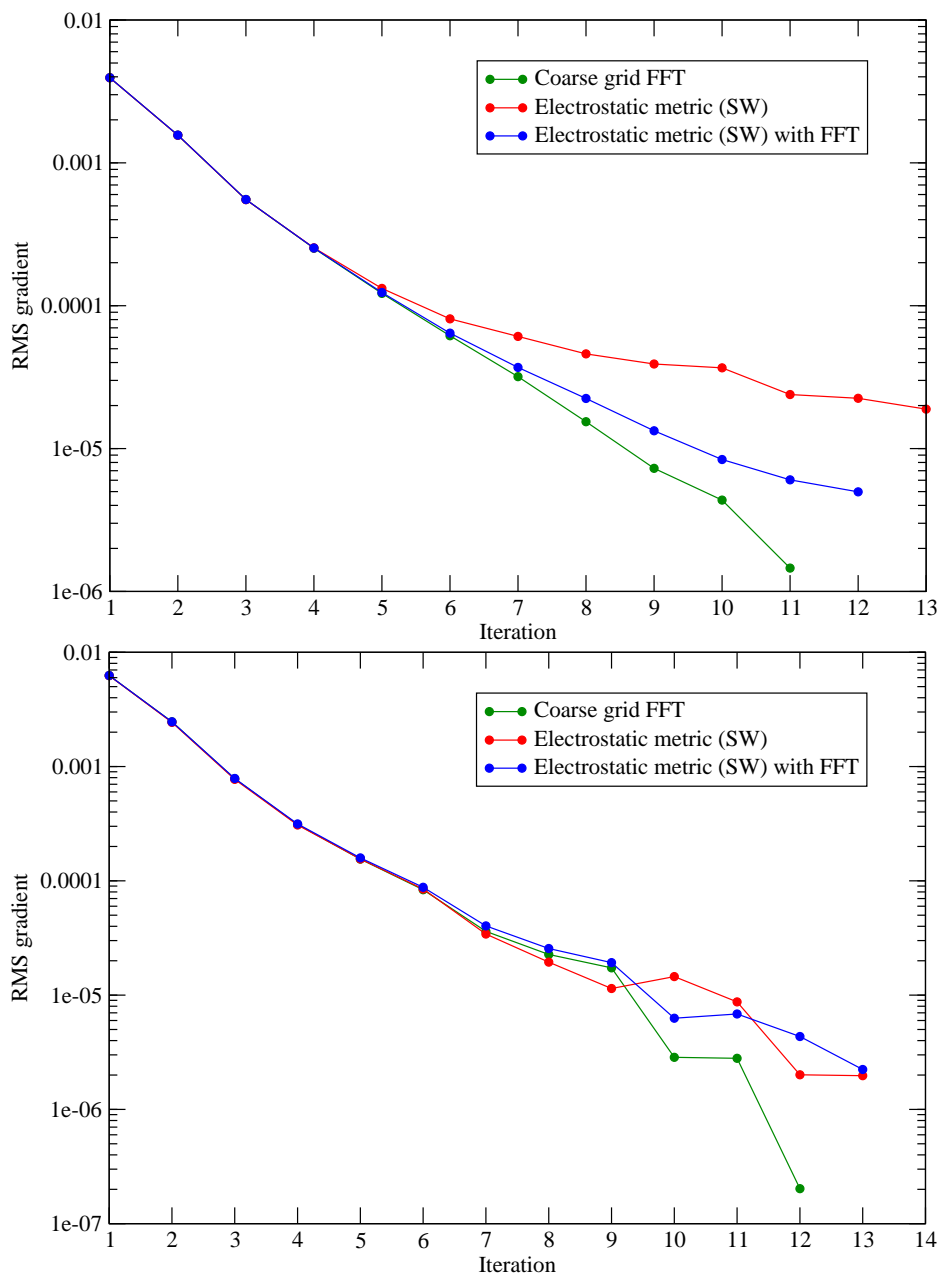
Figure 5.14: NGWF convergence of ONETEP Hartree-Fock single-point energy calculations for ethene (top) and water (bottom) using the reciprocal grid to calculate the spherical wave electrostatic metric matrix and optionally using the FFT method to calculate the diagonal atomblocks of the exchange matrix. Convergence for the coarse grid FFT method is included for comparison.
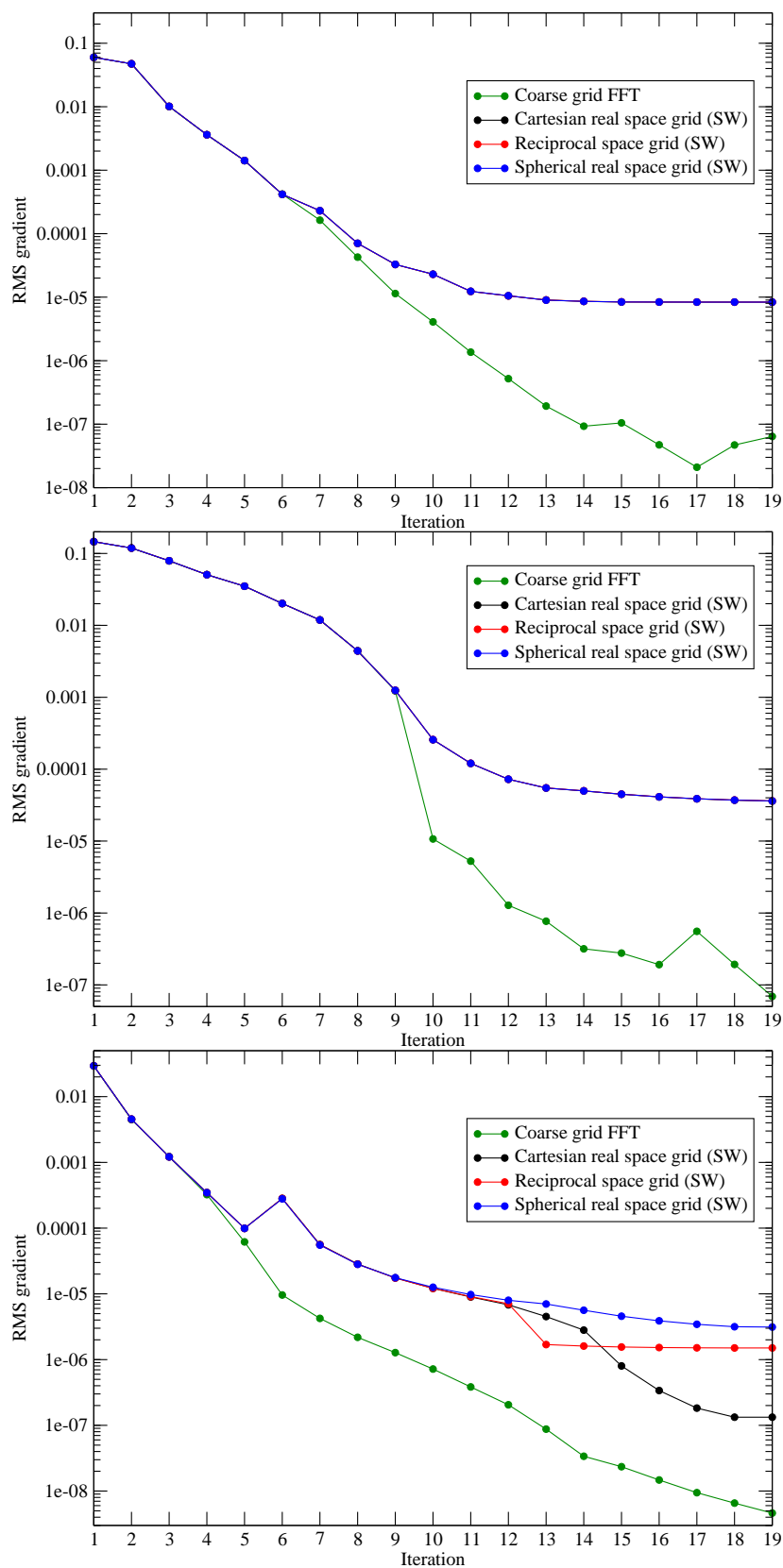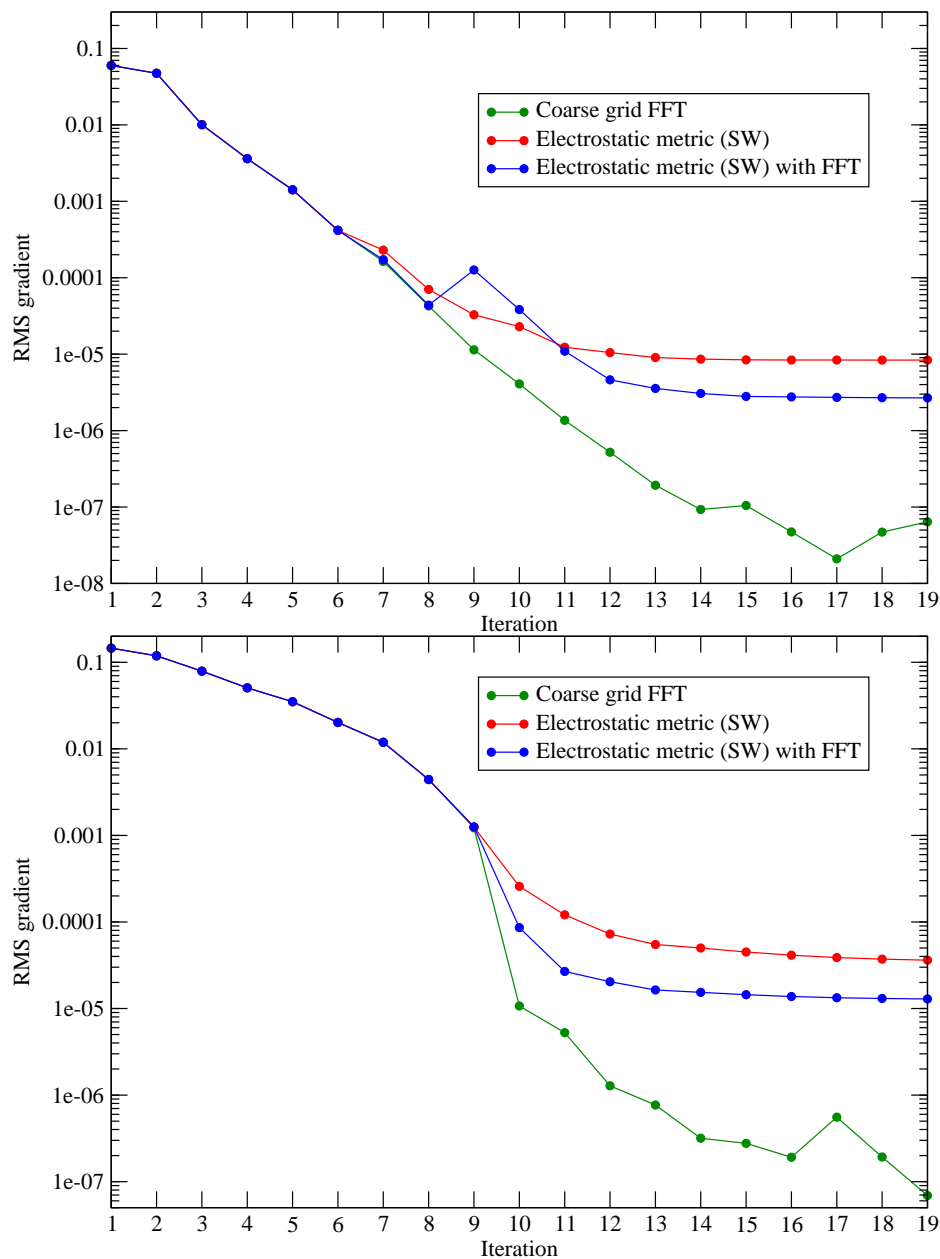
Figure 5.15: LNV convergence of ONETEP Hartree-Fock single-point energy calculations for ethene (top) and water (bottom) using the reciprocal grid to calculate the spherical wave electrostatic metric matrix and optionally using the FFT method to calculate the diagonal atomblocks of the exchange matrix. Convergence for the coarse grid FFT method is included for comparison.
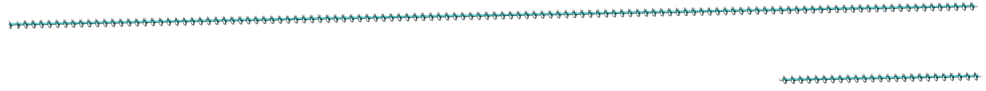
Figure 5.16: The 152 atom (25 ethene units) and 752 atom (125 ethene units) polythene chains.

small systems. The numerical pointwise approach is the slowest of all the methods. The overlap and electrostatic metrics completed the iteration in a similar amount of time with the former method being generally slightly quicker. Calculating the diagonal atomblocks of the Hartree-Fock exchange matrix by the Fourier transform method has a similar cost to the normal spherical wave electrostatic metric method. Due to the improvement in energies and convergence encountered in the previous section this is recommended as the preferred method.

Table 5.10: Comparison of times per iteration for Hartree-Fock exchange calculations using ONETEP for a selection of structures (see figure 4.7) using the methods described in this chapter. Coarse grid FFT timings are provided for comparison. EM indicates the electrostatic metric; NPA indicates the numerical pointwise approach; OM represents the overlap metric; SW stands for spherical wave.

| | Time / s | | | | | |
	FFT	NPA	SW OM	SW EM	SW EM with FFT	Cores
ammonia	102	141	129	141	144	2
benzene	2573	5727	2636	2965	3003	6
ethene	145	1413	583	741	594	2
H_2O-H_2CO	165	328	271	251	319	6
HCN	119	332	154	164	163	2
Si16	268	2325	3549	2783	268	16
water	60	246	109	104	96	2

5.9.3 Linear-scaling tests

To demonstrate linear scaling single LNV iteration calculations were performed for polythene chains of varying length ranging from 152 atoms to 752 atoms. The longest and shortest chains are illustrated in figure 5.16. This system was chosen since it is can be truncated at a variety of lengths and being a linear system allows the early onset of linear-scaling. A density kernel cutoff of $30a_0$ and a kinetic energy cutoff of 700eV has been used. The NGWF radii are all $7.0a_0$. Figure 5.17 demonstrates that the numerical pointwise scales linearly with system size. Figure 5.18 shows the linear-scaling behaviour of the spherical wave method for Hartree-Fock calculations in ONETEP. The spherical wave electrostatic metric method has been used, the metric matrix is generated using the reciprocal grid method. Both methods demonstrate linear scaling as would be expected.
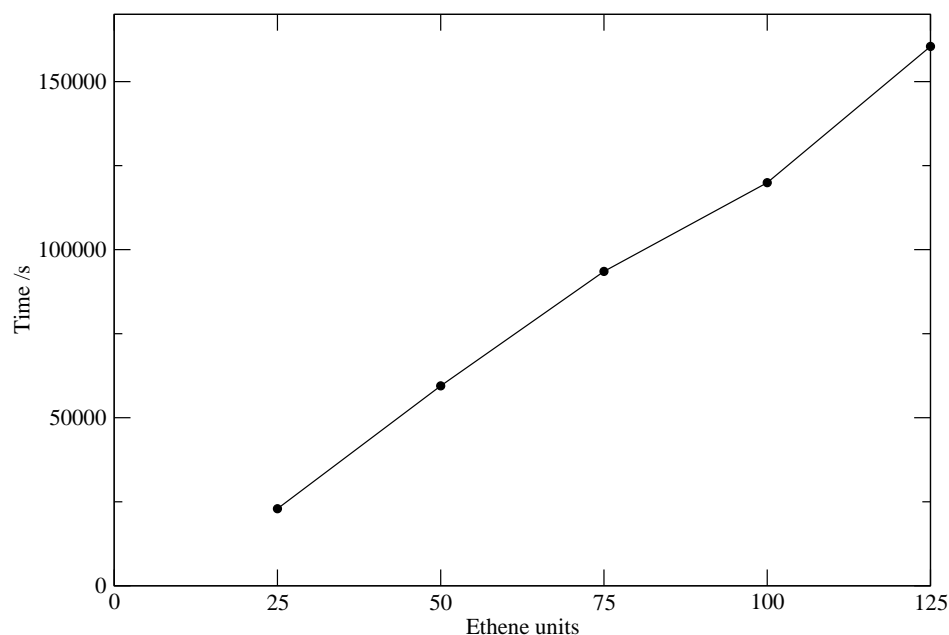
Figure 5.17: Time to perform a single LNV iteration in a ONETEP Hartree-Fock exchange calculation for various linear polythene chains on 48 cores. The numerical pointwise approach has been used.
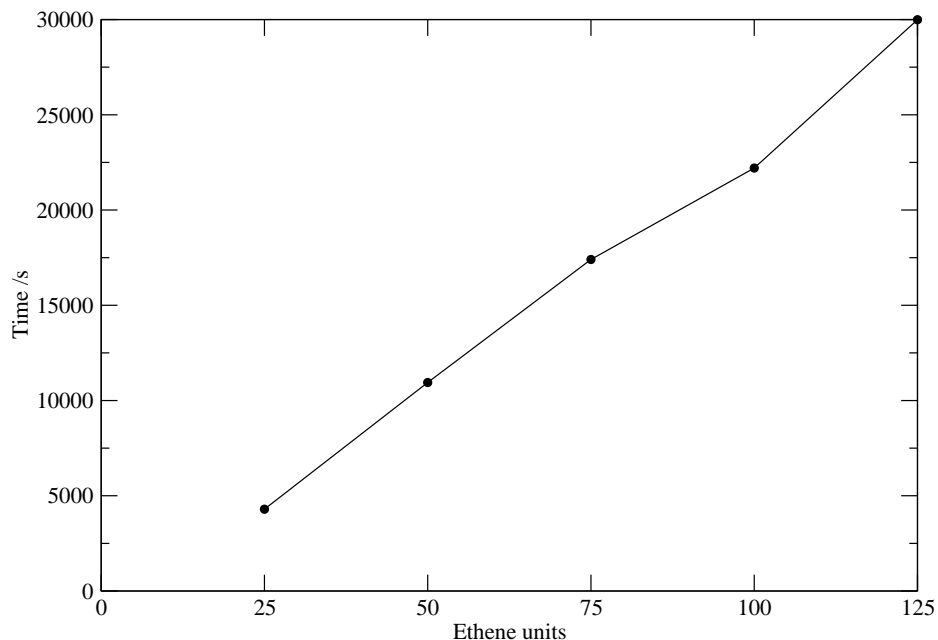
Figure 5.18: Time to perform a single LNV iteration in a ONETEP Hartree-Fock exchange calculation for various linear polythene chains on 48 cores. The spherical wave electrostatic metric method has been used, the metric matrix is generated using the reciprocal grid method (this step is included in the time). An $l_{\max}$ of 2 and a zero limit of 5 has been used.

5.10 Summary

Several methods for the calculation of Hartree-Fock exchange in a linear-scaling manner have been presented. The numerical pointwise approach provided an accurate but slow method to calculate this. Since this method was slow for large systems, methods using auxiliary basis sets to expand the NGWF product were investigated. The use of Gaussians as an auxiliary basis set gave unsatisfactory results. It was also difficult to chose a suitable basis set. As an alternative to Gaussians, spherical waves were introduced. The use of the overlap metric with spherical waves did not give the required improvement in convergence. The electrostatic metric improved convergence further. However there were difficulties in calculating the electrostatic metric matrix. Three different grids for the generation of spherical waves in the calculation of the metric matrix were considered. A Cartesian reciprocal space grid was the chosen grid for the calculation of the metric matrix. Using the Fourier transform method from chapter 4 to calculate the diagonal atomblocks of the exchange matrix lead to an improvement in accuracy and convergence with little extra cost. This is the preferred method for calculating the Hartree-Fock exchange energy using spherical waves. Linear-scaling has been demonstrated for the methods for the methods in this chapter.

Chapter 6

Hybrid functionals

In this chapter hybrid functionals will be considered. In section 6.1 the theoretical background of hybrid functionals will be explored and then the form of several hybrid functionals are introduced. In section 6.2 an organometallic reaction mechanism is investigated using hybrid functionals with NWChem [146]. This is to provide a comparison with the implementation of hybrid functionals in ONETEP in section 6.3.

6.1 Adiabatic connection methods

The Hellman-Feymann theorem can be used to provide an expression for the exchange-correlation energy in terms of a parameter that determines the level of electron-electron interaction in the system (also referred to as the inter-electronic coupling strength), running from zero for the non-interacting system (the Kohn-Sham reference system) to one for the exact interacting system. Using λ as such a parameter gives the adiabatic connection formula,

$$E_{\text{xc}} = \int_0^1 U_{\text{xc}}^\lambda d\lambda \qquad (6.1.1)$$

where U_{xc}^λ is the exchange-correlation potential energy at an intermediate inter-electronic coupling strength. (This can be derived from the usual Hellmann-Feymann theorem by integrating with respect to λ.) Though only the potential energy appears in the integrand of equation 6.1.1, the kinetic energy part of the exchange-correlation energy is obtained by the integration over λ [50].

Since the Slater determinant for the Kohn-Sham orbitals is the exact wavefunction for the non-interacting system, U_{xc}^0 will be the Kohn-Sham exact exchange for the non-interacting system. This can be computed as in Hartree-Fock theory but using the Kohn-Sham orbitals instead of Hartree-Fock orbitals. The first hybrid functional was the half and half functional introduced by Becke [50]. His functional approximates the λ dependence of the integrand in equation 6.1.1 using a linear interpolation. This results in the two point approximation,

$$E_{\text{xc}} \approx \frac{1}{2} U_{\text{xc}}^0 + \frac{1}{2} U_{\text{xc}}^1. \qquad (6.1.2)$$

Becke made the further approximation that U_{xc}^1 could be estimated by the potential energy

of the local (Spin-)Density Approximation (L(S)DA). He argued that L(S)DA is more accurate at the $\lambda = 1$ limit than the $\lambda = 0$ limit [50, 173]. The expression for the half and half functional is therefore,

$$E_{xc}^{HH} = \frac{1}{2}E_X^{HF} + \frac{1}{2}U_{xc}^{L(S)DA}. \tag{6.1.3}$$

(Note that the distinction between the potential and total energy of L(S)DA exchange correlation has been ignored by some authors [173,174]). In contrast to later hybrid functionals both the exchange and correlation of the L(S)DA functional are scaled.

6.1.1   Hybrid functionals

The introduction of the half and half functional by Becke led to the development of further hybrid functionals. Becke developed the 3-parameter functional expression by adding gradient corrections to the exchange and correlation energies and relaxing the linear λ dependence of the half and half model [51]. The three semi-empirical parameters are fitted to experimental data. The first such functional was B3PW91 [51],

$$E_{xc}^{B3PW91} = E_{xc}^{LSDA} + a_0(E_X^{HF} - E_X^{LSDA}) + a_X\Delta E_X^{Becke88} + a_c\Delta E_c^{PW91} \tag{6.1.4}$$

with $a_0 = 0.20$, $a_X = 0.72$ and $a_c = 0.81$. The extremely popular functional B3LYP was developed by replacing PW91 [45, 46] with LYP (Lee, Yang and Perdew) [42] but keeping the values of the three coefficients the same [52],

$$E_{xc}^{B3LYP} = E_{xc}^{LSDA} - a_0(E_X^{HF} - E_X^{LSDA}) + a_X\Delta E_X^{Becke88} + a_c(E_c^{LYP} - E_c^{LSDA}). \tag{6.1.5}$$

The correlation term changes in form since the LYP functional is a standalone correlation functional whereas PW91 is a correction to the LSDA correlation [175].

The X3LYP functional [72] involves both the Becke88 and PW91 exchange functionals, its formula is,

$$\begin{aligned} E_{xc}^{X3LYP} = {}&(1 - a_0 - a_X)\,E_X^{L(S)DA} - a_0 E_X^{HF} + a_X\left(0.765 E_X^{Becke88} + 0.235 E_X^{PW91}\right) \\ &+ (1 - a_c)E_c^{LYP} - a_c E_c^{VWN}, \end{aligned} \tag{6.1.6}$$

with $a_0 = 0.218$, $a_X = 0.709$ and $a_c = 0.12$ (note the full Becke88 functional has been used). The use of both Becke88 and PW91 in the exchange was expected to provide a more balanced functional. The functional was developed with reference to the form of the exact exchange density in Gaussian based codes. X3LYP was shown to offer greater accuracy for heats of formation, ionisation energies, electron affinities and total atomic energies than B3LYP. In particular the description of hydrogen-bonded and van der Waals complexes is more accurate [72].

Another common form for hybrid functionals incorporates 25% exact exchange and this single parameter is derived using perturbation theory (so there are no truly empirical parameters) [173]. The expectation is since the functional has not had parameters fitted to a particular set systems such a functional will be more generally applicable. Examples of
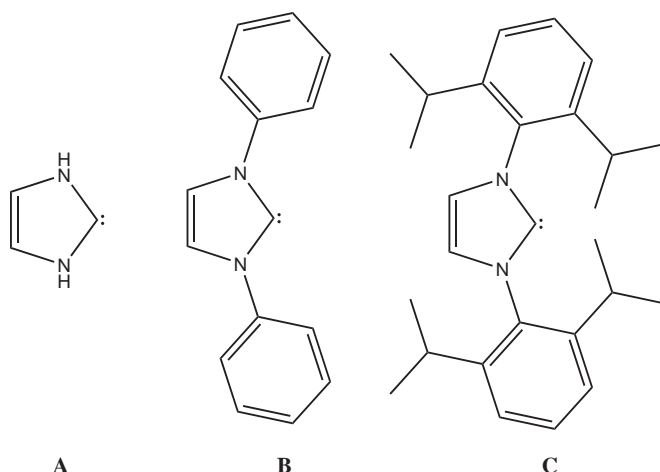
Figure 6.1: The three models used for the ligand.

such functionals are PBE0 [176], B1LYP [174], and B1PW91 [174]; these functionals can be expressed as,

$$E_{\mathrm{xc}} = \frac{1}{4}E_{\mathrm{X}}^{\mathrm{HF}} + \frac{3}{4}E_{\mathrm{X}}^{\mathrm{GGA}} + E_{\mathrm{c}}^{\mathrm{GGA}}. \tag{6.1.7}$$

Other hybrid functionals have been developed with exact exchange incorporated from the start rather then as a modification to an existing functional. Examples of such functionals are B97 [53] and B98 [177] which are hybrid GGAs incorporating 19.43% and 21.98% exact exchange respectively. Both these functionals were obtained by fitting parameters to a large set of complexes.

6.2 Application to an organometallic system

For organometallic systems hybrid functionals are considered to be a better choice than GGA functionals. A reaction involving a nickel complex with N-heterocyclic carbene (NHC) ligands and ethene has been studied. NWChem [146] has been used to study the reaction mechanism.

The substitution in $Ni(NHC)_2$ of one NHC and how this is affected by steric and electronic effects of different carbene ligands was explored. Three models for the ligand A, B and C, have been used as shown in figure 6.1 . In model A all NHC substituents were replaced by hydrogens, in model B the NHC substituents are phenyl rings and model C has di-isopropyl benzene side groups attached to the ligands.

For each of these three ligand models we have investigated the path connecting the reactants to the products via both an associative and a dissociative mechanism. The individual reactions of the mechanisms are shown in figure 6.2, where both mechanisms begin from compound **1**, the reactant. The dissociative mechanism then proceeds by loss of one of the NHC ligands to form the coordinatively unsaturated complex **3**. Then two ethene ligands bind in η^2 mode in two consecutive steps in order to form the product, **6**. In the associative mechanism an ethene is first added to **1** to form **7** and then a NHC ligand is
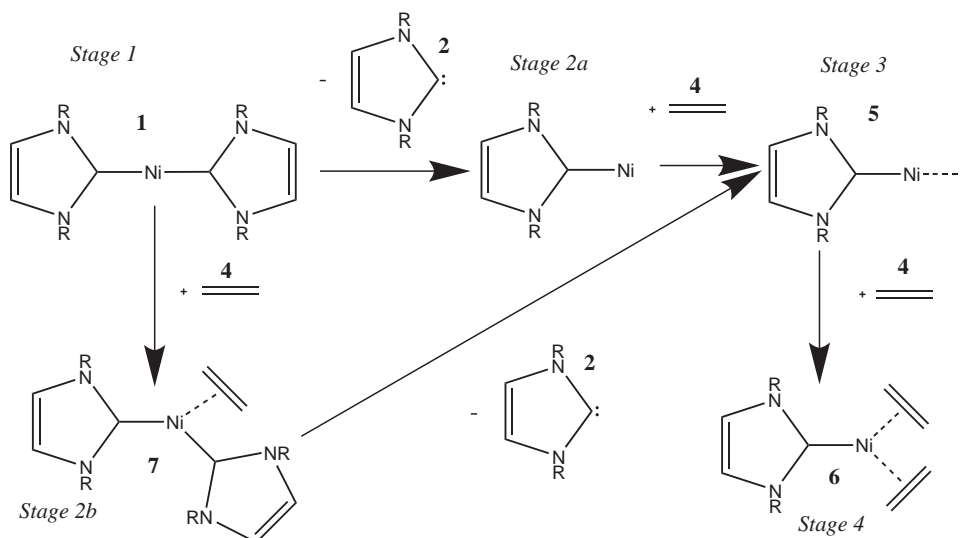
Figure 6.2: Dissociative and associative mechanisms for the substitution of one NHC ligand by ethene.
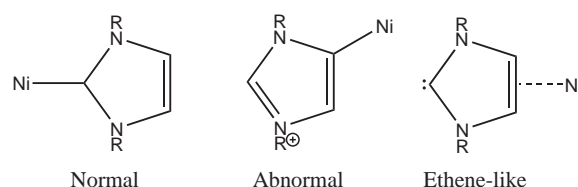
Figure 6.3: Possible ligand bindings to nickel.

dissociated to yield **5**. The final step involves the addition of a second ethene in the same way as in the dissociative mechanism.

The possible binding modes of an NHC ligand are shown in figure 6.3. These involve the "normal" mode which is part of the catalytic mechanism, the "abnormal" mode where the ligand still binds as a carbine and the ethene-like mode (for unsaturated NHCs only) where the ligand binds to the metal in an η^2 fashion.

The possibility of an NHC ligand disrupting the catalytic cycle, by binding to the Ni (in any of the three ways shown in 6.3) instead of ethene was considered. Figure 6.4 shows some of the potential structures which could arise. Structures **8**, **9** and **10** are analogues of structures **5**, **7**, and **6** respectively with the ethene ligands substituted for η^2 bound NHC ligands. In structure **11** three normally bound NHC ligands surround the nickel ion. Structures **12** and **13** both have a single abnormally bound NHC ligand. This abnormal binding could possibly occur directly from structures **3** and **5** (by a reaction with an abnormal ligand) or by a reaction converting structures **8** and **9** to abnormal binding.

6.2.1   Computational methods

Geometry optimisations of the model A structures **1** and **6** were performed using the B3LYP hybrid density functional using a variety of Gaussian basis sets (for carbon, nitrogen, oxygen and hydrogen). This pair of structures was chosen because experimental crystal structures [1] were available for the model C versions of these complexes. This allowed a suitable
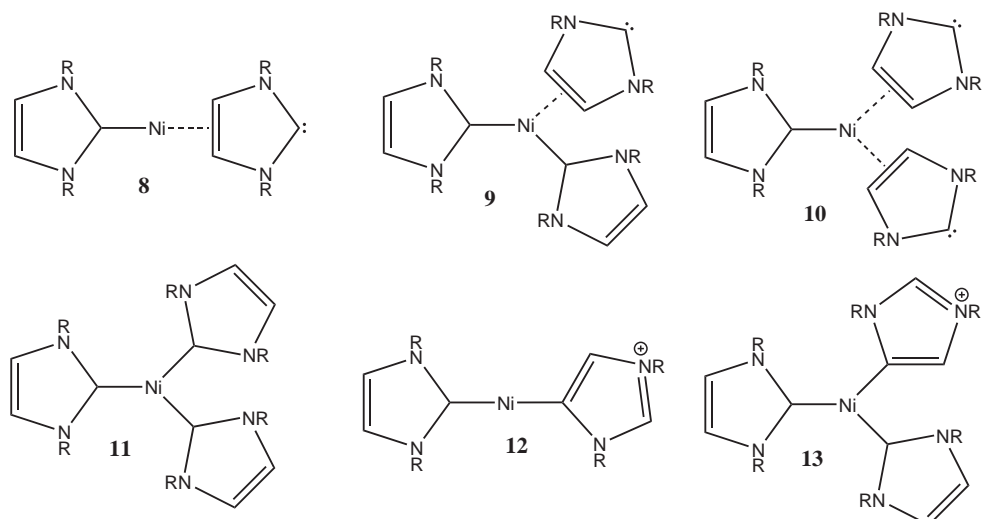
Figure 6.4: Alternative structures which arise if an additional NHC ligand binds to the Ni instead of the ethene.

Table 6.1: Basis set calibration for B3LYP, comparing with Andreas Danopoulos' crystal structures [1] for structure **1**.

Basis Set	Energy $/E_h$	Bond length / Å		Angle	
		Ni - C	C-N	Ni-C-N	Dihedral of ligands
6-31G**	−623.39	1.85	1.37	129.28	−90.01
6-31++G**	−623.41	1.85	1.39	129.32	−89.60
cc-pVDZ	−623.42	1.85	1.37	129.29	−88.93
cc-pVTZ	−623.55	1.85	1.37	129.27	−89.58
aug-cc-pVDZ	−623.45	1.85	1.38	129.33	−89.91
Crystal		1.85	1.38	129.15	−43.31

basis set to be chosen by comparison with the crystal structures. Table 6.1 found gives details of the basis set calibration for structure **1**. Since the ligands in model A are small compared to the model C ligands, there is a large discrepancy in the dihedral angle of the ligands, which will be discussed below. After comparison of bond lengths and angles the 6-31++G** [178, 179] basis set was selected for subsequent calculations. For the Ni atom the Stuttgart 1997 [180] relativistic effective core potential and complementary basis set were used. Geometry optimisations for all seven structures with model A and B were performed with B3LYP. Single point energy calculations have been performed with the COSMO implicit solvent model [181] using the vacuum geometry. The solvent modelled is tetrahydrofuran and a dielectric constant of 7.52 has been used. To confirm the validity of using the vacuum geometries, geometry optimisations with implicit solvent were performed with structures **2**, **5** and **7**. Using the vacuum geometry of each structure the energy difference was 17.5 kcal mol^{-1}, the structures with geometries obtained by optimising with implicit solvent, had an energy difference of 17.4 kcal mol^{-1}. This variation is not significant so the vacuum geometries have been used throughout.

The geometries of the model C structures have been optimised with ONETEP using the

Table 6.2: Relative energies of the stages in the two possible reaction mechanisms with B3LYP (with and without implicit solvent). Mechanism 1 goes through stage 2a; mechanism 2 goes through stage 2b.

	B3LYP ΔE / kcal mol^{-1}			B3LYP Implicit solvent ΔE / kcal mol^{-1}		
Step	Model A	Model B	Model C	Model A	Model B	Model C
1	0.0	0.0	0.0	0.0	0.0	0.0
2a	47.2	46.4	38.2	44.8	43.0	33.1
2b	−18.1	−11.8	41.6	−15.9	−11.6	44.3
3	12.6	9.6	4.9	10.1	7.3	1.9
4	−9.4	−11.0	−16.6	−8.8	−11.0	−17.0

Perdew, Burke and Ernzerhof (PBE) exchange correlation functional [47]. ONETEP was used because the time taken to perform the calculations with NWChem was prohibitive. This is caused because the time required to perform an NWChem calculation scales cubically with the number of atoms in the system; with ONETEP the cost of the calculation increases linearly with system size. PBE was chosen because it is a generally applicable generalised gradient approximation functional and has reasonable computational demands. The energies of structures **1-7** were computed for models A and B using ONETEP with PBE to verify that the results were quantitatively similar to those obtained with NWChem. Results are in table 6.3.

6.2.2   Results and discussion

The results in table 6.2 for models A and B show that the energy for the stage involving structure **7** has a very low energy and there is an energy barrier to be overcome to move from this stage to the expected product. While this shows a strong preference for the associative mechanism over the dissociative mechanism, neither mechanism looks feasible. With model C both stage 2a and stage 2b are higher in energy than stage 1. Also, the associative mechanism is no longer favoured as stage 2a is lower in energy than stage 2b. The bulk of the model C ligands make structure **7** difficult to realise, which is reflected by stage 2b having a higher energy than stage 1. In model C all the stages of the dissociative mechanism have a lower energy (relative to stage 1) than the corresponding stages in models A and B, so this makes this mechanism more feasible. This shows the importance of modelling the complete structure to account for all the steric effects, when comparing reaction mechanisms. For all three models the effect of the implicit solvent on the energy differences is minor. The energy difference between adjacent stages is slightly reduced. The energy of stage 2a is reduced by a larger amount than that of stage 2b. Structures **1** and **7** are the most stabilised, with structures **2**, **3**, **5** and **6** being stabilised by a smaller amount.

The alternative structures, shown in figure 6.4, have also been considered. The formation energy of these structures are shown in table 6.4. Comparing structures **1** and **8**, for models A, B and C structure **1** is favoured by 21.7, 15.9 and 23.5 kcal mol^{-1} respectively. These energies show no clear correlation with the size of the ligand. A similar comparison

Table 6.3: ONETEP PBE results for the mechanisms in figure 6.2.

| | ΔE / kcal mol^{-1} | | |
Step	Model A	Model B	Model C
1	0.0	0.0	0.0
2a	52.7	58.5	51.8
2b	−26.6	−15.9	32.5
3	9.5	14.7	11.9
4	−19.4	−2.0	−15.7

Table 6.4: Formation energies of the alternative structures shown in figure 6.4. (Structures **8** and **12** form from structures **2** and **3**; structures **9**, **11** and **13** form from structures **1** and **2**.) Dis indicated dissociation.

| | B3LYP ΔE / kcal mol^{-1} | | | B3LYP Implicit solvent ΔE / kcal mol^{-1} | | |
Structure	Model A	Model B	Model C	Model A	Model B	Model C
8	−25.5	−30.5	−14.7	−21.7	−29.1	0.0
9	−5.6	−5.2	Dis	0.1	−2.4	Dis
11	−12.5	−5.8	Dis	−5.9	0.3	Dis
12	−27.7	−30.0	−28.5	−32.4	−32.2	0.0
13	8.6	6.6	Dis	9.9	9.5	Dis

of structures **1** and **12** does suggest a pattern, as the ligand size increases structure **1** is less strongly favoured (structure **1** is favoured by 19.5, 16.4 and 9.7 kcal mol^{-1}, for models A, B and C respectively). However this pattern is more a reflection of the stability of structure **1** than that of structure **12** (its formation energy differs little between the three models). The normal reaction of an ethene binding to structure **3** to form structure **5** is more exothermic than the reactions that form structures **8** and **12**, but rebinding to an NHC ligand in the normal way to reform structure **1** lowers the energy more than any of these reactions. Therefore it is unlikely that the dissociative mechanism would be disrupted by these side reactions (from stage 2a). For the associative mechanism structures **9**, **11** and **13** are possible alternatives to structure **7**. When the energies are compared for model A structure **7** is enthalpically favoured and the reaction leading to structure **13** would be endothermic. Geometry optimisations of model C structures **9**, **11** and **13** resulted in one of the NHC ligands dissociating, therefore energies are not available for these structures. This mirrors the disfavouring of the associative mechanism for model C (see table 6.2). These large energy differences suggest that the associative mechanism should not be affected by these side reactions. It was not possible to obtain an optimised structure **10** with any of the three models. For all three models geometry optimisations of structure **10** resulted in one of the η^2 bound NHC ligands dissociating from the nickel. In the case of model A the dissociating ligand starts to rotate so that the carbene was closer to nickel than the C-C double bond.

Table 6.5 and table 6.6 compare the computed optimised structures with the available crystal structures. Most of the parameters are very similar between the three models and

Table 6.5: Comparison of structural parameters for the three models (see figure 6.1) and the crystal structure for structure **1** (see figure 6.2).

	Bond length				Angle		
	Ni - C	C-N	N-Ph	C=C	Ni-C-N	C-Ni-C	Dihedral of ligands
Model A	1.85	1.39	N/A	1.36	129.32	179.99	89.60
Model B	1.86	1.39	1.43	1.35	129.37	170.92	67.36
Model C	1.91	1.39	1.44	1.34	128/130	177.77	42.30
Crystal	1.85	1.38	1.44	1.32	129.15	177.23	43.31

Table 6.6: Comparison of structural parameters for the three models (see figure 6.1) and the crystal structure for structure **6** (see figure 6.2).

	Bond length						Angle
	Ni-C (Et)	Ni-C (L)	C-N	N-Ph	C=C (Et)	C=C (L)	Ni-C-N
Model A	2.01	1.89	1.37	N/A	1.40	1.36	129.1
Model B	2.02	1.90	1.38	1.43	1.40	1.35	128.62
Model C	2.06	1.95	1.38	1.44	1.39	1.34	129.43/127.81
Crystal	1.98	1.91	1.37/8	1.45	1.39	1.35	130.75/127.22

the experimental structures. The dihedral angle of the two ligands in the optimised model A structure **1** (89.6°) does not agree with the angle in the crystal structure (43.3°). This is because the steric hindrances from the large di-isopropyl phenyl groups are not accounted for in model A, the orientation of the ligands is, therefore, much more flexible. In model B the dihedral angle (67.4°) is closer to that observed in the crystal structure and with model C the dihedral angle (42.3°) is very close to the crystal structure dihedral.

6.2.3   Comparing hybrid functionals

As a further test the results of geometry optimisations of the three of the structures with different functionals were compared, this is presented in table 6.7. It was found that there was not a significant difference in the energies between the B3LYP geometries and the functionals native geometries. Therefore only single point energy calculations using the B3LYP geometries have been performed for the other truncated structures. In table 6.8 the results from several hybrid functionals are compared for the reaction mechanism. The results from BLYP and Hartree-Fock calculations are also included. It can be seen that some of the energy differences for Hartree-Fock are very large. The energy differences do vary between the different hybrid functionals.

6.3   Hybrid functionals in ONETEP

Several hybrid functionals have been implemented in ONETEP: B1LYP; B1PW91; B3LYP; B3PW91; PBE0; X3LYP. To facilitate the inclusion of these new functional several changes had to be made to the original implementation. Previously in ONETEP there was a sub-routine that implemented the calculation of equation 2.5.7 (including the calculation of

Table 6.7: Comparison of energy differences for model A of structure **7** and structures **2** and **5** (one of the steps in mechanism 2, see figure 6.2) using B3LYP geometries and (native) geometries obtained with the functional stated. NWChem was used for these calculations.

| | $\Delta E/$ kcal mol^{-1} | |
| | Geometry | |
Functional	Native	B3LYP
Hartree-Fock	−81.74	−90.40
BLYP (GGA)	−26.01	−25.88
PBE (GGA)	−33.16	−33.02
B3LYP (hybrid)	−26.66	−26.66
mPW1K (hybrid)	−32.15	−31.94
B97 (hybrid)	−28.61	−28.63
B98 (hybrid)	−29.67	−29.67

Table 6.8: Comparison of energy differences in kcal mol^{-1} for models A and B for the reaction mechanism with various hybrid functionals, Hartree-Fock and BLYP (a GGA functional). The geometries were obtained using B3LYP and the calculations were performed using NWChem.

Step	B3LYP A	B3LYP B	B97 A	B97 B	B98 A	B98 B	Hartree-Fock A	Hartree-Fock B	BLYP A	BLYP B
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2a	47.2	46.4	46.8	47.2	47.3	48.0	165.1	−13.4	49.3	49.0
2b	−18.1	−11.8	−22.3	−16.2	−23.3	−17.7	−15.0	−95.1	−15.3	−6.9
3	12.6	9.6	10.7	8.4	10.8	8.5	89.0	−79.1	14.4	13.1
4	−9.4	−11.0	−14.7	−15.5	−15.6	−16.4	107.4	−91.5	−4.7	−4.4

Table 6.9: Comparison of energy differences in kcal mol^{-1} for model A for the reaction mechanism with various hybrid functionals and Hartree-Fock. The geometries were obtained using B3LYP with NWChem and the calculations were performed using ONETEP.

	$\Delta E/$ kcal mol^{-1}				
Step	B3LYP	B1PW91	B3PW91	PBE0	HF
1	0.0	0.0	0.0	0.0	0.0
2a	47.1	52.8	51.4	56.6	82.6
2b	683.2	-29.8	-28.3	-33.8	67.7
3	17.2	15.5	16.4	15.2	75.6
4	-13.8	-21.3	-19.9	-25.8	-17.5

f_{xc}) for each functional. Two generic subroutines one for LDA functionals and the other for GGA and hybrid functionals to calculate equation 2.5.8 have now been implemented. These generic subroutines call other subroutines which calculate the value of f_x or f_c given the density and density gradient at the point and also the contribution to the potential from that point. So, for example, for BLYP the GGA subroutine would call a subroutine that calculates the Becke88 exchange energy and potential contributions at that point and another subroutine to calculate the LYP correlation energy and potential contributions at the same point. In the spin polarised case alternative subroutines are invoked at each point.

In the case of hybrid functionals the subroutine called by the generic subroutine calls the subroutines for the required components and applies the appropriate scaling to the contribution of every component functional. Taking B3LYP as an example, the subroutine which calculates f_x^{B3LYP}, calls subroutines which calculate the L(S)DA exchange and the Becke88 exchange energy and potential contributions, and then scales and sums these according to the B3LYP recipe (0.72 Becke88 + 0.08 LDA). The Hartree-Fock exchange contribution is calculated and scaled elsewhere in the code.

6.3.1   Tests on an organometallic system

Table 6.9 compares the performance of the various hybrid functionals in ONETEP. The spherical wave electrostatic metric method has been used to calculate the Hartree-Fock exchange energy, the metric matrix is generated using the reciprocal grid method; the FFT method was used to calculate the diagonal atomblocks of the Hartree-Fock exchange matrix. The four hybrid functional methods produced energies that are similar to each other. The exception to this is the energy for stage 2b for B3LYP, this calculation was very far from convergence. However the energy differences are rather different to those obtained with NWChem. This is in part due to the fact that the calculations were not fully converged. This lack of convergence can have a large effect on the energy differences.

Chapter 7

Conclusions

7.1 Summary

This dissertation sought to describe developments towards improving the accuracy of linear-scaling Density Functional Theory (DFT). Chapter 1 began with an introduction to quantum mechanics and continued by considering approximate methods for the solution of the Schrödinger equation. Hartree-Fock theory, DFT and density matrix theory were introduced. ONETEP [62] a linear-scaling density matrix method where the energy is minimised directly with respect to the density matrix was introduced in chapter 2. The localised orbitals and energy minimisation procedure were discussed.

In chapter 3 an empirical correction to improve the description of dispersion forces in DFT developed as part of this thesis is described. By optimising the parameters in this correction scheme for a large set of dispersion bound complexes, the treatment of dispersion interactions in ONETEP is much improved and validated on a variety of systems.

A method to calculate Hartree-Fock exchange with quadratic-scaling cost was presented in chapter 4. This method is accurate and converges well. However it has unfavourable computational scaling that makes it unsuitable for calculations on large systems. It was found that by representing the NGWF products on the coarse grid rather than the fine grid, significant speed improvements were obtained with minimal effect on accuracy. This also led to a reduction in the memory requirements of the calculations because fine grid arrays were no longer required.

In chapter 5 methods for calculating Hartree-Fock exchange with linear-scaling cost which were developed in this thesis were presented. The numerical pointwise approach converged well and gave accurate energies. Using Gaussians as an auxiliary basis set to expand the NGWF product did not give satisfactory results. Spherical waves are used as an alternative auxiliary basis set, with improved results. The electrostatic metric was found to be more suitable than the overlap metric due to the improved convergence that the former method demonstrates. Convergence appears to be sensitive to inaccuracies in the metric matrix and strategies to improve this were investigated. It was found that using a reciprocal Cartesian grid was the best method of calculating the metric matrix. Using the Fourier transform method from chapter 4 to calculate the diagonal atomblocks

of the exchange matrix lead to an improvement in accuracy and convergence with little extra cost. This is the preferred method for calculating the Hartree-Fock exchange energy using spherical waves. Linear-scaling has been demonstrated for a polythene chain for the numerical pointwise approach and the spherical wave electrostatic metric method.

The application of Hartree-Fock exchange in DFT to create hybrid functionals is discussed in chapter 6. Calculations with these hybrid functionals were applied in the study of the reaction paths of an organometallic system. Results obtained using the hybrid functionals that were implemented in ONETEP have been compared to those obtained using a Gaussian basis set approach on the same system.

7.2 Future work

The empirical dispersion correction could be improved by optimising the parameters for other sets of complexes. Currently there are only optimised parameters for hydrogen, carbon, nitrogen, oxygen and sulphur. It would be desirable to have optimised parameters for further elements, especially for the halogens and phosphorus. The ability to change the parameters used at runtime rather than compile time would be a valuable improvement. The treatment of hydrogen-bonded systems by the dispersion correction could be improved to remove the tendency to slightly over-bind in these cases. Another possibility is to calculate optimised parameters for carbon nanostructures.

The significant improvement in speed, with minimal effect on accuracy, provided by the use of the coarse grid instead of the fine grid in chapter 4 suggests that it would be worthwhile to investigate the applicability of this approximation to other areas of ONETEP such as the the calculation of the density. This approximation should lead to large time savings by avoiding Fourier interpolations and reducing the whole cell fine grid Fourier transform for the Hartree potential to a coarse grid Fourier transform. There would also be a reduction in memory use since the large arrays used to hold quantities on the fine grid would be superfluous. These savings would allow ONETEP to treat larger systems.

The method of using spherical waves as an auxiliary basis set to expand the NGWF product has potential for improvement. More sophisticated methods for calculating the metric matrix could be sought to improve convergence. Other strategies for improving the convergence can be explored. The efficiency of the manipulation of the tightboxes could be improved for example by avoiding work for points that are unimportant since they lay outside of the localisation region. The parallel strategy could be modified to reduce the communication demands. Reducing the overhead of communication and a more balanced distribution of the computational effort would improve the performance of this method. Also the use of alternative auxiliary basis sets could be investigated.

Initial tests on the polythene chain indicate that the exchange matrix sparsity currently set to the kernel sparsity could, as an approximation, be set to the GGA Hamiltonian sparsity, without major loss of accuracy. Hartree-Fock exchange is the only direct contributor to these elements in the density kernel gradient. This approximation would hasten the onset of linear scaling and reduce the cost of calculations on large systems. Tests on a

range of systems would be required to validate this approximation.

The effect of the pseudopotential approximation on the exchange energy should be assessed to ensure that energy differences obtained with ONETEP can be compared with all electron calculations.

Bibliography

[1] Andreas Danopoulos. Crystal structures. Private communication, 2009.

[2] Chris-Kriton Skylaris, Peter D. Haynes, Arash A. Mostofi, and Mike C. Payne. Implementation of linear scaling plane wave density functional theory on parallel computers. *Physica Status Solidi b*, 243(5):973–988, 2006.

[3] Quintin Hill and Chris-Kriton Skylaris. Including dispersion interactions in the ONETEP program for linear-scaling density functional theory calculations. *Proceedings of the Royal Society A*, 465(2103):669–683, 2009.

[4] Werner Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43(3-4):172–198, 1927.

[5] Ira N. Levine. *Quantum Chemistry*. Prentice-Hall, fifth edition, 2000.

[6] Robert G. Parr and Weitao Yang. *Density-Functional Theory of Atoms and Molecules*. International Series of Monographs on Chemistry. Oxford, 1989.

[7] E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Physical Review*, 28(6):1049–1070, 1926.

[8] Franz Mandl. *Quantum Mechanics*. Wiley, 1992.

[9] Max Born and Robert Oppenheimer. Zur Quantentheorie der Molekeln. *Annalen der Physik*, 389(20):457–464, 1927.

[10] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry*. Dover, 1996.

[11] Jiří Čížek. On the correlation problem in atomic and molecular systems. calculation of wavefunction components in ursell-type expansion using quantum-field theoretical methods. *The Journal of Chemical Physics*, 45(11):4256–4266, 1966.

[12] Chr. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Physical Review*, 46(7):618–622, 1934.

[13] Krishnan Raghavachari, Gary W. Trucks, John A. Pople, and Martin Head-Gordon. A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters*, 157(6):479–483, 1989.

[14] Christopher J. Cramer. *Essentials of Computational Chemistry*. Wiley, 2nd edition, 2004.

[15] L. Füsti-Molnár and Péter Pulay. Gaussian-based first-principles calculations on large sytems using the Fourier transform Coulomb method. *Journal of Molecular Structure (Theochem)*, 666-667:25–30, 2003.

[16] S. F. Boys and F. Bernardi. The calculation of small molecular interactions by the differences of separate total energies. some procedures with reduced errors. *Molecular Physics*, 19(4):553–566, 1970.

[17] Karl N. Kirschner, Jennifer B. Sorensen, and J. Phillip Bowen. Calculating interaction energies using first principle theories: Consideration of basis set superposition error and fragment relaxation. *Journal of Chemical Education*, 84(7):1225, 2007.

[18] John C. Slater. Atomic shielding constants. *Physical Review*, 36(1):57–64, 1930.

[19] Per-Olof Widmark, editor. *European Summerschool in Quantum Chemistry 2007 Book I*. Lund University, fifth edition, 2007.

[20] Ernest R. Davidson and David Feller. Basis set selection for molecular calculations. *Chemical Reviews*, 86(4):681–696, 1986.

[21] S. F. Boys. Electronic wave functions. i. a general method of calculation for the stationary states of any molecular system. *Proceedings of the Royal Society of London A*, 200(1063):542–554, 1950.

[22] Martijn Marsman, Andreas Grüneis, Joachim Paier, and Georg Kresse. Second-order Møller–Plesset perturbation theory applied to extended systems. i. within the projector-augmented-wave formalism using a plane wave basis set. *The Journal of Chemical Physics*, 130(18):184103, 2009.

[23] Andreas Grüneis, Martijn Marsman, and Georg Kresse. Second-order Møller–Plesset perturbation theory applied to extended systems. ii. structural and energetic properties. *The Journal of Chemical Physics*, 133(7):074107, 2010.

[24] Mike Finnis. *Interatomic Forces in Condensed Matter*. Oxford Series on Material Modelling. Oxford, 2003.

[25] Arash A. Mostofi. *On linear-scaling methods for quantum mechanical first-principles calculations*. PhD thesis, University of Cambridge, 2003.

[26] L. P. Bouckaert, R. Smoluchowski, and E. Wigner. Theory of Brillouin zones and symmetry properties of wave functions in crystals. *Physical Review*, 50(1):58–67, 1936.

[27] A. Baldereschi. Mean-value point in the Brillouin zone. *Physical Review B*, 7(12):5212–5215, 1973.

[28] Hendrik J. Monkhorst and James D. Pack. Special points for Brillouin-zone integrations. *Physical Review B*, 13(12):5188–5192, 1976.

[29] Richard M. Martin. *Electronic Structure*. Cambridge, 2004.

[30] C. J. Tymczak, Valéry T. Weber, Eric Raymond Schwegler, and Matt Challacombe. Linear scaling computation of the Fock matrix. VIII. Periodic boundaries for exact exchange at the Gamma point. *The Journal of Chemical Physics*, 122(12):124105, 2005.

[31] Lewellyn H. Thomas. The calculation of atomic fields. *Proceedings of the Cambridge Philosophical Society*, 23:542–548, 1927.

[32] E. Fermi. Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente. *Zeitschrift für Physik*, 48(1-2):73–79, 1928.

[33] P. A. M. Dirac. Note on exchange phenomena in the thomas atom. *Mathematical Proceedings of the Cambridge Philosophical Society*, 26(03):376–385, 1930.

[34] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B):B864–B871, 1964.

[35] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133–A1138, 1965.

[36] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Physical Review Letters*, 45(7):566–569, 1980.

[37] John P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Physical Review B*, 23(10):5048–5079, 1981.

[38] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics*, 58(8):1200–1211, 1980.

[39] Axel D. Becke. Density-functional thermochemistry. i. the effect of the exchange-only gradient correction. *The Journal of Chemical Physics*, 96(3):2155–2160, 1992.

[40] Axel D. Becke. Density-functional thermochemistry. ii. the effect of the perdew–wang generalized-gradient correlation correction. *The Journal of Chemical Physics*, 97(12):9173–9177, 1992.

[41] Axel D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, 38(6):3098–3100, 1988.

[42] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2):785–789, 1988.

[43] Burkhard Miehlich, Andreas Savin, Hermann Stoll, and Heinzwerner Preuss. Results obtained with the correlation energy density functionals of Becke and Lee, Yang and Parr. *Chemical Physics Letters*, 157(3):200–206, 1989.

[44] Axel D. Becke. Density functional calculations of molecular bond energies. *The Journal of Chemical Physics*, 84(8):4524–4529, 1986.

[45] John P. Perdew. *Electronic Stucture of Solids '91*, page 11. Akademie Verlag, Berlin, 1991.

[46] John P. Perdew, J. A. Chevary, S. H. Vosko, Koblar A. Jackson, Mark R. Pederson, D. J. Singh, and Carlos Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical Review B*, 46(11):6671–6687, 1992.

[47] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, 1996.

[48] John P. Perdew, Stefan Kurth, Aleš Zupan, and Peter Blaha. Accurate density functional with correct formal properties: A step beyond the generalized gradient approximation. *Physical Review Letters*, 82(12):2544–2547, 1999.

[49] Jianmin Tao, John P. Perdew, Viktor N. Staroverov, and Gustavo E. Scuseria. Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Physical Review Letters*, 91(14):146401, 2003.

[50] Axel D. Becke. A new mixing of Hartree–Fock and local density-functional theories. *The Journal of Chemical Physics*, 98(2):1372–1377, 1993.

[51] Axel D. Becke. Density-functional thermochemistry. III. the role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, 1993.

[52] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and Michael J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *Journal of Physical Chemistry*, 98(45):11623–11627, 1994.

[53] Axel D. Becke. Density-functional thermochemistry. v. systematic optimization of exchange-correlation functionals. *The Journal of Chemical Physics*, 107(20):8554–8560, 1997.

[54] B. O. Roos, editor. *Lecture Notes in Quantum Chemistry II*, volume 2 of *Lecture Notes in Chemistry*. Springer-Verlag, 1994.

[55] N. Troullier, J. R. Chelikowsky, and Y. Saad. Calculating large systems with plane-waves - is it a n^3 or n^2 scaling problem. *Solid State Communication*, 93(3):225–230, 1995.

[56] W. Kohn. Density functional and density matrix method scaling linearly with the number of atoms. *Physical Review Letters*, 76(17):3168–3171, 1996.

[57] E. Prodan and W. Kohn. Nearsightedness of electronic matter. *Proceedings of the National Academy of Sciences*, 102(33):11635–11638, 2005.

[58] Jacques Des Cloizeaux. Energy bands and projection operators in a crystal: Analytic and asymptotic properties. *Physical Review*, 135(3A):A685–A697, 1964.

[59] Weitao Yang. Direct calculation of electron density in density-functional theory. *Physical Review Letters*, 66(11):1438–1441, 1991.

[60] Weitao Yang and Tai-Sung Lee. A density-matrix divide-and-conquer approach for electronic structure calculations of large molecules. *The Journal of Chemical Physics*, 103(13):5674–5678, 1995.

[61] X.-P. Li, R. W. Nunes, and David Vanderbilt. Density-matrix electronic-structure method with linear system-size scaling. *Physical Review B*, 47(16):10891–10894, 1993.

[62] Chris-Kriton Skylaris, Peter D. Haynes, Arash A. Mostofi, and Mike C. Payne. Introducing ONETEP: Linear-scaling density functional simulations on parallel computers. *The Journal of Chemical Physics*, 122:084119, 2005.

[63] Gregory H. Wannier. The structure of electronic excitation levels in insulating crystals. *Physical Review*, 52(3):191–197, 1937.

[64] Jacques Des Cloizeaux. Analytical properties of n-dimensional energy bands and Wannier functions. *Physical Review*, 135(3A):A698–A707, 1964.

[65] Arash A. Mostofi, Peter D. Haynes, Chris-Kriton Skylaris, and Mike C. Payne. Preconditioned iterative minimization for linear-scaling electronic structure calculations. *The Journal of Chemical Physics*, 119(17):8842–8848, 2003.

[66] Chris-Kriton Skylaris, Arash A. Mostofi, Peter D. Haynes, Oswaldo Diéguez, and Mike C. Payne. Nonorthogonal generalized Wannier function pseudopotential planewave method. *Physical Review B*, 66(3):035119, 2002.

[67] Chris-Kriton Skylaris, Arash A. Mostofi, Peter D. Haynes, Chris J. Pickard, and Mike C. Payne. Accurate kinetic energy evaluation in electronic structure calculations with localized functions on real space grids. *Computer Physics Communications*, 140(3):315–322, 2001.

[68] Arash A. Mostofi, Chris-Kriton Skylaris, Peter D. Haynes, and Mike C. Payne. Total-energy calculations on a real space grid with localized functions and a plane-wave basis. *Computer Physics Communications*, 147(3):788–802, 2002.

[69] Peter D. Haynes, Chris-Kriton Skylaris, Arash A. Mostofi, and Mike C. Payne. ONETEP: linear-scaling density-functional theory with local orbitals and plane waves. *Physica Status Solidi b*, 243(11):2489–2499, 2006.

[70] Nick D. M. Hine, Peter D. Haynes, Arash A. Mostofi, Chris-Kriton Skylaris, and Mike C. Payne. Linear-scaling density-functional theory with tens of thousands of atoms: Expanding the scope and scale of calculations with ONETEP. *Computer Physics Communications*, 180(7):1041–1053, 2009.

[71] J. A. White and D. M. Bird. Implementation of gradient-corrected exchange-correlation potentials in Car-Parrinello total-energy calculations. *Physical Review B*, 50(7):4954–4957, 1994.

[72] Xin Xu and William A. Goddard. The X3LYP extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. *Proceedings of the National Academy of Sciences*, 101(9):2673–2677, 2004.

[73] Leonard Kleinman and D. M. Bylander. Efficacious form for model pseudopotentials. *Physical Review Letters*, 48(20):1425–1428, 1982.

[74] R. McWeeny. Some recent advances in density matrix theory. *Rev. Mod. Phys.*, 32(2):335–369, 1960.

[75] Peter D. Haynes, Chris-Kriton Skylaris, Arash A. Mostofi, and Mike C. Payne. Density kernel optimization in the ONETEP code. *Journal of Physics: Condensed Matter*, 20(29):294207, 2008.

[76] Adam H. R. Palser and David E. Manolopoulos. Canonical purification of the density matrix in electronic-structure theory. *Physical Review B*, 58(19):12704–12711, 1998.

[77] T. Ozaki. Efficient recursion method for inverting an overlap matrix. *Physical Review B*, 64(19):195110, 2001.

[78] Emilio Artacho and Lorenzo Miláns del Bosch. Nonorthogonal basis sets in quantum mechanics: Representations and second quantization. *Physical Review A*, 43(11):5770–5777, 1991.

[79] Christopher A. White, Paul Maslen, Michael S. Lee, and Martin Head-Gordon. The tensor properties of energy gradients within a non-orthogonal basis. *Chemical Physics Letters*, 276(1-2):133–138, 1997.

[80] D. R. Bowler and Mike J. Gillan. Length-scale ill conditioning in linear-scaling DFT. *Computer Physics Communications*, 112(2-3):103–111, 1998.

[81] Mike C. Payne, Michael P. Teter, Douglas C. Allan, T. A. Arias, and J. D. Joannopoulos. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.*, 64(4):1045–1097, 1992.

[82] Chris-Kriton Skylaris, Peter D. Haynes, Arash A. Mostofi, and Mike C. Payne. Recent progress in linear-scaling density functional calculations with plane waves and pseudopotentials: the ONETEP code. *Journal of Physics: Condensed Matter*, 20(6):064209 (9pp), 2008.

[83] Matt Challacombe. A general parallel sparse-blocked matrix multiply for linear scaling SCF theory. *Computer Physics Communications*, 128(1-2):93–107, 2000.

[84] Stefan Grimme. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry*, 25(12):1463–1473, 2004.

[85] F. London. Zur Theorie und Systematik der Molekularkräfte. *Zeitschrift für Physik*, 63(3-4):245–279, 1930.

[86] F. London. Über einige Eigenschaften und Anwendungen der Molekularkräfte. *Zeitschrift für Physikalische Chemie B*, 11:222–251, 1930.

[87] J. E. Lennard-Jones. Cohesion. *Proceedings of the Physical Society*, 43(5):461–482, 1931.

[88] Sándor Kristyán and Péter Pulay. Can (semi)local density functional theory account for the London dispersion forces? *Chemical Physics Letters*, 229(3):175–180, 1994.

[89] José M. Pérez-Jordá and Axel D. Becke. A density-functional study of van der Waals forces: rare gas diatomics. *Chemical Physics Letters*, 233:134–137, 1995.

[90] Evert Jan Meijer and Michiel Sprik. A density-functional study of the intermolecular interactions of benzene. *The Journal of Chemical Physics*, 105(19):8684–8689, 1996.

[91] Urs Zimmerli, Michele Parrinello, and Petros Koumoutsakos. Dispersion corrections to density functionals for water aromatic interactions. *The Journal of Chemical Physics*, 120(6):2693–2699, 2004.

[92] Marcus Elstner, Pavel Hobza, Thomas Frauenheim, Sándor Suhai, and Efthimios Kaxiras. Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. *The Journal of Chemical Physics*, 114(12):5149–5155, 2001.

[93] Qin Wu and Weitao Yang. Empirical correction to density functional theory for van der Waals interactions. *The Journal of Chemical Physics*, 116(2):515–524, 2002.

[94] H. Rydberg, N. Jacobson, P. Hyldgaard, S. I. Simak, B. I. Lundqvist, and David C. Langreth. Hard numbers on soft matter. *Surface Science*, 532-535:606–610, 2003.

[95] Petr Jurečka, Jirí Černý, Pavel Hobza, and Dennis R. Salahub. Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. *Journal of Computational Chemistry*, 28(2):555–569, 2007.

[96] Takeshi Sato, Takao Tsuneda, and Kimikihiko Hirao. Van der Waals interactions studied by density functional theory. *Molecular Physics*, 103(6-8):1151–1164, 2005.

[97] Takeshi Sato, Takao Tsuneda, and Kimikihiko Hirao. A density-functional study on π-aromatic interaction: Benzene dimer and naphthalene dimer. *The Journal of Chemical Physics*, 123:104307, 2005.

[98] Yue Zhang and Dennis R. Salahub. A reparametrization of a meta-GGA exchange-correlation functional with improved descriptions of van der Waals interactions. *Chemical Physics Letters*, 436(4-6):394–399, 2007.

[99] David C. Langreth, M. Dion, H. Rydberg, Elsebeth Schröder, P. Hyldgaard, and B. I. Lundqvist. Van der Waals density functional theory with applications. *International Journal of Quantum Chemistry*, 101(5):599–610, 2004.

[100] Hans-Joachim Böhm and Reinhart Ahlrichs. A study of short-range repulsions. *The Journal of Chemical Physics*, 77(4):2028–2034, 1982.

[101] Wijnand T.M. Mooij, Frans B. van Duijneveldt, Jeanne G. C. M. van Duijneveldt-van de Rijdt, and Bouke P. van Eijck. Transferable ab initio intermolecular potentials. 1. derivation from methanol dimer and trimer calculations. *Journal of Physical Chemistry A*, 103(48):9872–9882, 1999.

[102] Thomas A. Halgren. The representation of van der Waals (vdw) interactions in molecular mechanics force fields: potential form, combination rules, and vdw parameters. *Journal of the American Chemical Society*, 114(20):7827–7843, 1992.

[103] John C. Slater and John G. Kirkwood. The van der Waals forces in gases. *Physical Review*, 37(6):682–697, 1931.

[104] Stefan Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry*, 27(15):1787–1799, 2006.

[105] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Physical Chemistry Chemical Physics*, 10:6615–6620, 2008.

[106] Petr Jurečka, Jirí Černý, Pavel Hobza, and Jirí Šponer. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Physical Chemistry Chemical Physics*, 8:1985–1993, 2006.

[107] Claudio A. Morgado, Jonathan P. McNamara, Ian H. Hillier, Neil A. Burton, and Mark A. Vincent. Density functional and semiempirical molecular orbital methods including dispersion corrections for the accurate description of noncovalent interactions involving sulfur-containing molecules. *Journal of Chemical Theory and Computation*, 3(5):1656, 2007.

[108] Yingkai Zhang and Weitao Yang. Comment on "Generalized gradient approximation made simple". *Physical Review Letters*, 80(4):890, 1998.

[109] B. Hammer, L. B. Hansen, and J. K. Nørskov. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Physical Review B*, 59(11):7413–7421, 1999.

[110] Maxim Tafipolsky and Rochus Schmid. A general and efficient pseudopotential Fourier filtering scheme for real space methods using mask functions. *The Journal of Chemical Physics*, 124(17):174102, 2006.

[111] M.O. Sinnokrot and C. David Sherrill. Highly accurate coupled cluster potential energy curves for the benzene dimer: Sandwich, T-shaped, and parallel-displaced configurations. *Journal of Physical Chemistry A*, 108(46):10200–10207, 2004.

[112] Thom H. Jr. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90(2):1007–1023, 1989.

[113] Rick A. Kendall, E. Apra, et al. High performance computational chemistry: An overview of NWChem a distributed parallel application. *Computer Physics Communications*, 128(1-2):260–283, 2000.

[114] M. R. Jarvis, I. D. White, R. W. Godby, and Mike C. Payne. Supercell technique for total-energy calculations of finite charged and polar systems. *Physical Review B*, 56(23):14972–14978, 1997.

[115] Carlo A. Rozzi, Daniele Varsano, Andrea Marini, Eberhard K. U. Gross, and Angel Rubio. Exact Coulomb cutoff technique for supercell calculations. *Physical Review B*, 73(20):205119, 2006.

[116] S. J. A. van Gisbergen, J. G. Snijders, and E. J. Baerends. A density functional theory study of frequency-dependent polarizabilities and van der Waals dispersion coefficients for polyatomic molecules. *The Journal of Chemical Physics*, 103(21):9347–9354, 1995.

[117] Jiyun Kuang and C. D. Lin. Two-centre exchange integrals for complex exponent Slater orbitals. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 29(24):L889–L895, 1996.

[118] Mark A. Watson, Nicholas C. Handy, and Aron J. Cohen. Density functional calculations, using Slater basis sets, with exact exchange. *The Journal of Chemical Physics*, 119(13):6475–6481, 2003.

[119] Lilian Berlu and Hassan Safouhi. Multicentre two-electron Coulomb and exchange integrals over Slater functions evaluated using a generalized algorithm based on nonlinear transformations. *Journal of Physics A: Mathematical and General*, 37(10):3393–3410, 2004.

[120] J. J. Fernández, R. López, I. Ema, G. Ramírez, and J. Fernández Rico. Auxiliary functions for molecular integrals with Slater-type orbitals. i. translation methods. *International Journal of Quantum Chemistry*, 106(9):1986–1997, 2006.

[121] Reinhart Ahlrichs, Michael Bär, Marco Häser, Hans Horn, and Christoph Kölmel. Electronic structure calculations on workstation computers: The program system turbomole. *Chemical Physics Letters*, 162(3):165–169, 1989.

[122] Gunnar Karlström, Roland Lindh, Per-Åke Malmqvist, Björn O. Roos, Ulf Ryde, Valera Veryazov, Per-Olof Widmark, Maurizio Cossi, Bernd Schimmelpfennig, Pavel Neogrady, and Luis Seijo. MOLCAS: a program package for computational chemistry. *Computational Materials Science*, 28(2):222–239, 2003. Proceedings of the Symposium on Software Development for Process and Materials Design.

[123] K. Doll, N. M. Harrison, and V. R. Saunders. Analytical Hartree-Fock gradients for periodic systems. *International Journal of Quantum Chemistry*, 82(1):1–13, 2001.

[124] Reinhart Ahlrichs. Methods for efficient evaluation of integrals for Gaussian type basis sets. *Theoretical Chemistry Accounts*, 33(2):157–167, 1974.

[125] M. Bénard and M. Barry. Efficient evaluation of molecular integrals over s,p,d,f Gaussian basis sets. *Computers & Chemistry*, 3(2-4):121–124, 1979.

[126] Martin Head-Gordon and J. A. Pople. A method for two-electron Gaussian integral and integral derivative evaluation using recurrence relations. *The Journal of Chemical Physics*, 89(9):5777–5786, 1988.

[127] Peter M. W. Gill, Martin Head-Gordon, and J. A. Pople. Efficient computation of two-electron - repulsion integrals and their nth-order derivatives using contracted Gaussian basis sets. *Journal of Physical Chemistry*, 94(14):5564–5572, 1990.

[128] Peter M. W. Gill, Martin Head-Gordon, and J. A. Pople. An efficient algorithm for the generation of two-electron repulsion integrals over Gaussian basis functions. *International Journal of Quantum Chemistry*, 36(S23):269–280, 1989.

[129] Peter M. W. Gill and J. A. Pople. The prism algorithm for two-electron integrals. *International Journal of Quantum Chemistry*, 40(6):753–772, 1991.

[130] Roland Lindh, U. Ryu, and B. Liu. The reduced multiplication scheme of the Rys quadrature and new recurrence relations for auxiliary function based two-electron integral evaluation. *The Journal of Chemical Physics*, 95(8):5889–5897, 1991.

[131] Roland Lindh. The reduced multiplication scheme of the Rys-Gauss quadrature for 1st order integral derivatives. *Theoretical Chemistry Accounts*, 85(6):423–440, 1993.

[132] Kazuya Ishimura and Shigeru Nagase. A new algorithm of two-electron repulsion integral calculations: a combination of Pople–Hehre and McMurchie–Davidson methods. *Theoretical Chemistry Accounts*, 120(1-3):185–189, 2008.

[133] Sanjay Chawla and Gregory A. Voth. Exact exchange in ab initio molecular dynamics: An efficient plane-wave based algorithm. *The Journal of Chemical Physics*, 108(12):4697–4700, 1998.

[134] James Spencer and Ali Alavi. Efficient calculation of the exact exchange energy in periodic systems using a truncated Coulomb potential. *Physical Review B*, 77(19):193110, 2008.

[135] François Gygi and A. Baldereschi. Self-consistent Hartree-Fock and screened-exchange calculations in solids: Application to silicon. *Physical Review B*, 34(6):4405–4408, 1986.

[136] Michael C. Gibson, Stuart Brand, and Stewart J. Clark. Screened-exchange stress tensor in density functional theory. *Physical Review B*, 73(12):125120, 2006.

[137] Joachim Paier, Martijn Marsman, K. Hummer, Georg Kresse, Iann C. Gerber, and János G. Ángyán. Screened hybrid density functionals applied to solids. *The Journal of Chemical Physics*, 124(15):154709, 2006.

[138] Martijn Marsman, Joachim Paier, A Stroppa, and Georg Kresse. Hybrid functionals applied to extended systems. *Journal of Physics: Condensed Matter*, 20(6):064201 (9pp), 2008.

[139] Iann C. Gerber, János G. Ángyán, Martijn Marsman, and Georg Kresse. Range separated hybrid density functional with long-range Hartree-Fock exchange applied to solids. *The Journal of Chemical Physics*, 127(5):054101, 2007.

[140] Carsten Rostgaard. Exact exchange in density functional calculations (an implementation in the projector augmented wave method). Master's thesis, Technical University of Denmark, 2006.

[141] P. E. Blöchl. Projector augmented-wave method. *Physical Review B*, 50(24):17953–17979, 1994.

[142] J Ihm, Alex Zunger, and M L Cohen. Momentum-space formalism for the total energy of solids. *Journal of Physics C: Solid State Physics*, 12(21):4409, 1979.

[143] Takeshi Yanai, David P. Tew, and Nicholas C. Handy. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (cam-b3lyp). *Chemical Physics Letters*, 393(1-3):51–57, 2004.

[144] Stewart J. Clark, Matthew D. Segall, Chris J. Pickard, Phil J. Hasnip, Matt. I. J. Probert, Keith Refson, and Mike C. Payne. First principles methods using CASTEP. *Zeitschrift für Kristallographie*, 220(5-6):567–570, 2005.

[145] Chris-Kriton Skylaris, Peter D. Haynes, Arash A. Mostofi, and Mike C. Payne. Using onetep for accurate and efficient o(n) density functional calculations. *Journal of Physics: Condensed Matter*, 17(37):5757–5769, 2005.

[146] E. J. Bylaska, W. A. de Jong, et al. "NWChem, a computational chemistry package for parallel computers, version 5.1.1", 2009.

[147] G. Reza Ahmadi and Jan Almlöf. The Coulomb operator in a Gaussian product basis. *Chemical Physics Letters*, 246(4-5):364–370, 1995.

[148] Brett I. Dunlap. Variational fit to the nonlocal electron-electron exchange potential: Is an accurate n^3 Hartree-Fock method possible? *Physical Review Letters*, 51(7):546–549, 1983.

[149] Marco Häser and Reinhart Ahlrichs. Improvements on the direct SCF method. *Journal of Computational Chemistry*, 10(1):104–111, 1989.

[150] Eric Raymond Schwegler and Matt Challacombe. Linear scaling computation of the Hartree–Fock exchange matrix. *The Journal of Chemical Physics*, 105(7):2726–2734, 1996.

[151] Eric Raymond Schwegler, Matt Challacombe, and Martin Head-Gordon. Linear scaling computation of the Fock matrix. II. Rigorous bounds on exchange integrals and incremental Fock build. *The Journal of Chemical Physics*, 106(23):9708–9717, 1997.

[152] Christian Ochsenfeld, Christopher A. White, and Martin Head-Gordon. Linear and sublinear scaling formation of Hartree–Fock-type exchange matrices. *The Journal of Chemical Physics*, 109(5):1663–1669, 1998.

[153] Eric Raymond Schwegler. *Linear scaling computation of the Hartree-Fock exchange matrix*. PhD thesis, Los Alamox National Laboratory, 1998.

[154] Frank Neese, Frank Wennmohs, Andreas Hansen, and Ute Becker. Efficient, approximate and parallel Hartree-Fock and hybrid DFT calculations. a 'chain-of-spheres' algorithm for the Hartree-Fock exchange. *Chemical Physics*, 356(1-3):98–109, 2009.

[155] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. Hybrid functionals based on a screened Coulomb potential. *The Journal of Chemical Physics*, 118(18):8207–8215, 2003.

[156] Artur F. Izmaylov, Gustavo E. Scuseria, and Michael J. Frisch. Efficient evaluation of short-range Hartree-Fock exchange in large molecules and periodic systems. *The Journal of Chemical Physics*, 125(10):104103, 2006.

[157] John C. Burant, Gustavo E. Scuseria, and Michael J. Frisch. A linear scaling method for Hartree–Fock exchange calculations of large molecules. *The Journal of Chemical Physics*, 105(19):8969–8972, 1996.

[158] Herbert A. Früchtl, Rick A. Kendall, Robert J. Harrison, and Kenneth G. Dyall. An implementation of ri–scf on parallel computers. *International Journal of Quantum Chemistry*, 64(1):63–69, 1197.

[159] Brett I. Dunlap, J. W. D. Connolly, and J. R. Sabin. On some approximations in applications of X alpha theory. *The Journal of Chemical Physics*, 71(8):3396–3402, 1979.

[160] Karin Eichkorn, Oliver Treutler, Holger Ošhm, Marco Hašser, and Reinhart Ahlrichs. Auxiliary basis sets to approximate Coulomb potentials. *Chemical Physics Letters*, 240(4):283–289, 1995.

[161] Chris-Kriton Skylaris, Laura Gagliardi, Nicholas C. Handy, Andrew G. Ioannou, Steven Spencer, and Andrew Willetts. On the resolution of identity Coulomb energy approximation in density functional theory. *Journal of Molecular Structure (Theochem)*, 501-502:229–239, 2000.

[162] Florian Weigend. A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. *Physical Chemistry Chemical Physics*, 4:4285–4291, 2002.

[163] Alex Sodt and Martin Head-Gordon. Hartree-Fock exchange computed using the atomic resolution of the identity approximation. *The Journal of Chemical Physics*, 128(10):104106, 2008.

[164] Robert Polly, Hans-Joachim Werner, Frederick R. Manby, and Peter J. Knowles. Fast Hartree-Fock theory using local density fitting approximations. *Molecular Physics*, 102(21-22):2311–2321, 2004.

[165] Ken F. Riley, Michael P. Hobson, and Stephen J. Bence. *Mathematical Methods for Physics and Engineering*. Cambridge University Press, 1998.

[166] Brett I. Dunlap. Robust variational fitting: Gáspár's variational exchange can accurately be treated analytically. *Journal of Molecular Structure (Theochem)*, 501-502:221–228, 2000.

[167] Z. Su and P. Coppens. Rotation of real spherical harmonics. *Acta Cryst.*, A50:636–643, 1994.

[168] George B. Arfken and Hans J. Weber. *Mathematical Methods for Physicists International Edition*. Academic Press, 4th edition, 1995.

[169] Florian Weigend. Hartree-Fock exchange fitting basis sets for H to Rn. *Journal of Computational Chemistry*, 29(2):167–175, 2007.

[170] Peter D. Haynes and Mike C. Payne. Localised spherical-wave basis set for O(N) total-energy pseudopotential calculations. *Computer Physics Communications*, 102(1-3):17–27, 1997.

[171] Christopher W. Murray, Nicholas C. Handy, and Gregory J. Laming. Quadrature schemes for integrals of density functional theory. *Molecular Physics*, 78(4):997–1014, 1993.

[172] Milton Abramowitz and Irene Stegun, editors. *Handbook of mathematical functions.* Dover, 1972.

[173] John P. Perdew, Matthias Ernzerhof, and Kieron Burke. Rationale for mixing exact exchange with density functional approximations. *The Journal of Chemical Physics*, 105(22):9982–9985, 1996.

[174] Carlo Adamo and Vincenzo Barone. Toward reliable adiabatic connection models free from adjustable parameters. *Chemical Physics Letters*, 274(1-3):242–250, 1997.

[175] John P. Perdew and Yue Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B*, 45(23):13244–13249, 1992.

[176] Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: The pbe0 model. *The Journal of Chemical Physics*, 110(13):6158–6170, 1999.

[177] Hartmut L. Schmider and Axel D. Becke. Optimized density functionals from the extended g2 test set. *The Journal of Chemical Physics*, 108(23):9624–9631, 1998.

[178] J. A. Pople and P. C. Hariharan. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretical Chemistry Accounts*, 18(3):213–222, 1973.

[179] W. J. Hehre, R. Ditchfield, and J. A. Pople. Self—consistent molecular orbital methods. xii. further extensions of Gaussian—type basis sets for use in molecular orbital studies of organic molecules. *The Journal of Chemical Physics*, 56(5):2257–2261, 1972.

[180] Andreas Bergner, Michael Dolg, Wolfgang Kuchle, Hermann Stoll, and Heinzwerner Preuss. Ab initio energy-adjusted pseudopotentials for elements of groups 13-17. *Molecular Physics*, 80(6):1431–1441, 1993.

[181] A. Klamt and G. Schüürmann. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2*, 0(5):799 – 805, 1993.