Department of Chemistry
Nanochemistry Research Institute

Linear-Scaling Techniques for First Principles Calculations of Stationary and Dynamic Systems

Burak O. Cankurtaran

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University of Technology

December 2010

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: 

Date: 29 December 2010

Acknowledgements

A sincere thank you to my supervisors, Prof. Julian Gale and Assoc. Prof. Michael Ford. Firstly, for allowing me to reside in Sydney at the Institute for Nanoscale Technology, and more importantly for their invaluable help and guidance on my way to completion of this research work. The leeway afforded to me by my supervisors was invaluable for my development as an independent researcher.

Prof. Michael Cortie deserves a big thanks for always putting on a friendly smile and lots of support for my pipe dreams. I have had the privilege of belonging to two world class institutes, the first is the Nanochemistry Research Institute at Curtin University of Technology and the Institute for Nanoscale Technology at the University of Technology, Sydney. A special thanks to my fellow students, who are also completing or have completed their Phd degree; Donald (Easy-Way) Maclurcan, Nadine Harris, Nicholas Stokes, Dakrong Pissuwan, Dylan Riessen, Benjamin Soule de Bas, Rainer Hoft, Xiaoda (David) Xu, Vijay Bhatia, Tim Lucey and to my fiery friend Jonathan Edgar. Not to forget my comrade in arms Martin Blaber, especially for his *awk*ness. Finally, my family and friends for putting up with a slightly *mature* student.

My PhD was made possible by funding from the Australian Research Council. Very generous computational resources were made available by the Australian Centre for Advanced Computing and Communications (AC3), the Australian Partnership for Advanced Computing (APAC), the Mare Nostrum supercomputer in Barcelona, Spain, and most importantly from the iVEC facility in Western Australia.

Contents

List of Figures

List of Tables

Glossary of Acronyms

BNNT	Boron-Nitride Nanotube
BSSE	Basis Set Superposition Error
CCDC	Cambridge Crystallographic Data Centre
D&C	Divide-and-Conquer
DCTDDFT	Divide-and-Conquer Time-Dependent Density Functional Theory
DFT	Density Functional Theory
DIIS	Direct Inversion of the Iterative Subspace
DMM	Density Matrix Minimisation
DNA	DeoxyriboNucleic Acid
DOS	Density of States
DZ	Double-ζ
DZP	Double-ζ plus Polarisation
EOM	Equation of Motion
FDM	Frozen Density Matrix
GGA	Generalized Gradient Approximation
KB	Klienman and Bylander
KLI	Krieger-Li-Iafrate
KMG	Kim-Mauri-Galli
LCAO	Linear Combination of Atomic Orbitals
LDA	Local Density Approximation
LDM	Localised Density Matrix
MD	Molecular Dynamics
MM	Molecular Mechanics
MOF	Metal Organic Framework
mGGA	meta-GGA
NA	Neutral Atom
OEP	Optimised Effective Potential
order-N	O (N)

PBE	Perdew-Burke-Ernzerhof (GGA parametrisation)
PDOS	Projected Density Of States
QM	Quantum Mechanical
RT	Room Temperature
SCF	Self-Consistent Field
SWNT	Single Walled Carbon Nanotube
SZ	Single-ζ
SZP	Single-ζ plus Polarisation
SIESTA	Spanish Initiative for Electronic Simulations with Thousands of Atoms
TDDFT	Time Dependent Density Functional Theory
UFF	Universal Force Field
XC	Exchange-Correlation
ZIF	Zeolitic Imidazolate Framework

Software

The SIESTA (Spanish Initiative for Electronic Simulations with Thousands of Atoms) package [2–6] is used for density functional theory calculations and as a code base for all implementations of linear-scaling methods presented in this thesis. For molecular mechanics simulations the General Utility Lattice Program (GULP) [7] is used. Visualisation of atomic structures and electronic structure properties are rendered using GDIS [8] and Visual Molecular Dynamics (VMD) [9]. Ray traced images are rendered using POV-ray [10]. Tachyon [11] is used for images rendered with ray tracing and ambient occlusion methods. Scientific plots are created with the Kaleidagraph application from Synergy Software. The generation and editing of atomic structures was carried out using the Atomic Simulation Environment Python library [12].

Abstract

First principles calculations can be a computationally intensive task when studying large systems. Linear-scaling methods must be employed to find the electronic structure of systems consisting of thousands of atoms and greater. The goal of this thesis is to combine the linear-scaling divide-and-conquer (D&C) method with the linear-scaling capabilities of the SIESTA (Spanish Initiative for Electronic Simulations with Thousands of Atoms) density functional theory (DFT) methodology and present this union as a viable approach to large-scale first principles calculations. In particular, the density matrix version of the D&C method is implemented into the SIESTA package. This implementation can accommodate high quality calculations consisting of atom numbers in the tens of thousands using moderate computing resources. Low quality calculations have been tested up to half million atoms using reasonably sized computing resources. The D&C method is extended to better handle atomic dynamics simulations. First, by alleviating issues caused by discontinuities in the potential energy surface, with the application of a switching function on the Hamiltonian and overlap matrices. This allows for a smooth potential energy surface to be generated. The switching function has the additional benefit of accelerating the self-consistent field (SCF) process. Secondly, the D&C frozen density matrix (FDM) is modified to allow for improved charge transfer between the active and constrained regions of the system. This modification is found to reduce both the number of SCF iterations required for self-consistency and the number of relaxation steps in a local geometry optimisation. The D&C paradigm is applied to the real-time approach of time-dependent density functional theory (TDDFT). The method is tested on a linear alkane molecule with varying levels of success. Divergences in the induced dipole moment occur when the external excitation field is aligned parallel to the axis of the molecule. The method succeeds in producing accurate dipole moments when the external field is aligned perpendicular to the molecule. Various techniques are tested to improve the proposed method. Finally, the performance and effectiveness of the current D&C implementation is evaluated by studying three current systems. The first two systems consist of two different DNA sequences and the last system is the large ZIF-100 zeolitic imidazolate framework (ZIF).

The gods confound the man who first found out
How to distinguish hours!
Confound him, too,
Who in this place set up a sun-dial,
To cut and hack my days so wretchedly
Into small portions.

- Titus Maccius Plautus

1 Introduction

Finding exact analytical solutions to quantum mechanical (QM) equations is difficult due to the many-body problem. The quantum mechanical problem is generally solved using numerical methods, such as wavefunction-based post-Hartree-Fock methods [13] or DFT methods [14, 15]. Post-Hartree-Fock methods can be limited to systems consisting of a few atoms (for full configuration interaction), up to approximately hundred atoms (for second-order Møller-Plesset perturbation theory). A standard implementation of DFT can accommodate hundreds of atoms[a]. Computational scientists studying systems of interest are limited to small systems and with the advent of fields such as nanotechnology, it is not possible to study the larger systems found in these fields. Efficient methods must be found and employed to handle larger systems and then possibly large-scale first principles based predictions (the *holy grail* of any computational scientist) can be made in a routine fashion.

The reason for the above size limitation is due to the inherent scaling of the methods. The scaling refers to a theoretical measure of the execution of the method, in terms of either the computation time or the required memory. Using Big-O notation (also known as Landau notation [16]), in terms of the system size, N, DFT methods typically scale up to $\mathrm{O}\left(N^4\right)$ for the calculation time and $\mathrm{O}\left(N^2\right)$ for the memory. With special techniques, such as density fitting [17], the calculation time scaling is reduced to $\mathrm{O}\left(N^3\right)$, where the scaling is dominated by

[a]Some implementations can handle up to a thousand atoms with generous computing resources.

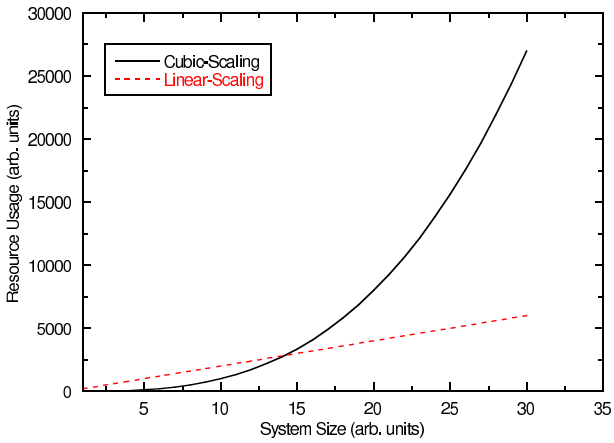


Figure 1.1: A comparison between cubic-scaling and linear-scaling with resource usage (either computation time or memory usage). With a gradient of 200 in this case, the cross-over point occurs at a system size of approximately 14 arb. units.

matrix diagonalisation. What is ideal is to have an algorithm that scales in a linear fashion, $\mathrm{O}(N)$ (order-N), in both computational time and memory. This will allow first principles calculations of much larger systems than are currently feasible with standard algorithms.

The impact of scaling is shown in Figure 1.1 for a hypothetical example. The figure shows a comparison between an arbitrary cubic-scaling method and an arbitrary linear-scaling method in terms of resource usage. One can see that as the system size increases, the cubic-scaling method quickly becomes computationally demanding. While on the other hand, the linear-scaling method can easily handle much larger systems. With cubic-scaling, as the system size doubles, the resource usage will increase by a factor of 8.

For a small system size, it is actually more efficient to use the cubic-scaling method. The point at which it becomes beneficial to use the linear-scaling algorithm is called the cross-over point. In the example shown in Figure 1.1, this occurs at a system size of approximately 14 arb. units. The applicability of the linear-scaling method is largely determined by the cross-over point. Reduction of the cross-over point is desirable and is accomplished by reducing the gradient. The gradient of the linear-scaling method corresponds to what is termed the prefactor, where $R = c_1 N$, R being the resource usage, c_1 is the prefactor

and N is the measure of the system size. Actually, all methods have a prefactor, where for higher-order scaling, $R = c_m N^m$, m being the order of the scaling. In summary, linear-scaling methods should have a small prefactor in order to reduce the cross-over point.

1.1 Principle of Locality

The concept of locality within quantum mechanics can be used to derive linear-scaling methods for many-atom systems. Locality within quantum mechanics, in respect to this study, refers to an object or property being strongly influenced by its immediate surroundings and only weakly influenced by factors some distance away. An example, to highlight the locality principle, can be found with a system consisting of highly localised sigma bonds, in that the properties of the bond are strongly determined *only* by its neighbouring atoms.

Locality was qualitatively defined by Walter Kohn [18] as the concept of "near-sightedness" of electrons in many-body systems. The concept proposes that locality exists for properties, such as the electron charge density[b], and that the dependence of the surrounding environment (e.g. external potential) at a particular point of interest of the property is significant only at nearby points. The dependence is reduced the further one is from the point of interest. In fact, changes in the surrounding environment beyond a certain radius have severely limited influence at the point of interest.

Because the dependence rapidly decays to zero as a function of the radius, the computational effort needs only be concerned with the local surrounding environment. The global problem of considering the complete system is avoided. This forms the basis of the majority of order-N methods [19]. Generally, within most linear-scaling methods, atom-centred localisation regions are assembled to represent the local properties of the atom. Regardless of the method, the computational effort is concentrated on each localisation region separately, then by combining the information gained from each localisation region a solution to the global problem can be found.

[b]The electron charge density refers to the probability of the finding an electron at a particular point in space. See section 2.2 for details.

1.2 Present Study

The work presented in this thesis concentrates on the density matrix version of the D&C linear-scaling method [20–22] and its application to the ground and excited states in the framework of (TD)DFT. The D&C method was one of the first proposed linear-scaling techniques for density-based quantum mechanical methods, relying on the locality found in the electron charge density [18] for its scaling. In D&C, the locality is represented by localisation regions termed subsystems. By calculating the density matrix of each individual subsystem and then combining the subsystem density matrices to form the complete density matrix, the global problem of finding a solution is avoided. The main advantage of the technique is that it is applicable to all system types, where as other linear-scaling methods are generally difficult to apply to systems with a small or non-existent band-gap. This is not say that the D&C method does not have issues with systems that have small or non-existent band-gaps. The decay length of the locality dependence, mentioned in section 1.1, will increase as the band-gap decreases, which in turn will increase the size of the localisation regions needed to represent the required interactions. The requirement of large localisation regions amount to large prefactors.

The D&C method has been relatively neglected until recently [23] within the condensed matter physics field, though in the past it has found significant use within the quantum chemistry community, who have concentrated on semi-empirical QM methods for biological systems [24–27]. The post-Hartree-Fock community have also shown interest in the method recently [28–31].

For this thesis, it is proposed that a highly efficient and parallel order-N code can be achieved by using the D&C method in conjunction with the numerical atomic basis set of the DFT SIESTA methodology. The aim here is to have the ability to handle systems with the number of atoms in the tens of thousands with modest computing resources and in the hundred of thousands with larger computing capacity. This is achieved in ground state calculations and with further development for the time-dependent case, where a new method based on the D&C method is proposed that will be able to handle atom numbers in the

thousands. A brief summary of the work done in each chapter follows.

In chapter 2, the background theory required for the understanding of the subsequent chapters is presented. The quantum many-body problem is posed in the time-dependent and stationary formulations, with the description of some standard approximations that help reduce the complexity. The proofs and derivations of DFT and TDDFT are shown. The linear combination of atomic orbitals (LCAO) basis set computational interpretation of DFT and TDDFT are explained with an emphasis on the SIESTA methodology. The formulation of the linear-scaling Kim-Mauri-Galli (KMG) energy functional orbital minimisation [32] method is also presented.

Chapter 3 describes the parallel implementation of the D&C method into the SIESTA package. The linear-scaling ability of the implementation is presented for insulating, semi-conducting and near-metallic systems. The total energy convergence with respect to basis set size and subsystem size is investigated. Comparisons with the KMG method are also included. The parallel scaling of the implementation is tested on bulk silicon. First steps towards a new linear-scaling method are also made here, where the D&C method is used to seed the KMG-based orbital minimisation method. The hybrid method should have a lower prefactor than both the D&C method and the KMG method.

Extensions to the standard D&C method in regards to molecular dynamics (MD) simulations and local geometry optimisations are examined in chapter 4. Specifically, the D&C method produces energy discontinuities in the potential energy surface as atoms enter or leave subsystems. To alleviate the possible problems the discontinuities can pose to dynamics simulations, a switching function is applied to the boundary of the subsystems to taper any interactions between atoms at the boundary and the rest of the atoms in the subsystem. This has the effect of smoothing the energy discontinuity. An outer buffer region [25] is also applied and, when used in conjunction with the switching function, produces very smooth energy surfaces. A desirable side effect of the tapering is that the number of SCF iterations to reach self-consistency is reduced. The second part of this chapter looks at the FDM [27,33] version of the D&C method. In particular, it is proposed that the use of multiple regions (instead of just a frozen and ac-

tive region) allows for faster SCF convergence and faster geometry optimisation convergence. The FDM method is also implemented to work in parallel.

Application of the D&C method to the TDDFT formulation is shown in chapter 5. The formulation of electron dynamics is shown and the real-time propagation method [34] is described. The D&C TDDFT (DCTDDFT) formulation is also presented. The DCTDDFT method is tested on a linear alkane molecule, with results suggesting that either the subsystem boundaries or the stability of the method is the cause of the eventual divergence of the dipole moment. The investigation into its causes should allow for future solutions. The method is expected to be able to handle the same system size of atoms as the ground state calculation, though, due to its non-linear memory scaling, only thousands of atoms are possible at this stage.

Chapter 6 highlights both the capabilities and limitations of the present D&C implementation. Three systems are examined using the D&C implementation. The first system considered is λ-DNA [1], where the effectiveness of the tapering mechanism is investigated. A comparison with a second DNA system is made against published results of Otsuka *et al* [35], who used the density matrix minimisation (DMM) [36] linear-scaling technique implemented in CONQUEST [37–39]. A comparison of the electronic structure for wet and dry DNA structures is also made. The full power of the D&C implementation is utilised for the third system under investigation, where the electronic structure of the large 13,584-atom ZIF-100 [40] structure is found using only comparatively moderate computing resources.

Finally, conclusions are drawn for the work presented in this thesis. Possible future improvements to the current D&C implementation, as well as suggestions for future developments are discussed.

2

Background Theory

The dynamics of quantum systems are governed by the Schrödinger equation. The mechanics of the quantum particles are coupled with each other leading to difficulties in finding analytical solutions. Generally this is referred to as the many-body problem, as found in many other fields of the physical sciences. DFT and TDDFT are methods with which to overcome the many-body problem. They replace the many-body wavefunction solutions of the Schrödinger equation with the electron density, that depends on only three spatial coordinates plus a spin coordinate. This chapter introduces the methods and key concepts of (TD)DFT and is the basis for any other theories and methods presented in this thesis[a].

2.1 Introduction

The dynamics of quantum systems is governed by the evolution of the Schrödinger equation (2.1) in time. This initial-valued, first order in time partial differential equation describes the propagation of a wavefunction solution, also referred to as a quantum state. The Schrödinger equation describes the behaviour of most subatomic particles, where this study is predominantly concerned with the motion and properties of electrons. The following is the general formulation of the time-

[a]The theory and methods described in this chapter provide general background knowledge for the understanding of the current thesis, and of any theory, presented in later chapters. It is by no means complete; for a detailed description refer to references [41–43].

dependent Schrödinger equation;

$$i\frac{\partial\Psi(\mathbf{x},t)}{\partial t} = \hat{H}\Psi(\mathbf{x},t) \tag{2.1}$$

where atomic units ($\hbar = m_e = e = 1$) have been used and will be continued to be used in the rest of this work. The wavefunction solution, Ψ, is a function of variables $\mathbf{x} = \{\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N, \boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_M\}$ as the coordinates of N electrons and M nuclei. The coordinates comprise of spatial coordinates $\boldsymbol{r}_i$ for electrons and $\boldsymbol{R}_I$ for the nuclei. Spin coordinates have been excluded for brevity. The Hamiltonian operator $\hat{H}$ is expanded as;

$$\hat{H} = \hat{T} + \hat{T}_{\mathcal{N}} + \hat{V} \tag{2.2}$$

which consists of the electron kinetic energy operator, $\hat{T}$, the nuclear kinetic energy operator, $\hat{T}_{\mathcal{N}}$, and the potential, $\hat{V}$, felt by each particle. The electron kinetic energy is defined as;

$$\hat{T} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 \tag{2.3}$$

and the nuclei kinetic energy as;

$$\hat{T}_{\mathcal{N}} = -\frac{1}{2}\sum_{I=1}^{M}\frac{1}{M_I}\nabla_I^2 \tag{2.4}$$

where M_I is the mass of nucleus I relative to the mass of an electron. The potential $\hat{V}$ contains the Coulomb interaction with other particles within the system and from any external potentials (e.g. applied electric field), defined as:

$$\hat{V} = \hat{V}_{ee} + \hat{V}_{\mathcal{N}e} + \hat{V}_{\mathcal{N}\mathcal{N}} + \hat{V}_{ext} \tag{2.5}$$

The components of $\hat{V}$ are defined as the following; the electron-electron repulsion operator;

$$\hat{V}_{ee} = \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \tag{2.6}$$

the electron-nucleus attraction operator;

$$\hat{V}_{\mathcal{N}e} = \sum_{i=1}^{N}\sum_{I=1}^{M} -\frac{Z_I}{|\boldsymbol{r}_i - \boldsymbol{R}_I|} \tag{2.7}$$

and the internuclear repulsion operator:

$$\hat{V}_{\mathcal{N}\mathcal{N}} = \sum_{I=1}^{M}\sum_{J>I}^{M} \frac{Z_I Z_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|} \tag{2.8}$$

The external potential operator, $\hat{V}_{\text{ext}}$, describes any potentials sourced outside the system and can be time-dependent. The variables Z_I and Z_J are the positive charges of nuclei I and J, respectively.

The Schrödinger equation (2.1) describes the dynamics of wavefunctions during a period of time, although it is not always necessary to take into consideration time-dependent processes. A time-independent, stationary Schrödinger equation can also be derived and is stated as:

$$\hat{H}\Psi(\mathbf{x}) = E\Psi(\mathbf{x}) \tag{2.9}$$

The time-dependence has been removed and the Schrödinger equation has become an eigenvalue equation. The set of eigenvalues, $E = \{e_1, e_2, \ldots, e_k\}$, represent the energy level values of the system corresponding to the set of eigenfunctions, $\Psi = \{\Psi_1, \Psi_2, \ldots, \Psi_k\}$, where k is the number of eigenstates[b]. This equation describes all ground state properties of the system. It also describes excited state properties, although extracting the excited state properties from the ground state Schrödinger equation is difficult and mostly unknown. The Hamiltonians for various systems differ only by the number of particles and the external potential; all other components of the Hamiltonian are independent of the system of interest.

The wavefunction, Ψ, characterises the behaviour of all atomic particles. Ψ must be well behaved everywhere, obeying any boundary conditions imposed on (2.1) or (2.9). The Born interpretation of the wavefunction is considered to be that the probability, $P(\boldsymbol{r}, t)$, of finding the particle at time, t, in a volume element,

[b]In principle, within the continuum limit, there are an infinite number of states.

$d\boldsymbol{r}^k$, located at point, $\boldsymbol{r}$, is given by the equation;

$$P(\boldsymbol{r},t) = C|\Psi(\boldsymbol{r},t)|^2 d\boldsymbol{r}^k \tag{2.10}$$

where C is a normalisation constant. The probability of finding a particle anywhere in space at time, t, is equal to unity (i.e. $P = 1$):

$$\int |\Psi(\boldsymbol{r},t)|^2 \, \mathrm{d}\boldsymbol{r}^k = \frac{1}{C} \tag{2.11}$$

All integrals are assumed to be evaluated over all space unless stated otherwise. Within operator theory in quantum mechanics, when the set of quantum states is complete and orthonormal then the eigenfunctions, Ψ_i, in accordance with (2.10), can be shown to obey;

$$\int \Psi_i^* \Psi_j \, \mathrm{d}\boldsymbol{r}^k = \langle \Psi_i | \Psi_j \rangle = \delta_{ij} \tag{2.12}$$

where Dirac notation is introduced and Ψ_i^* is the complex conjugate of Ψ_i. The expectation value[c] of an observable is the Hermitian linear operator for the observable, O, defined as the operator, $\hat{O}$, operating on a normalised eigenstate, Ψ:

$$\langle \hat{O} \rangle = \langle \Psi | \hat{O} | \Psi \rangle \tag{2.13}$$

$$= \int \Psi^* \hat{O} \Psi \, \mathrm{d}\boldsymbol{r} \tag{2.14}$$

If Ψ is not normalised then the above equation becomes:

$$\langle \hat{O} \rangle = \frac{\langle \Psi | \hat{O} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{2.15}$$

Finding solutions to either the stationary, or particularly the time-dependent, Schrödinger equation is not a trivial task, due to the many-body problem, found in many varying fields of the physical sciences. To solve these non-trivial, many-body systems, approximations need to be made and the Schrödinger equation

[c]Expectation values are the predicted mean value that correspond to measurements made in experiments, which are statistical in nature.

solved in a numerical fashion. The following sections describe certain approximations and strategies to overcome the many-body problem, thereby providing solutions to the Schrödinger equation.

2.1.1 Born-Oppenheimer approximation

The Born-Oppenheimer approximation [44, 45] takes advantage of the fact that the electrons orbiting nuclei move at much greater velocities than the nuclei due to their smaller mass. At the time-scale of electron motion the nuclei can be treated as stationary, where the contribution to the Hamiltonian is considered as an effective potential, $v_{\text{nuc}}(\boldsymbol{r})$, felt by the electrons. This uncoupling of nuclear and electron motion serves as an effective and often accurate method in reducing the number of particles in the Schrödinger equation and hence reducing the complexity and the time required to solve the system. This makes the problem separable, allowing the wavefunction to be written as:

$$\Psi = \Psi_{electronic}\Psi_{nuclear} \tag{2.16}$$

The new Hamiltonian (2.17) describing the motion of *only* the electrons is known as the electronic Hamiltonian, $\hat{H}_{\text{el}}$;

$$\begin{aligned} \hat{H}_{\text{el}} &= -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} + \sum_{i=1}^{N} v_{\text{nuc}}(\boldsymbol{r}_i) && (2.17)\\ &= \hat{T} + \hat{V}_{\text{ee}} + \hat{V}_{\mathcal{N}\text{e}} && (2.18) \end{aligned}$$

where the external potential due to the field of the nuclei is given by;

$$v_{\text{nuc}}(\boldsymbol{r}) = -\sum_{I=1}^{M}\frac{Z_I}{|\boldsymbol{r} - \boldsymbol{R}_I|} \tag{2.19}$$

$$\hat{H}_{\text{el}}\Psi_{\text{el}} = \text{E}_{\text{el}}\Psi_{\text{el}} \tag{2.20}$$

and the total energy defined as;

$$\text{E}_{\text{total}} = \text{E}_{\text{el}} + \text{E}_{\text{nuc}} \tag{2.21}$$

where E_{nuc} is the constant internuclear repulsion energy:

$$E_{nuc} = \sum_{I=1}^{M} \sum_{J>1}^{M} \frac{Z_I Z_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|} \tag{2.22}$$

From now on only the electronic problem is considered and the subscript 'el' is dropped:

$$\hat{H} \equiv \hat{H}_{el} \tag{2.23}$$

In order to solve the Schrödinger equation for an arbitrary system the Hamiltonian[d] shown in equation (2.17) must be computed for the particular system. The difficult task is then to find a suitable wavefunction solution. As stated before, there exists only a few trivial known solutions. To overcome this problem, the variational principle can be used to systematically find the wavefunction of the ground state.

2.1.2 Variational principle

The variational principle[e] states that the energy, E_{trial}, given by the expectation value of the Hamiltonian operator when calculated using a trial wavefunction, Ψ_{trial}, will always be an upper bound to the true ground state energy, E_0, given by the true ground state wavefunction, Ψ_0. Here the expectation value of the energy is given by;

$$E[\Psi_{trial}] = \frac{\langle \Psi_{trial} | \hat{H} | \Psi_{trial} \rangle}{\langle \Psi_{trial} | \Psi_{trial} \rangle} \tag{2.24}$$

and the variational principle states that:

$$\frac{\langle \Psi_{trial} | \hat{H} | \Psi_{trial} \rangle}{\langle \Psi_{trial} | \Psi_{trial} \rangle} = E_{trial} >= E_0 \tag{2.25}$$

In theory, the ground state wavefunction can be found using the variational principle, although in practice searching through all possible wavefunctions is near

[d]This Hamiltonian is the starting point for a majority of first principles methods and will be used as the starting point for the theorems of DFT.

[e]The variational principle will be used with the proofs associated with the Hohenberg-Kohn theorems.

impossible; thus the search is limited to a small subset of possible wavefunctions. The space of the possible wavefunctions can be spanned by a finite basis set that restricts the search for the ground state to a finite number.

2.2 Density Functional Theory

DFT is a method used to determine the electronic structure of quantum systems. The main premise of the method is to replace the $4N$-dimensional wavefunction solution (spin coordinates included) to the Schrödinger equation with the electronic charge density which is dependent on only three spatial coordinates and a spin coordinate. In principle, the theory leads to exact results although in practice approximations are made for certain quantum aspects of the system. Even with approximations, DFT provides acceptably accurate results for large systems which other *ab initio* methods find computationally intractable.

Early attempts at using the electron density to describe the kinetic energy of a system were made in 1927 by Thomas and Fermi [46, 47]. Using the electron density gave way to the popular orbital formulations of the Hartree-Fock method and its many derivatives (post-Hartree-Fock methods) [13]. It took 37 years, after Thomas and Fermi, for Hohenberg and Kohn in 1964 [48] to produce the theorems of DFT that placed the use of the electron density on a firm theoretical foundation. A year later Kohn and Sham [15] formulated a version of the Hohenberg-Kohn theorems that allowed the direct ground state calculation of many systems within a one-particle framework that provided an expression for the kinetic energy term. It is the Kohn-Sham equations that are the foundation for all DFT work in this study.

The electron density, ρ, is defined as the number of electrons per unit volume for a given state. In terms of the normalised wavefunction, Ψ, it is stated as;

$$\rho(\mathbf{x}) = N \int \cdots \int |\Psi(\mathbf{x}, \mathbf{x_2}, \ldots, \mathbf{x_N})|^2 \, \mathrm{d}\mathbf{x_2} \ldots \mathrm{d}\mathbf{x_N} \tag{2.26}$$

integrating to the number of electrons, N, (spin coordinates removed for brevity):

$$\int \rho(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} = N \tag{2.27}$$

In the above case for ρ, the set of variables, x_N, are $4N$ dimensional where there are $3N$ spatial coordinates and N coordinates to describe the spin of the electron.

Within the context of DFT, the external potential, v_{ext}, combines interactions with the nucleus and any potentials from outside the system. The expectation value of v_{ext} can be written in terms of ρ:

$$\begin{aligned}
\langle \hat{V}_{\text{ext}} \rangle &= \langle \Psi | \sum_{i=1}^{N} v_{\text{ext}}(\boldsymbol{r}_i) | \Psi \rangle \\
&= \int \Psi^* \left(\sum_{i=1}^{N} v_{\text{ext}}(\boldsymbol{r}_i) \right) \Psi \, \mathrm{d}\mathbf{x}^N \\
&= \sum_{i=1}^{N} \int \Psi^* v_{\text{ext}}(\boldsymbol{r}_i) \Psi \, \mathrm{d}\mathbf{x}^N \\
&= \sum_{i=1}^{N} \int \frac{1}{N} \rho(\boldsymbol{r}_i) v_{\text{ext}}(\boldsymbol{r}_i) \, \mathrm{d}\boldsymbol{r}_i \\
&= \int \rho(\boldsymbol{r}) v_{\text{ext}}(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} \qquad (2.28)
\end{aligned}$$

The expressions for the $\langle \hat{T} \rangle$ and $\langle \hat{V}_{ee} \rangle$ cannot be written in this manner.

2.2.1 Hohenberg-Kohn Theorems

The first Hohenberg-Kohn theorem [48] shows that the density, $\rho(\boldsymbol{r})$, of a nondegenerate ground state uniquely determines the external potential, v_{ext}, to within an additive constant. The universality of the kinetic and electron-electron repulsion operators in the Hamiltonian defines a mapping $v_{\text{ext}} \xrightarrow{map} \rho(\boldsymbol{r})$. To complete the mapping an inverse map $\rho(\boldsymbol{r}) \xrightarrow{map} v_{\text{ext}}$ can be proven by the technique of *reductio ad absurdum*. This is done by showing that two external potentials, v_{ext} and v'_{ext}, that give rise to the same nondegenerate ground state electron density, $\rho_0(\boldsymbol{r})$, will differ by more than a constant and not lead to the same ground state wavefunction, Ψ_0. The two external potentials, v_{ext} and v'_{ext}, are associated with the Hamiltonians, $\hat{H}$ and $\hat{H}'$, respectively. The ground state wavefunctions, Ψ_0 and Ψ'_0, and energies, E_0 and E'_0, for the two external potentials are found by

solving the ground state Schrödinger equation:

$$\hat{H}\,|\Psi_0\rangle = E_0\,|\Psi_0\rangle \tag{2.29}$$

$$\hat{H}'\,|\Psi_0'\rangle = E_0'\,|\Psi_0'\rangle \tag{2.30}$$

Using the variational principle (see section 2.1.2) and (2.28) leads to:

$$\begin{aligned} E_0 &\leq \langle\Psi_0'|\hat{H}|\Psi_0'\rangle \\ &= \langle\Psi_0'|\hat{H}'|\Psi_0'\rangle + \langle\Psi_0'|\hat{H}-\hat{H}'|\Psi_0'\rangle \\ &= E_0' + \langle\Psi_0'|\hat{V}_{\text{ext}}-\hat{V}_{\text{ext}}'|\Psi_0'\rangle \\ &= E_0' + \int \rho_0(\boldsymbol{r})\,[v_{\text{ext}}(\boldsymbol{r}) - v_{\text{ext}}'(\boldsymbol{r})]\,\mathrm{d}\boldsymbol{r} \end{aligned} \tag{2.31}$$

Similarly, by interchanging primed and unprimed quantities leads to:

$$\begin{aligned} E_0' &\leq \langle\Psi_0|\hat{H}'|\Psi_0\rangle \\ &= E_0 + \int \rho_0(\boldsymbol{r})\,[v_{\text{ext}}'(\boldsymbol{r}) - v_{\text{ext}}(\boldsymbol{r})]\,\mathrm{d}\boldsymbol{r} \\ &= E_0 - \int \rho_0(\boldsymbol{r})\,[v_{\text{ext}}(\boldsymbol{r}) - v_{\text{ext}}'(\boldsymbol{r})]\,\mathrm{d}\boldsymbol{r} \end{aligned} \tag{2.32}$$

Adding the energies from (2.31) and (2.32) leads to;

$$E_0 + E_0' \leq E_0 + E_0' \tag{2.33}$$

where this can only be true if $\Psi_0 = \Psi_0'$. This creates a contradiction in that the Hamiltonians associated with the wavefunctions differ by more than a constant and cannot possibly be equal. Hence, there cannot exist two different external potentials corresponding to the same nondegenerate ground state electron density, completing the one-to-one mapping between v_{ext}, Ψ_0 and $\rho(\boldsymbol{r})$.

The ground state density now uniquely defines the external potential and the ground state wavefunction. A universal functional can now be written in terms of the density:

$$F[\rho_0(\boldsymbol{r})] = \langle\Psi|\hat{T}+\hat{V}_{\text{ee}}|\Psi\rangle \tag{2.34}$$

This functional is independent of the external potential and hence is valid for any external potential with any number of electrons (2.27 fixes the number of electrons). The total ground state energy functional can now be written in terms of $F[\rho_0(\boldsymbol{r})]$:

$$E_0[\rho_0(\boldsymbol{r})] = \langle \Psi_0[\rho_0(\boldsymbol{r})] | \hat{H} | \Psi_0[\rho_0(\boldsymbol{r})] \rangle = \int \rho_0(\boldsymbol{r}) v_{\text{ext}}(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} + F[\rho_0(\boldsymbol{r})] \qquad (2.35)$$

If the explicit form of 2.35 was known then this functional could be used for any system.

The second Hohenberg-Kohn theorem states that the exact ground state electron density in a given external potential minimises the energy functional 2.35, that is;

$$E_0^v = \min_{\rho_0(\boldsymbol{r})} \left\{ \int \rho_0(\boldsymbol{r}) v_{\text{ext}}(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r} + F[\rho_0(\boldsymbol{r})] \right\} \qquad (2.36)$$

where E_0^v is the ground state energy in a given external potential, v_{ext}. To prove this, Hohenberg and Kohn used the variational principle again. Defining two ground state densities such that $\rho_0'(\boldsymbol{r}) \neq \rho_0(\boldsymbol{r})$, corresponding to two external potentials $v'(\boldsymbol{r}) \neq v(\boldsymbol{r})$, with Hamiltonians, $\hat{H}$ and $\hat{H}'$, and their corresponding ground state energies, E_0 and E_0', then via the variational principle:

$$\begin{aligned} E_0^v[\rho_0'(\boldsymbol{r})] &= \langle \Psi' | \hat{H} | \Psi' \rangle \\ &\geq \langle \Psi | \hat{H} | \Psi \rangle \\ &= E_0^v[\rho_0(\boldsymbol{r})] \\ &= E_0 \end{aligned} \qquad (2.37)$$

Although the Hohenberg-Kohn theorems provide an exact framework for the calculation of any quantum system, the unknown exact form of the universal functional, $F[\rho_0(\boldsymbol{r})]$, requires the use of approximations to find appropriate solutions.

2.2.2 Kohn-Sham Formalism

The Hohenberg-Kohn theorems simplify the calculation of the ground state electronic structure of a many-body system by using the electron density instead of

the many-body wavefunction. They do not, however, provide a method of applying the theorems in practice where the form of the functional F is unknown. A practical framework to approximate the energy functional was proposed by Kohn and Sham [15] reintroducing the use of orbitals into DFT. They created a fictitious system of non-interacting electrons moving in an effective external potential, $v_s(\boldsymbol{r})$, which reproduces the same ground state density as in the corresponding interacting system. The first Hohenberg-Kohn theorem guarantees the uniqueness of $v_s(\boldsymbol{r})$. The primary achievement of the Kohn-Sham formalism is the calculation of the kinetic energy, which is made possible in the Kohn-Sham one-electron framework. This is accomplished through the use of Slater determinants, which are defined below.

Slater determinants are the simplest possible wavefunctions that can represent non-interacting electrons and also obey the antisymmetry requirement due to the Pauli principle. The determinants are composed of one electron wavefunctions typically known as orbitals and are used in the Kohn-Sham formalism. An example of a normalised Slater determinant for N non-interacting electrons can be written as:

$$\Psi_{\mathrm{S}}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \ldots & \psi_N(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \ldots & \psi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_N) & \psi_2(\mathbf{x}_N) & \ldots & \psi_N(\mathbf{x}_N) \end{vmatrix}. \tag{2.38}$$

The electron density is defined as:

$$\rho(\boldsymbol{r}) = \sum_{i=1}^{N} |\psi_i(\boldsymbol{r})|^2 \tag{2.39}$$

Now the representation of the fictitious non-interacting electron system is defined, the next step is find a way to calculate the kinetic energy term.

The kinetic energy functional is separated into a classical and non-classical part. The classical part is a major contribution to the total system energy and can be calculated exactly. The non-classical part has to be approximated as no exact forms are currently known. The separation of the kinetic energy functional

is defined as;

$$\begin{aligned} T[\rho] &= \langle\Psi|\sum_{i=1}^{N} -\frac{1}{2}\nabla_i^2|\Psi\rangle \\ &\equiv \widetilde{T}[\rho] + \langle\Psi_{\mathrm{S}}|\sum_{i=1}^{N} -\frac{1}{2}\nabla_i^2|\Psi_{\mathrm{S}}\rangle \\ &= \widetilde{T}[\rho] + \sum_{i=1}^{N}\langle\psi_i(\boldsymbol{r})| - \frac{1}{2}\nabla^2|\psi_i(\boldsymbol{r})\rangle \\ &\equiv \widetilde{T}[\rho] + T_{\mathrm{S}}[\rho] \end{aligned} \tag{2.40}$$

where T_{S} is the classical kinetic energy of an non-interacting electron and $\widetilde{T}$ is the non-classical part of the kinetic energy (usually termed as the kinetic energy "correction" term). Now that the kinetic energy is defined, the energy functional, E_0^v, is redefined as;

$$E_0^v[\rho] = T[\rho] + E_H[\rho] + V_{\mathrm{ext}}[\rho] + E_{\mathrm{nc}}[\rho] \tag{2.41}$$

$$E_H[\rho] = \frac{1}{2}\iint \frac{\rho(\boldsymbol{r})\rho(\boldsymbol{r'})}{|\boldsymbol{r}-\boldsymbol{r'}|}\mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{r'} \tag{2.42}$$

$$V_{\mathrm{ext}}[\rho] = \int \rho(\boldsymbol{r})v_{\mathrm{ext}}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} \tag{2.43}$$

$$E_{\mathrm{nc}}[\rho] = E_{\mathrm{x}}[\rho] + E_{\mathrm{c}}[\rho] \tag{2.44}$$

where again T is the kinetic energy functional and V_{ext} the potential energy due to the electron - nuclei interaction and any other external field. The electron-electron interaction energy is also divided in two parts. The first part is the classically defined Coulomb contribution called the Hartree energy, E_H; the second part, E_{nc}, contains the non-classical many-body contributions. The non-classical contributions are further divided into the exchange energy, E_{x}, and correlation energy, E_{c}. The exchange energy is due to the antisymmetric properties of fermions and arises due to the Pauli exclusion principle. The correlation energy encompasses all many-body effects due to the instantaneous Coulomb repulsion not considered in E_H and E_{x} and all non-classical effects not considered in the kinetic energy functional, $T[\rho]$. There is a non-physical self-interaction contribution in E_H that E_{nc} has to account for. This form of the energy functional allows for the explicit

expressions for the classic Coulomb and external potential energies, although the exact form for the kinetic energy and non-classical energy functional are not known. The Kohn-Sham formalism provides a good first approximation to these unknown functionals as it produces desirable kinetic energies.

Creating a new functional;

$$E_{\mathrm{XC}}[\rho] = \widetilde{T}[\rho] + E_{\mathrm{nc}}[\rho] \tag{2.45}$$

where $\widetilde{T}$ is absorbed into the non-classical functional, $E_{\mathrm{nc}}[\rho]$. Now all non-classical many-body contributions that cannot be calculated explicitly are accounted for in the exchange-correlation (XC) energy, $E_{\mathrm{XC}}[\rho]$, producing a new total energy functional:

$$\begin{aligned} E_v[\rho] = & - \frac{1}{2}\sum_{i=1}^{N}\langle\psi_i(\boldsymbol{r})|\nabla^2|\psi_i(\boldsymbol{r})\rangle + \frac{1}{2}\iint\frac{\rho(\boldsymbol{r})\rho(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{r}' \\ & + \int\rho(\boldsymbol{r})v_{\mathrm{ext}}(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} + E_{\mathrm{XC}}[\rho] \end{aligned} \tag{2.46}$$

The XC potential, V_{XC}, is defined as functional derivative of the XC energy with respect to the electron density:

$$V_{\mathrm{XC}}(\boldsymbol{r}) \equiv \frac{\delta E_{\mathrm{XC}}}{\delta\rho} \tag{2.47}$$

The final Kohn-Sham equations can be written now as;

$$\left(-\frac{1}{2}\nabla^2 + V_{\mathrm{eff}}(\boldsymbol{r})\right)\psi_i(\boldsymbol{r}) = \varepsilon_i\,\psi_i(\boldsymbol{r}) \tag{2.48}$$

where an effective one-particle potential has been defined:

$$\begin{aligned} V_{\mathrm{eff}}(\boldsymbol{r}) &\equiv v_{ext}(\boldsymbol{r}) + V_H(\boldsymbol{r}) + V_{\mathrm{XC}}(\boldsymbol{r}) \\ V_H(\boldsymbol{r}) &\equiv \int\frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\mathrm{d}\boldsymbol{r}' \end{aligned} \tag{2.49}$$

Using $\hat{f}^{\mathrm{KS}} = \left(-\frac{1}{2}\nabla^2 + V_{\mathrm{eff}}(\boldsymbol{r})\right)$ the Kohn-Sham equations can be neatly written

in their final canonical form as:

$$\hat{f}^{\mathrm{KS}}\psi_i(\boldsymbol{r}) = \varepsilon_i\psi_i(\boldsymbol{r}) \tag{2.50}$$

This eigenvalue equation is solved self-consistently in an effective one-particle field (SCF). The fictitious non-interacting Kohn-Sham orbital solutions, ψ_i, produce the same density as the real interacting system and hence find the same total energy. The only approximation is the form of the XC potential $V_{\mathrm{XC}}(\boldsymbol{r})$.

2.2.2.1 Exchange-Correlation Approximations

There are many approximations to the XC potential with the simplest being the local density approximation (LDA) [15], which is based on the uniform electron gas model. Here the XC potential is equal to the XC of the uniform electron gas with constant density. The approximation is local in that the XC potential at a specific point in the density equals the XC of the uniform electron gas at its corresponding point in the uniform electron gas density, defined as;

$$E_{XC}^{\mathrm{LDA}}\left[\rho(\boldsymbol{r})\right] = \int \rho(\boldsymbol{r})\,\varepsilon_{XC}^{LDA}\left[\rho(\boldsymbol{r})\right]\mathrm{d}\boldsymbol{r} \tag{2.51}$$

where $\varepsilon_{XC}^{\mathrm{LDA}}$ is the XC energy per particle in the uniform electron gas and can be separated into components for the exchange energy, ε_X^{LDA}, and correlation energy, $\varepsilon_C^{\mathrm{LDA}}$. An exact expression for $\varepsilon_X^{\mathrm{LDA}}$ is defined by Dirac [49] as:

$$\varepsilon_x^{\mathrm{LDA}}\left[\rho(\boldsymbol{r})\right] = -\frac{3}{4}\left(\frac{3\rho(\boldsymbol{r})}{\pi}\right)^{\frac{1}{3}} \tag{2.52}$$

The correlation term, $\varepsilon_C^{\mathrm{LDA}}$, does not have an exact expression, thus what is generally used is an interpolation or parameterisation of highly accurate quantum Monte Carlo simulations of a homogeneous electron gas performed by Ceperley and Alder [50]. Researchers, such as Perdew and Wang [51], Perdew and Zunger [52], and Vosko, Wilk and Nusair [53] have published popular parameterised forms of $\varepsilon_C^{\mathrm{LDA}}$.

The success of the LDA is predominantly due to the cancellation of exchange and correlation errors; where the exchange energy is underestimated and the

correlation energy is overestimated. Despite the cancellation of errors, the LDA is still an approximation which can produce large errors for certain system types. Improvements or alternatives to the LDA are necessary.

The generalized gradient approximation (GGA), which is based on the slowly varying electron gas expansion, is an improvement on the LDA. In this case the gradient of the density is also considered at each point in space. Generally the form of the GGA is defined as:

$$E_{\mathrm{XC}}^{\mathrm{GGA}}[\rho^{\alpha},\rho^{\beta},\nabla\rho^{\alpha},\nabla\rho^{\beta}] = \int \rho(\boldsymbol{r})\varepsilon_{\mathrm{XC}}^{\mathrm{GGA}}(\rho^{\alpha},\rho^{\beta},\nabla\rho^{\alpha},\nabla\rho^{\beta})\mathrm{d}\boldsymbol{r} \qquad (2.53)$$

Here the spin density is considered, where ρ^{α} is the majority spin and ρ^{β} is the minority spin. Although, the formal derivation of DFT does not depend on the spin of the electrons, in practice these XC potential approximations are improved when they become functionals of the spin density, especially in cases with odd number of electrons or the treatment of multiplet states. As with the LDA there are several parameterisations available to use. The work in this thesis uses the parameterisation of Perdew, Burke and Ernzerhof (PBE) [54]. In general, these approximations are quite simple and in the majority of cases produce reasonable results.

The next level of approximation above the GGA functional is the meta-GGA (mGGA) functional [55, 56]. These functionals are an extension and improvement to the GGA functional in which the non-interacting kinetic energy density (Laplacian) enters the equation. Earlier mGGA functionals were partially empirical based [57–59], although Tao *et al* have formulated a non-empirical version [60] valid for periodic and molecular systems. Their version is a very important step towards a general density functional. The primary form of the meta-GGA functional is defined as;

$$E_{\mathrm{XC}}^{\mathrm{mGGA}}[\rho^{\alpha},\rho^{\beta},\nabla\rho^{\alpha},\nabla\rho^{\beta},\tau_s^{\alpha},\tau_s^{\beta}] = \int \rho(\boldsymbol{r})\varepsilon_{\mathrm{XC}}^{\mathrm{mGGA}}(\rho^{\alpha},\rho^{\beta},\nabla\rho^{\alpha},\nabla\rho^{\beta},\tau_s^{\alpha},\tau_s^{\beta})\mathrm{d}\boldsymbol{r} \qquad (2.54)$$

where τ_s is the one-electron kinetic energy density (spin-independent), defined

as;

$$\tau_{sj} = \frac{1}{2}\sum_{i}^{occ} |\nabla\phi_{ij}(\boldsymbol{r})|^2 \tag{2.55}$$

in the space of occupied Kohn-Sham orbitals, $\phi_{ij}(\boldsymbol{r})$. Tao *et al* [60] refers to this functional as the third rung of the "Jacob's ladder"of XC functional approximations. Where the first rung is found by removing τ_s and $\nabla\rho$ producing the LDA functional, and the second rung is found by removing only τ_s to produce the GGA functional.

Another type of XC functional comes in the form of hybrid functionals. These functionals take advantage of the exact exchange energy produced from Hartree-Fock calculations by combining it with an LDA or GGA functional. Due to the exchange energy being much larger than the correlation energy, the exact exchange can lead to more accurate results than just LDA or GGA functionals. Early attempts produced unphysical XC holes resulting in poor performance. Modern hybrid functionals use certain amounts of exact exchange in mixtures with the exchange and correlation energies of the LDA or GGA functionals. The weighting of the exact exchange allows for a systematic cancellation of error. For example, because Hartree-Fock typically overestimates band gaps by 200% and the LDA typically underestimates band gaps by 50%, the weighting can take account of this to produce accurate band gaps. Currently, the most popular of these hybrid functionals is B3LYP [61–64].

Finally, a short mention of orbital-dependent based XC functionals [65] is appropriate. The potentials corresponding to the orbital-dependent XC energies, such as the exact exchange energy, are constructed with the optimised effective potential (OEP) [66] or the much simpler, but highly accurate, Krieger-Li-Iafrate (KLI) scheme [67]. The constructed XC potentials are self-interaction free, have the correct long-range -1/r asymptotic behaviour for finite systems and generally produce higher-quality eigenvalues and orbitals. The LDA and GGA potentials exhibit a much faster decaying long-range behaviour, that leads to the underestimation of the band gap and lower quality orbitals due to the fictitious electron interacting with itself; where on the other hand the orbital dependent schemes produce high-quality band structures [65].

2.3 Time-Dependent Density Functional Theory

TDDFT extends the ground state (stationary) DFT method where calculations of time-dependent and excited state phenomena are made possible[f]. Just as the Hohenberg-Kohn theorems, section 2.2.1, establish a one-to-one mapping between the external potential, V_{ext}, and the electron density, $\rho(\boldsymbol{r})$, the Runge-Gross theorem [72] provides a comparable mapping when the external potential and electron density are time-dependent.

According to Runge and Gross there is an unique one-to-one mapping between the time-dependent external potential, $V_{\text{ext}}(\boldsymbol{r}, t)$, and the time-dependent electron density, $\rho(\boldsymbol{r}, t)$, for a given initial state. The Runge-Gross theorem states that two densities, $\rho(\boldsymbol{r}, t)$ and $\rho'(\boldsymbol{r}, t)$, evolving from the same initial state, $\Psi_0(t = t_0 = 0)$, under the influence of two separate potentials, $v_{ext}(\boldsymbol{r}, t)$ and $v'_{ext}(\boldsymbol{r}, t)$, will differ over time if the potentials differ by more than a purely time-dependent function (with the condition that both $v_{ext}(\boldsymbol{r}, t)$ and $v'_{ext}(\boldsymbol{r}, t)$ are Taylor expandable about the initial time, t_0):

$$\Delta v_{ext}(\boldsymbol{r}, t) = v_{ext}(\boldsymbol{r}, t) - v'_{ext}(\boldsymbol{r}, t) \neq c(t) \tag{2.56}$$

Proving this theorem involves showing that the mapping $v_{\text{ext}}(\boldsymbol{r}, t) \xrightarrow{map} \Psi(\boldsymbol{r}, t) \xrightarrow{map} \rho(\boldsymbol{r}, t)$ is invertible, of course up to the purely time-dependent function, $c(t)$. The time-dependent function appears as a phase factor in the wavefunctions as $\Psi(\boldsymbol{r}, t) = e^{-ic(t)}\phi(\boldsymbol{r}, t)$. The phase factor cancels out in the evaluation of the expectation value, allowing the expectation value to be a functional of the density.

The first part of the proof shows that the current densities, $\mathbf{j}(\boldsymbol{r}, t)$ and $\mathbf{j}'(\boldsymbol{r}, t)$, which are also equal at t_0, differ infinitesimally later than t_0. The current density is given by;

$$\mathbf{j}\,(\boldsymbol{r}, t) = N \int d^3r_2 \ldots \int d^3r_N \Im\{\psi(\boldsymbol{r}, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N, t)\nabla\psi^*(\boldsymbol{r}, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N, t)\} \tag{2.57}$$

where $\Im$ signifies the imaginary part. Using the following formulation of the

[f]For recent reviews on TDDFT see Marques *et al* [43], Gross *et al* [68], van Leeuwen [69], Burke *et al* [70] and Elliot *et al* [71].

density matrix operator;

$$\hat{\rho}(\boldsymbol{r}) = \sum_{j=1}^{N} \delta(\boldsymbol{r} - \boldsymbol{r}_j) \tag{2.58}$$

the current density (2.57) is equivalent to;

$$\mathbf{j}\left(\boldsymbol{r}, t\right) = \langle \Psi(\boldsymbol{r}, t) | \hat{\mathbf{j}}(\boldsymbol{r}) | \Psi(\boldsymbol{r}, t) \rangle \tag{2.59}$$

where the current density operator, $\hat{\mathbf{j}}(\boldsymbol{r})$, is defined as:

$$\hat{\mathbf{j}}(\boldsymbol{r}) = \frac{1}{2i} \sum_{j=1}^{N} \left(\nabla_{\boldsymbol{r}j} \, \delta\left(\boldsymbol{r} - \boldsymbol{r}_j\right) + \delta\left(\boldsymbol{r} - \boldsymbol{r}_j\right) \nabla_{\boldsymbol{r}j} \right) \tag{2.60}$$

By examining the equation of motion of the difference of the two current densities;

$$i \frac{\partial}{\partial t} \left[\mathbf{j}(\boldsymbol{r}, t) - \mathbf{j}'(\boldsymbol{r}, t) \right]_{t=0} = i\rho(\boldsymbol{r}, t_0) \nabla \left[v_{\text{ext}}(\boldsymbol{r}, t_0) - v'_{\text{ext}}(\boldsymbol{r}, t_0) \right] \tag{2.61}$$

one can see that if at the initial time the two potentials differ by more than just a constant, then the first derivatives of the current will also differ. Differences in the first derivatives will produce differences in the currents, $\mathbf{j}(\boldsymbol{r}, t) \neq \mathbf{j}'(\boldsymbol{r}, t)$, thus proving that the external potential has a one-to-one correspondence, and is a functional of the current density.

The second part of the proof requires the same to be said for the densities. Using continuity;

$$\frac{\partial \rho(\boldsymbol{r}, t)}{\partial t} = -\nabla \cdot \mathbf{j}\left(\boldsymbol{r}, t\right) \tag{2.62}$$

an equation of motion is defined by taking the gradient of (2.61);

$$\frac{\partial^2}{\partial t^2} \left[\rho(\boldsymbol{r}, t) - \rho'(\boldsymbol{r}, t) \right]_{t=0} = \nabla \cdot \left(\rho(\boldsymbol{r}, t_0) \nabla \left[v_{\text{ext}}(\boldsymbol{r}, t_0) - v'_{\text{ext}}(\boldsymbol{r}, t_0) \right] \right) \tag{2.63}$$

where, similar to the current densities, the densities $\rho(\boldsymbol{r}, t)$ and $\rho'(\boldsymbol{r}, t)$ will differ infinitesimally later than t_0. This proves the mapping $v_{\text{ext}}(\boldsymbol{r}, t) \xrightarrow{map} \Psi(\boldsymbol{r}, t) \xrightarrow{map} \rho(\boldsymbol{r}, t)$ is invertible and that the external potential is a functional of the density and the initial wavefunction, $V_{\text{ext}}\left[\rho, \Psi_0\right](\boldsymbol{r}, t)$.

2.3.1 Time-dependent Kohn-Sham Equations

The same non-interacting electron system used in the ground state Kohn-Sham formalism can also be used in the context of TDDFT. Here the time-dependent Kohn-Sham equations take on the following form;

$$i\frac{\partial \Phi_{ks}}{\partial t} = \hat{H}_{ks}\Phi_{ks} \tag{2.64}$$

where Φ_{ks} are the Kohn-Sham orbitals and $\hat{H}_{ks}$ is the time-dependent Kohn-Sham Hamiltonian. This Hamiltonian is described as follows;

$$\hat{H}_{ks} = -\frac{1}{2}\nabla^2 + V_{\text{ext}}(\boldsymbol{r},t) + \int \frac{\rho(\boldsymbol{r}',t)}{|\boldsymbol{r}-\boldsymbol{r}'|}d\boldsymbol{r}' + V_{xc}\left[\rho\right](\boldsymbol{r},t) \tag{2.65}$$

where the first term on the right describes the majority of the kinetic energy, the second term is the external potential, the third term is the Coulombic repulsion felt between the electrons and the last term is the time-dependent XC functional. The time-dependent XC functional is much more complex than the ground state version. For low lying states and small external perturbations, the time-dependent XC functional can be replaced with the ground state XC functional. This is termed as the adiabatic approximation;

$$V_{xc}\left[\rho\right](\boldsymbol{r},t) \simeq \frac{\delta E_{xc}^{LDA}\left[\rho_t\right]}{\delta\rho_t(\boldsymbol{r})} = V_{xc}^{LDA}\left[\rho_t\right](\boldsymbol{r}) \tag{2.66}$$

where ρ_t is the instantaneous density. The time-dependent density is defined as;

$$\rho(\boldsymbol{r},t) = \sum_{i,j}\rho_{ij}(t)\phi_i(\boldsymbol{r})\phi_j(\boldsymbol{r}) \tag{2.67}$$

where the time-dependent density matrix $\rho_{ij}(t)$ is defined as:

$$\rho_{ij}(t) = \sum_{m}^{occ} C_{mi}(t)C_{mj}(t) \tag{2.68}$$

There are a variety of methods by which to solve (2.64). The approach taken for this study evolves the Kohn-Sham eigenfunctions, Φ_{ks}, in real time [34]. This

method is described in detail in section 5.3. The complete knowledge of $V_{\text{ext}}(\boldsymbol{r}, t)$ and $V_{xc}[\rho](\boldsymbol{r}, t)$ implies solutions of all time-dependent Coulomb interacting systems.

2.4 Computational Implementation

Finding solutions to the Kohn-Sham equations (2.50) would be more appropriate to be solved on a computer if the components of the equations were formulated in an algebraic form. This can be accomplished by expanding the Kohn-Sham orbitals in a set of basis functions. A feature of DFT is that the Kohn-Sham equations are expressible in a wide variety of basis functions, such as planewaves [73], Gaussians [74], wavelets [75], grids [76], B-splines [77], psincs [78], and numerical orbitals [79]. In this present thesis the SIESTA package [2] is used for the construction of the Hamiltonian, bringing with it a focus on the use of real-space localised pseudo-atomic orbital methods (PAOs), while recognising this is just one of many possible approaches. To put the use of PAOs in context, a quick overview of planewave and atomic orbital (Gaussians) approaches are summarised[g].

In the planewave approach [73] the Kohn-Sham orbitals are expanded out with a planewave basis, that is the planewaves extend throughout space. They are suited for the calculation of systems in condensed phases, whereby the planewaves automatically obey Bloch's theorem. Advantageous features of the planewave approach include the use of fast Fourier transform techniques, basis set independence on atomic positions and, most importantly, the systematic control of the convergence of the basis set. Planewaves become disadvantageous when representing inner-shell states and finite systems, where for the rapidly varying inner-shell states, a large number of planewaves are required to describe those states. For finite systems the extra expense of using planewaves to represent the vacuum within the supercell can be large. Generally, a large number of planewaves are required to represent the Kohn-Sham orbitals, which can make this approach computationally expensive.

[g] A concise review of planewave, atomic orbital and real-space techniques can be found in the review article by Beck [80].

On the other hand, the atom-centred atomic orbitals of Gaussian functions and Slater-type orbitals are ideal for finite systems [13, 81]. Atomic orbital functions are more representative of Kohn-Sham orbitals (within atoms only) as they are based on chemically intuitive reasoning. The multicenter integrals in the Kohn-Sham Hamiltonian can be analytically solved using Gaussians, while Slater-type orbitals have to solve many of these integrals numerically. Gaussians come with a price, in that they do not correctly describe the behaviour of the orbitals close to and far away from the nucleus, which in turn requires a larger number of basis functions to accurately describe the states. The Slater-type orbitals do not suffer from this problem. Atomic orbital-based approaches suffer from basis-set superposition error (BSSE), especially when investigating the interaction energies from a chemical reaction. The BSSE can be alleviated with the counterpoise correction due to Boys and Bernadi [82]. Using non-orthogonal basis functions also leads to linear dependence issues with large basis set sizes. These issues, unlike the planewave approach, inhibit a systematic way of obtaining convergence[h] by increasing the number of basis functions i.e. basis set limit.

The number of basis functions in the atomic orbital approach can be still quite large and the computational scaling high. To reduce the cost of using atomic orbitals, PAOs [3–5, 84] can be used, which are purely numerical functions and have arbitrary shape. The shape of the basis function can be chosen as to more accurately represent the actual system wavefunctions. By doing this, the expansion of the wavefunctions can be accomplished with a smaller basis set. SIESTA incorporates the use of PAOs in its methodology [2]. The shape of the PAOs used in SIESTA are physically intuitive as they are derived from the solutions of an atom in a spherical potential [2, 4, 5]. The strict confinement (localisation) of the PAOs [3] lead to linear-scaling in computational work and storage, allowing this approach to be used on large systems. The SIESTA methodology is explained in section 2.4.3, but first, a reformulation of the Kohn-Sham equations in a basis set of a linear combination of atomic orbitals (LCAO) is presented.

[h] Well-tempered basis sets [83], with their large-exponent Gaussians and diffuse orbitals, are designed for more exact excitation energies for finite systems and can in theory be made to converge in a systematic way.

2.4.1 Linear Combination of Atomic Orbitals

The Kohn-Sham equations can be formulated to use a LCAO approach (see [42] for a simple, general overview). First, a Kohn-Sham orbital, $\Phi_i(\boldsymbol{r})$, is expanded as a LCAO by;

$$\psi_i(\boldsymbol{r}) = \sum_{\mu=1}^{L} c_{\mu i}\phi_\mu(\boldsymbol{r}) \tag{2.69}$$

where there are L basis functions, $\phi_\mu(\boldsymbol{r})$. By substituting (2.69) into the canonical Kohn-Sham equations (2.50), multiplying both sides with an arbitrary basis function, $\phi_\nu(\boldsymbol{r})$, and then integrating both sides leads to:

$$\sum_{\mu=1}^{L} c_{\mu i} \int \phi_\nu(\boldsymbol{r})\hat{f}^{\mathrm{KS}}\phi_\mu(\boldsymbol{r})\mathrm{d}\boldsymbol{r} = \varepsilon_i \sum_{\mu=1}^{L} c_{\mu i} \int \phi_\nu(\boldsymbol{r})\phi_\mu(\boldsymbol{r})\mathrm{d}\boldsymbol{r} \tag{2.70}$$

The sums have been left out of the integrals for the purpose of defining the Hamiltonian and overlap matrices:

$$H_{\nu\mu} = \int \phi_\nu(\boldsymbol{r})\hat{f}^{\mathrm{KS}}\phi_\mu(\boldsymbol{r})\mathrm{d}\boldsymbol{r} \tag{2.71}$$

$$S_{\nu\mu} = \int \phi_\nu(\boldsymbol{r})\phi_\mu(\boldsymbol{r})\mathrm{d}\boldsymbol{r} \tag{2.72}$$

This matrix representation of the Kohn-Sham equations is particularly useful for implementation on computers. The Hamiltonian matrix is a real symmetric (Hermitian) matrix. The overlap matrix describes the amount of overlap between the basis functions. For an orthonormal basis set the overlap matrix becomes the identity matrix. Using the Hamiltonian and overlap matrices, (2.70) becomes:

$$\sum_{\mu=1}^{L} c_{\mu i} H_{\nu\mu} = \varepsilon_i \sum_{\mu=1}^{L} c_{\mu i} S_{\nu\mu} \tag{2.73}$$

For all states, i, and basis functions, μ, the coefficients, $c_{\mu i}$, can be represented in matrix form;

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1L} \\ c_{21} & c_{22} & \cdots & c_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ c_{L1} & c_{L2} & \cdots & c_{LL} \end{pmatrix} \tag{2.74}$$

along with a diagonal matrix for the energy eigenvalues, ε_i:

$$\varepsilon = \begin{pmatrix} \varepsilon_1 & 0 & \cdots & 0 \\ 0 & \varepsilon_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \varepsilon_L \end{pmatrix} \equiv \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ddots \\ \varepsilon_L \end{bmatrix} \tag{2.75}$$

Using all the above mentioned matrices and (2.73), a set of linear equations in matrix form can be defined as:

$$(H - \varepsilon S)C = 0 \tag{2.76}$$

There are two primary aspects relating to finding solutions to (2.76). The first aspect is the construction of the matrices, H and S. The construction of the matrices is defined by the type of basis functions. As mentioned in the introduction chapter (see section 1.1), this process is made to scale in a linear-scaling fashion in SIESTA through the strict confinement of PAOs. The second aspect is to solve for C and ε, which currently in SIESTA, is possible through two separate methods. The first method, is standard matrix diagonalisation, which can only be used for small-to-moderately sized systems due to the $\mathrm{O}(N^3)$ computational scaling and $\mathrm{O}(N^2)$ memory scaling. The second method involves using a linear-scaling method based on minimising an energy functional.

2.4.2 Locality for Linear-Scaling

Order-N methods [19] are generally devised based on the decay properties of either the density matrix or the localised Wannier-type function [85–87] repre-

sentation of the density matrix. Localised Wannier functions are a transformation of extended Bloch eigenfunctions from crystalline periodic solids. Typically, any legitimate, localised wave function that can be used to represent the density matrix is referred to as a Wannier-type function. The interaction between any two elements of the density matrix is known to decay exponentially for insulators and algebraically for the metals [19].

By invoking the concept of "near-sightedness" [18], the density matrix can be partitioned into localisation regions. The range of the Wannier-type states will be dependent on the decay properties of the system. For example, Wannier-type states with longer tails will be required to capture the behaviour at a surface of a system with a small or non-existent band-gap.

Solutions to the localisation regions do not depend on the global solution, and thus can be solved separately and combined to form the global solution in a linear-scaling fashion. Even though, the "near-sightedness" principle decouples the very small interactions between distant particles, there still exits an interaction, which by employing localisation regions is in fact an approximation. This leads to the accuracy of the final global solution being dependent on the size of the localisation regions. Obviously, the accuracy is also tied to the decay properties of the density matrix, with smaller localisation regions being required for insulating systems and larger regions for more metallic systems. Eliminating the interaction information outside a given localisation region enables the storage of the density matrix in a linear scaling fashion, as only non-zero terms (the actual interaction terms) need to be stored.

The next section details the adaption of the Kohn-Sham formalism to form the SIESTA methodology which allows a fast computational implementation and solution of the Kohn-Sham equations.

2.4.3 SIESTA Methodology

The SIESTA code [2–6] is a self-consistent DFT package that has been developed primarily to handle systems with large numbers of atoms. All components within SIESTA are designed to scale linearly in computing time and memory usage except for the case of the direct diagonalisation path, where the computational

time scales as $\mathrm{O}(N^3)$ and memory usage as $\mathrm{O}(N^2)$. It is this package that is used for all upcoming DFT calculations and as a base framework for any methods implemented in this work. The following formalism is a *summary* of the SIESTA methodology described by Soler *et al* [2].

The SIESTA methodology mainly deals with the definition and assembly of the Hamiltonian. The SIESTA Hamiltonian is designed to be assembled and stored in a linear-scaling fashion. There are two aspects to the Hamiltonian which will be covered here. The first is the use of a pseudopotential to remove any core electrons from the Hamiltonian which are considered not to be involved in chemical activity. This is another level of approximation, similar to the Born-Oppenheimer approximation (section 2.1.1), that will speed up the calculations and allow for a smooth charge density near the core of the atom. The second aspect is the type of basis function used to represent the orbitals. The type will determine the quality and scaling behaviour in the construction of the Hamiltonian. Solutions to the Kohn-Sham equations within SIESTA are found either using standard matrix diagonalisation methods or the linear-scaling energy functional minimisation method. The solution methods will be dealt with in section 2.4.3.1 and later chapters.

SIESTA uses norm-conserving pseudopotentials [88] originally in a semi-local form where each angular momentum, l, has a different radial potential, $V_l(r)$. The pseudopotential will need to perform well in different environments and due to using norm-conserving pseudopotentials, each angular momentum channel will see a different potential. This leads to dividing the pseudopotential into local (long-range) and semi-local (short-range) components, which in turn makes the pseudopotential easier to handle. The local component is chosen as to have correct asymptotic (i.e. Coulombic) behaviour and the semi-local component is projected out onto each angular momentum channel from the pseudo wavefunctions, to handle the angular momentum dependent potentials. With this in mind, the semi-local form allows for a different pseudopotential for each angular momentum channel. The pseudopotentials can be generated to take into account scalar relativistic effects [89,90] and typically use the Troullier-Martins construction scheme [91]. The semi-local form is transformed to a fully nonlocal form by

the method proposed by Kleinman and Bylander (KB) [92];

$$\hat{V}^{PS} = V_{local}(r) + \hat{V}^{KB} \tag{2.77}$$

$$\hat{V}^{KB} = \sum_{l=0}^{l_{max}^{KB}} \sum_{m=-l}^{l} \sum_{n=1}^{N_l^{KB}} |\chi_{lmn}^{KB}\rangle \, v_{ln}^{KB} \, \langle \chi_{lmn}^{KB}| \tag{2.78}$$

$$v_{ln}^{KB} = \langle \phi_{ln} | \delta V_l(r) | \phi_{ln} \rangle \tag{2.79}$$

where $r = |\boldsymbol{r}|$. $V_{local}(r)$ is the local part of the pseudopotential and has an arbitrary form with the condition that it must connect with the semi-local potential, $V_l(r)$, where it becomes equal to the all-electron potential beyond a specified pseudopotential core radius, r_{core}:

$$\delta V_l(r) = V_l(r) - V_{local}(r) \tag{2.80}$$

This means that $\delta V_l(r)$ is equal to zero for $r > r_{core}$. Finally, the KB projectors $\chi_{lmn}^{KB}(\boldsymbol{r})$ are defined as;

$$\chi_{lmn}^{KB}(\boldsymbol{r}) = \chi_{ln}^{KB}(r) Y_{lm}(\hat{\boldsymbol{r}}) \tag{2.81}$$

where $\chi_{ln}^{KB}(r)$ are the KB projection functions (that are also zero for $r > r_{core}$), Y_{lm} is a spherical harmonic and $\hat{\boldsymbol{r}} = \boldsymbol{r}/r$ (unit vector). The eigenstates, ϕ_{ln}, are found by solving for the eigenfunctions of the semi-local pseudopotential with an all-electron Schrödinger equation which are then transformed using the orthogonalisation scheme proposed by Blöchl [93].

Solving the Kohn-Sham equations (2.50) can be made more tractable to be solved on a computer by expanding out the eigenfunctions, Ψ, with a set of basis functions. In the case of SIESTA, the expansion is a linear combination of (pseudo) atomic orbitals. The numerical pseudo atomic orbitals (PAOs) have the form;

$$\phi_{Ilmn}(\boldsymbol{r}) = \phi_{Iln}(\boldsymbol{r}_I) Y_{lm}(\hat{\boldsymbol{r}}_I) \tag{2.82}$$

where for atom, I, located at $\boldsymbol{R}_I$ with $\boldsymbol{r}_I = \boldsymbol{r} - \boldsymbol{R}_I$ the basis orbitals are a product of a numerical radial function, $\phi_{Iln}(\boldsymbol{r}_I)$, and a spherical harmonic, $Y_{lm}(\hat{\boldsymbol{r}})$.

The orbitals are strictly confined to a cutoff radius, r_l^c, where $\phi_{Ilmn}(\boldsymbol{r}) = 0$

when $r > r_l^c$. By setting the orbital values to zero beyond a confinement radius creates sparsity in the Hamiltonian and overlap matrices. This sparsity is a property which is crucial for the linear-scaling construction and storage of the Hamiltonian matrix, as well as other matrices dependent on the PAOs. More specifically the number of operations required for a particular orbital in the construction of the Hamiltonian matrix remains constant as the system size is increased unless orbitals are added within the vicinity, (r_l^c), of the orbital in question.

Generally there will be a number of orbitals (representing the principal quantum number), n, with an arbitrarily large number of angular momenta, l, and $(2l+1)$ magnetic quantum numbers, m. The multiple-ζ basis sizes used in quantum chemistry are created by using multiple orbitals with the same angular momentum dependence but with different radial dependence. The radial functions are defined by a cubic spline interpolation from values calculated on a logarithmic radial grid where the size and shape of the radial functions is arbitrary. SIESTA does provide default procedures to create these functions explained below.

For the generation of a single-ζ (SZ) basis set SIESTA uses the method of Sankey and Niklewski [84, 94]. Here the angular momentum dependent eigenfunctions, $\phi_{Ilmn}(r)$, of a pseudo atom in a spherical box potential including the atomic pseudopotential, $V_l(\boldsymbol{r})$ (radial potential), are used as the basis orbitals. The eigenfunctions are found by solving the following equation for an energy, $\epsilon_l + \delta\epsilon_l$, chosen so the first node occurs at a cut-off, r_l^c:

$$\left(-\frac{1}{2r}\frac{d^2}{dr^2}r + \frac{l(l+1)}{2r^2} + V_l(r)\right)\phi_{Ilmn}(r) = (\epsilon_l + \delta\epsilon_l)\phi_{Ilmn}(r) \qquad (2.83)$$

It is desirable for the effects of the orbital confinement to be similar for each of the orbitals. To accomplish this, a common energy shift, $\delta\epsilon$, is used rather than a common confinement radius. The strictly localised orbitals found from the above-mentioned approach are confined by an infinite potential. This approach generates orbitals with a discontinuous derivative at the cut-off, r_l^c. Using smaller cut-off distances the discontinuity will become larger. The discontinuity can have an effect on the calculation of the forces and stresses. Use of a soft-confinement potential [3] for the pseudo atom Hamiltonian (2.83) will remove the discontinuity. The soft-confinement potential is zero in the core region, with no discontinuous

derivatives and diverges at r_l^c for strict confinement.

For the generation of multiple-ζ basis sets a *split-valence* method [5, 95] adapted to numerical PAOs [4] is used. In the standard split-valence scheme the higher-ζ orbitals are generated by splitting the slowest decaying Gaussian orbital from the combination of primitive Gaussian functions. Control of the orbital confinement through the common energy shift is not possible with a Gaussian orbital. Hence, numerical orbitals are preferred rather than Gaussian functions in SIESTA's *split-valence* scheme [2–4]. The numerical orbital reproduces the tail of the first-ζ numerical orbital, $\phi_l^{1\zeta}(r)$, at a given split radius, r_l^s, and changes to a polynomial inside the radius that smoothly goes to the origin, namely;

$$\phi_l^{2\zeta}(r) = \begin{cases} r^l(a_l - b_l r^2) & \text{if } r < r_l^s \\ \phi_l^{1\zeta}(r) & \text{if } r \geq r_l^s \end{cases} \tag{2.84}$$

where $\phi_l^{2\zeta}(r)$ is the second-ζ orbital and the coefficients, a_l and b_l, ensure continuity and a continuous derivative at r_l^s. Optimisation of r_l^s is achieved via fixing the norm of $\phi_l^{1\zeta}(r)$ beyond r_l^s. This fixed amount of norm is termed the *split norm* and is usually set at 0.15 for most systems [4]. To reduce the number of Hamiltonian elements and still keep variational freedom, the second-ζ orbital is subtracted from the first-ζ orbital, which ensures that the new orbital is zero beyond r_l^s. The newly found orbital, still in the same Hilbert space as the first-ζ orbital, is renormalised to take into the account the loss of norm beyond r_l^s. Higher ζ orbitals can be generated by repeating the procedure at higher shells.

For a higher quality result polarisation orbitals are also included. Polarisation orbitals are used to describe the deformation produced by bond formation. PAOs with a higher angular momentum, $l+1$ (or higher l for extra orbitals), can be used as the polarisation orbitals. However, these types of orbitals do not perform well as they are typically too extended and can even be unbound for large l. SIESTA creates polarisation orbitals by using a first-order perturbation calculation of an orbital with a small applied electric field [2,4]. They have the same cut-off radius as the orbitals from which they are constructed.

With the inclusion of the KB pseudopotentials the Kohn-Sham SIESTA Hamil-

tonian can be written as;

$$\hat{H} = \hat{T} + \sum_I V_I^{local}(\boldsymbol{r}) + \sum_I \hat{V}_I^{KB} + V^H(\boldsymbol{r}) + V_{xc}(\boldsymbol{r}) \tag{2.85}$$

where $\hat{T}$ is the kinetic energy operator, $V^H(\boldsymbol{r})$ and $V^{xc}(\boldsymbol{r})$ are the total Hartree and XC potentials and $V_I^{local}(\boldsymbol{r})$ and $\hat{V}_I^{KB}(\boldsymbol{r}, \boldsymbol{r'})$ and the local and non-local (KB) parts of the pseudopotential of the atom I.

Screening with a potential, V_I^{atom}, created from an atomic electron density, ρ_I^{atom}, can eliminate the long-range aspects of $V_I^{local}(\boldsymbol{r})$. The atomic densities are assembled by appropriately populating the set of valence basis functions with single atom charges. The new screened *neutral-atom* (NA) potential, $V_I^{NA} = V_I^{local} + V_I^{atom}$, has its long-range interactions removed and due to the confinement of the atomic orbitals, V_I^{NA} is also zero beyond r_l^c. Defining $\delta\rho(\boldsymbol{r})$ as the difference between the self-consistent electron density, $\rho(\boldsymbol{r})$, and the sum of atomic densities, $\rho^{atom} = \sum_I \rho_I^{atom}$, and $\delta V^H(\boldsymbol{r})$ as the electrostatic potential generated by $\delta\rho(\boldsymbol{r})$ the final SIESTA Hamiltonian [2] can be written as:

$$\hat{H} = \hat{T} + \sum_I \hat{V}_I^{KB} + \sum_I V_I^{NA}(\boldsymbol{r}) + \delta V^H(\boldsymbol{r}) + V^{xc}(\boldsymbol{r}) \tag{2.86}$$

When constructing the above Hamiltonian, the first term is a two-centre integral which can be evaluated as a convolution. The term is transformed into Fourier space which is then treated as a simple product and tabulated as a function of interatomic distance. The second term is treated in a similar fashion as the first term, additionally with the local part of the pseudopotential, $V_{local}(r)$, found in $\hat{V}_I^{KB}$, integrated on a grid as it depends on the position of the atom with the pseudopotential as well. All other terms are calculated on a three-dimensional real-space integration grid. As mentioned earlier for complete details of the integration methods and for other details pertaining to SIESTA see references [2,4,5].

Solutions to the SIESTA Kohn-Sham equations can be found through either conventional matrix diagonalisation methods or via an energy functional minimisation approach. A summary of the energy functional minimisation method, as used in later chapters of this thesis, is provided in the following section.

2.4.3.1 Energy Functional Minimisation

The basis of energy functional minimisation methods is to reformulate the Kohn-Sham equations into a functional of an energy (e.g band structure or total energy). Use of the energy functional provides an avenue for the Kohn-Sham system to be solved in order-N, when the energy functional uses localised Wannier-type states. The energy functional is then minimised with respect to variations of the wavefunctions expanded in the LCAO basis, using a common optimisation technique such as the conjugate-gradient minimisation (CG) algorithm [96]. Currently within the SIESTA package there are two different energy functionals implemented. The first is a functional from Ordejón *et al* [6, 97, 98] which uses a fixed number of occupied states equal to the number of electron pairs. This functional has been found to have a large number of local minima [98, 99] leading to issues with the minimisation process (this does not occur with extended Bloch-like states). The second functional within SIESTA overcomes this problem. This functional by KMG [32] directly finds the ground state configuration when using localised wavefunctions, avoiding the multiple minima problem by using a larger number of states than electron pairs. Other advantages of the KMG functional are the decrease in the error of the estimate of the ground state energy, E_o, and improvement of the conservation of energy within MD [32]. Only the KMG functional is considered in this thesis due to the above mentioned issues with the Ordejón *et al* functional.

The derivation of the KMG functional [32] begins by starting with the energy functional defined by Mauri and Galli [99]. For an N-electron system, this functional depends on $N/2$ occupied orbitals. The Mauri and Galli functional is then generalised to account for an arbitrary number, M, of orbitals. There are no restrictions on M and in practice M is larger than the number of occupied states. The KMG functional is defined as;

$$\mathbf{E}[\{\phi\}, \eta, M] = 2 \sum_{ij=1}^{M} Q_{ij} \langle \phi_j | \hat{H} - \eta | \phi_i \rangle + \eta N \tag{2.87}$$

where $\{\phi\}$ is a set of M overlapping orbitals, η is a parameter representing the chemical potential of the electrons, $\hat{H}$ is the Hamiltonian, and $\mathbf{Q}$ is an efficiently

calculated truncated series expansion of the inverse of the overlap matrix $\mathbf{S}$;

$$\mathbf{Q} = \mathbf{S}^{-1} = [\mathbf{I} - (\mathbf{I} - \mathbf{S})]^{-1} \approx \sum_{n=0}^{k} (\mathbf{I} - \mathbf{S})^n \tag{2.88}$$

where the overlap matrix is defined as $S_{ij} = \langle \phi_i | \phi_j \rangle$ and $\mathbf{I}$ is the identity matrix. $\mathbf{Q}$ is used instead of $\mathbf{S}^{-1}$ because the explicit calculation of the $\mathbf{S}^{-1}$ occurs in $\mathrm{O}\left(N^3\right)$; conversely $\mathbf{Q}$ can be readily calculated in order-N when $\mathbf{S}$ is sparse. One other reason is that orbitals with a vanishing norm, $\langle \phi_i | \phi_i \rangle \to 0$, produce a non-zero contribution to the energy, since the eigenvalues of $\mathbf{S}^{-1}$ go to infinity [32]. Mauri *et al* [100] showed that an expansion to an odd number has the property of having a global minimum at a stable point. Hence, to guarantee finding a valid solution via standard minimisation techniques, an odd number expansion is required. What is chosen in practice is to use a first order expansion, as higher order expansions will lead to less sparse matrices which will produce higher scaling in the number of matrix operations. In the first order, $k = 1$, $\mathbf{Q}$ becomes:

$$\mathbf{Q} = 2\mathbf{I} - \mathbf{S} \tag{2.89}$$

Using $\mathbf{Q}$, the electron density is defined as:

$$\rho(\boldsymbol{r}) = 2 \sum_{ij=1}^{M} \langle \phi_j | \boldsymbol{r} \rangle \langle \boldsymbol{r} | \phi_i \rangle Q_{ij} \tag{2.90}$$

Equation (2.87) can now be defined in terms of a density matrix, $\hat{\sigma}\left[\{\phi\}\right] = \sum_{ij=1}^{M} |\phi_i\rangle \, Q_{ij} \, \langle \phi_j|$ (in the notation of [32]):

$$\mathbf{E}[\{\phi\}, \eta, M] = 2\mathrm{Tr}\left[\left(\hat{H} - \eta\right)\hat{\sigma}\right] + \eta N \tag{2.91}$$

Minimising the energy functional (2.91) with respect to the variation of the wavefunctions, and for a given η, will produce the orthonormal Kohn-Sham solution. The minimisation itself is unconstrained as there is not requirement of orthonormality of the wavefunctions. Instead, the energy functional itself will penalise any nonorthogonality and the final solutions will become orthonormalised. Issues can

arise with the electron number not being conserved. Within a few iterations the wave functions will become more orthonormal, which in turn stabilises the process.

In its original form the KMG functional is not solved in a linear-scaling fashion; the method achieves linear-scaling by using localised Wannier-like states. The Wannier-like states are constrained to an atom-centred spherical localisation region, where each Wannier-like state only interacts with its neighbouring Wannier-like states up to a specified radius. By using localised states, the determination of any one state only depends on the states surrounding environment and not the complete system. Obviously, larger localisation regions will increase the accuracy of the calculation and also the prefactor. Each atom, I, is assigned a number of Wannier-type states, N_{WS}:

$$N_{WS} = int\,(Z_I/2 + 1) \tag{2.92}$$

When doubly occupied, the total number of states can accommodate two extra electrons. These states can be partially filled allowing the excess state to be empty during the minimisation process. The accuracy of the calculation is controlled by the localisation region size of the Wannier-like states. For unconstrained states, the Kohn-Sham ground state energy is also the minimum of the KMG functional. By constraining the wavefunctions via localisation, the final energy will be higher than the unconstrained case, due to the orthogonality not being exact.

The major problem with this method is that an *a priori* knowledge of the chemical potential (i.e. Fermi level) is required. This introduces difficulties in estimating the chemical potential, where an educated guess will have to be made. For insulating and semiconducting systems the chemical potential can be located anywhere within the band gap. Although, even if the chemical potential is specified correctly at the beginning of the SCF calculation, due to the consequences of the mixing of the density matrices between SCF cycles the band gap of the system will shift. In some cases the shifting of the band gap will be large enough to cause the specified chemical potential to be invalid, causing the SCF calculation to diverge. These instabilities limit the use of this method to insulating and wide band gap semiconductors. Systems with small band gaps or no band

gaps (e.g. metals) will not work without an iterative refinement of η. The lack of k-point sampling hinders the use of the method for metals (and other periodic structures), where large supercells are required to capture the behaviour of the material at the gamma point. The decay length of metals is also much larger than finite band gap materials, requiring the use of larger localisation regions which in turn increase the prefactor of the method. Linear-scaling techniques requiring no *a priori* knowledge of any property are desirable.

2.5 Concluding Remarks

The theory of (TD)DFT and the SIESTA methodology has been presented within this chapter in a concise manner. Details on different basis set expansions show that PAOs are indeed an efficient choice compared to other atomic orbitals and planewaves. A linear-scaling energy functional method based on the KMG band structure energy [32] is presented. The SIESTA methodology is expanded upon in the next chapter, where details of a linear-scaling divide-and-conquer (D&C) implementation are provided for the solution of the Kohn-Sham equations. The D&C method is shown to complement the linear-scaling aspects of the SIESTA Hamiltonian assembly and the localised nature of PAOs.

divide et impera.

- An ancient Roman

3

Divide-and-Conquer Electronic Structure Theory

The linear-scaling density matrix D&C method has been implemented within the SIESTA package. When the D&C method is coupled with SIESTA's linear combination of pseudo numerical atomic orbitals, the results suggest that the D&C method can prove to be a very efficient first principles quantum mechanical method. The implementation is tested on a range of systems with varying band gaps, with comparisons being presented to the Kim-Mauri-Galli (KMG) energy functional minimisation method. The performance of the parallel computation is investigated. A first attempt at combining the D&C method with the KMG energy functional minimisation method to produce a more efficient linear-scaling method is shown.

3.1 Introduction

Electronic structure calculations, based on first principles quantum mechanics, provide reliable physical and chemical descriptions of atomistic, molecular and crystalline systems. However, practical calculations are often limited to fairly small systems (< 500 atoms) due to both theoretical difficulties and limitations in available computational resources. The theoretical difficulties arise from the high order, $O(N^3)$ and greater, scaling which is inherit within all *ab initio* quantum mechanical methods in the absence of approximations, where N is a measure

of the system size and usually most critically depends on the number of basis functions. To date, DFT [15] has proven to be a reliable and efficient choice in the study of medium-sized quantum systems. Only approximations, such as the Born-Oppenheimer approximation and the XC functional in Kohn-Sham theory, lead to deviation from a numerically converged calculation of the ground state Schrödinger equation, however, often the accuracy of the calculated physical and chemical properties is sufficient for practical use. It is desirable to be able extend DFT to calculate the properties of very large systems. The work presented in this chapter will allow the electronic structure of systems consisting of tens of thousands atoms to be calculated using DFT.

Solution of the Kohn-Sham equations (2.50) consists of two key steps - the assembly of the Hamiltonian and the determination of the orthogonal Kohn-Sham eigenstates. In the worst case scenario, construction of the DFT Hamiltonian matrix can scale as $O(N^4)$ due to the Coulomb term, though the use of density fitting [17] in an auxiliary basis can reduce this to $O(N^3)$. Diagonalisation of the Hamiltonian matrix will similarly scale as $O(N^3)$. Thus the assembly and diagonalisation of the Hamiltonian matrix are considered to be the major bottlenecks of any conventional implementation. Although DFT is considered relatively efficient it is still computationally prohibitive for the study of systems consisting of atom numbers in the thousands and greater. To overcome this barrier, techniques have been developed and employed to reduce the scaling of the computational cost to the linear regime, $O(N)$. In the same way, memory usage must also scale linearly, instead of as $O(N^2)$, in order to avoid another potential bottleneck.

Enforcing locality in all phases of the calculation is the key strategy to achieving complete linear-scaling. Using basis functions that are strictly local in real space allows the assembly of the Hamiltonian to be accomplished in a linear-scaling fashion. Even with localised basis functions, the long-range Coulomb energy term requires special consideration where it can be made to scale linearly by employing techniques such as fast multipole methods [101], or full multigrid methods [102]. The Hamiltonian matrix and the overlap matrix become sparse in a set of strictly localised basis functions, which allows the memory to scale linearly too when using sparse matrix techniques. In the present thesis, focus is

placed on the SIESTA methodology [2] to define the Hamiltonian and overlap matrices. Here norm-conserving pseudopotentials [88] are used to replace the core electrons and nuclei with a non-local potential, while the valence states are expanded in a set of PAOs [94]. A similar approach with the use of the PAOs can be found within the PLATO code [103]. Recapping section 2.4.3, the PAOs are the numerical solutions to the atomic pseudized problem, represented as a tabulation on a radial grid and multiplied by the appropriate spherical harmonic. The basis functions are made to be strictly confined by solving the atomic problem within a confining potential that becomes instantaneously or asymptotically infinite at a given radius [3]. By localising the basis functions, an approximation is made within the basis set, differing from methods where the Hamiltonian is directly made sparse through thresholding of integrals involving infinitely ranged basis functions [104].

With the assembly of the Hamiltonian computed in linear-scaling time and memory, the second key step of diagonalisation of the Hamiltonian matrix will have to be performed in a linear-scaling fashion. It is necessary to replace the standard cubic-scaling matrix diagonalisation with an approach to obtaining the self-consistent density that enforces localised solutions without explicit calculation of all Kohn-Sham states. This exploits the fact that the density is known to decay exponentially in materials with a band gap, while metals exhibit an algebraic power-law decay [19]. One of the first linear-scaling methods to be proposed in this context for DFT was the D&C approach, proposed by Yang in 1991 [21,22] and then subsequently reformulated for use within the density matrix framework in 1995 [20]. This method reduces the $O(N^3)$ scaling inherit with the diagonalisation of the Hamiltonian matrix to the linear-scaling regime by using partition functions to subdivide the electron density of the complete system. Each subsystem is then solved separately and the electron charge density of each subsystem is found. The sum of the corresponding contributions from all subsystems is used to obtain the total electron density and the energy of the system. This is possible due to the fact that the electron density is a local property within DFT.

Following the proposal of the D&C approach, there was extensive interest in

other linear-scaling approaches within the field[a]. This included methods based on functional orbital minimisation with respect to localised Kohn-Sham states [100], while avoiding explicit orthogonalisation, and techniques that operate directly on the density matrix with sparsity imposed [105–107], namely density matrix minimisation (DMM). Codes which employ DMM include ONETEP [78] and CONQUEST [37–39]. Another linear-scaling code is OPENMX [108] which uses either a linear-scaling Krylov-subspace method [109] or the D&C method with the SIESTA PAO basis set.

Although being one of the earliest so-called order-N methods, D&C has been relatively neglected until recently [23] within the condensed matter physics field, though it has found significant use within the quantum chemistry community due to the greater focus on localised basis sets and semi-empirical QM methods [24–26]. A few researchers have extended the D&C method to handle large-scale MD simulations using the FDM approach [27, 33] and to solid state systems [110, 111]. An implementation of a FDM method can found in section 4.2. Warschkow implemented a version to work with discrete variational DFT [112]. Pan *et al* [113] have also extended implementation to work in parallel. Another early implementation can be found in the DeFT software package implemented by Shaw and St-Amant [114]. Recently, the group of Vashishta and Shimojo used their hierarchical real-space D&C scheme [23] for large scale MD simulations [115–119], validating the use of D&C for possible calculations up to millions of atoms. Another recent investigation of the D&C method was made by Zhao *et al* [120], where they altered the standard method by using so-called positive and negative fragments instead of spatial partition functions, which when combined in a specific way can cancel out artificial boundary effects. In parallel to the first principles community, there is a large push to use the D&C method by the Hartree-Fock and post-Hartree-Fock community [28–31, 121–126].

The relative simplicity and robust nature of D&C with respect to the size and position of the band gap, combined with the successful implementations of various groups, suggests a reevaluation of the method is merited.

What is proposed for this study, is combining the fast Hamiltonian assem-

[a]For a more in depth analysis of the field please refer to the thorough review on the matter written by Goedecker [19] and for a review of the CONQUEST code see Bowler *et al* [38].

bly provided by the SIESTA package, through the use of localised PAOs, with an efficient, parallel implementation of the D&C method. This chapter contains a thorough explanation of the D&C method and the details of the current implementation. Tests are carried out on an insulating, semi-conducting and (near)metallic system, where comparisons to the KMG orbital minimisation method are made. The parallelisation scheme is tested along with a first attempt at combining the D&C method with the KMG orbital minimisation method [32].

3.2 Divide-and-Conquer Overview

The D&C scheme is related to the principle that the electronic structure for a particular region of a quantum system, to a good approximation, only depends significantly on the external potential due to nearby atoms, while those further away are rapidly screened with increasing distance. This principle was formalised and coined "near-sightedness" by Kohn [18]. The D&C method, first proposed by Yang [21, 22], was arguably the first practical linear-scaling scheme for first principles methods and while it precedes the work of Kohn, it builds on the prior knowledge of localisation through construction of Wannier functions [127, 128].

The D&C method involves dividing a system into a set of smaller overlapping subsystems. The speedup in calculation time occurs because each subsystem is solved separately with a cost that no longer depends on the size of the global problem. The individual subsystems are coupled to each other by a common Fermi level allowing electrons to flow until equilibrium is achieved. The electronic information obtained for each subsystem is then combined in a specific way so as to provide an approximation to the global (complete system) density matrix.

The present implementation treats each subsystem as consisting of a core region that is surrounded by a buffer region, as per the original work of Yang [21]. The atom(s) found in the core region are those whose localised electronic states are to be determined, while the atoms within the buffer region are required to correctly describe the electronic states of the core atoms within the local subsystem. Figure 3.1 shows an example of a possible subsystem that could be used for a graphene sheet. The atoms within the inner ring (red ring) are

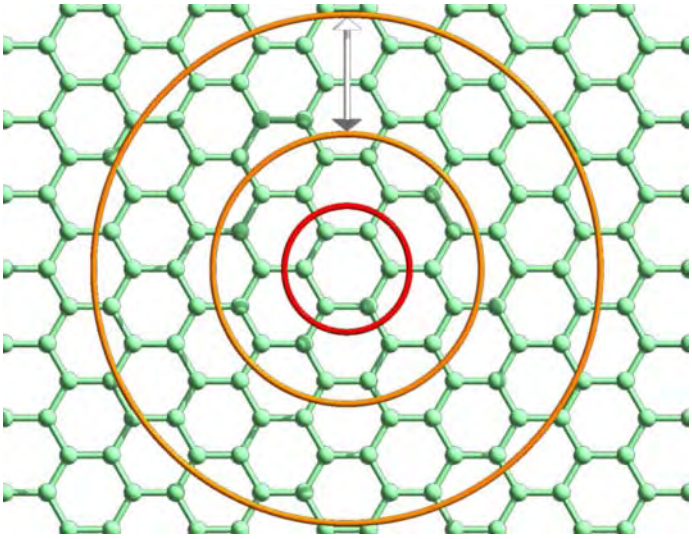


Figure 3.1: Core and buffer regions within D&C for an infinite graphene sheet. In this case the core region is indicated with the inner red circle. The buffer region, indicated with the outer orange ring, can be increased in size to provide a more accurate calculation.

designated as the core region while the atoms in between the inner and first outer ring (orange ring) are designated as the buffer region. By increasing the radius of the buffer region (the outermost ring) a more accurate computation is achieved at the cost of computational time. For the purposes of the present chapter, the majority of the calculations will focus on the situation where the core region holds one atom, while the buffer region can include as many atoms as required. Each atom in the system will become a core atom of a single subsystem. The size of the buffer region depends on the decay length (between any elements in the density matrix) within the material of interest and controls the degree of deviation from the unconfined Kohn-Sham solutions. Within the SIESTA methodology, an initial guideline as to the radius needed is given by the distance at which the Hamiltonian matrix elements go exactly to zero[b]. However, the buffer size may need to exceed this distance since the density matrix will usually decay at a slower rate than the Hamiltonian. Despite this, it is found that using smaller buffer radii than the Hamiltonian cut-off can also produce reasonable qualitative results for certain systems, as will be shown in section 3.6.1.

[b]This distance will always be greater than the equivalent distance for the overlap matrix as a

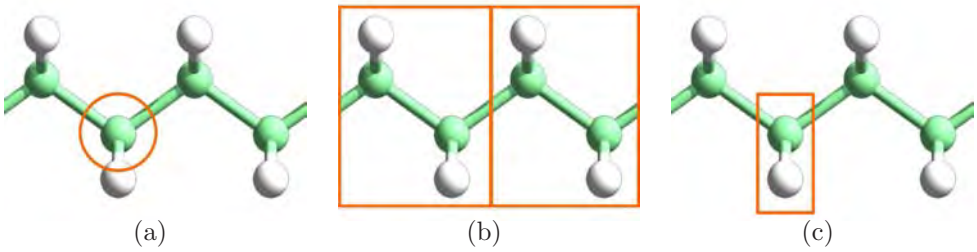


Figure 3.2: Different types of subsystems. a) Subsystem with a single core atom. b) Subsystem with the core atoms as a functional group of the system. c) Subsystem merging the hydrogen atom with the heavier carbon atom for this particular system.

Although, the present focus is on the situation where there is a subsystem centred on each individual atom, this need not be the case. For example, where atoms are closely linked, such as in a functional group or small covalent molecule, this entity could be treated with a single subsystem. Figure 3.2 shows possible choices one can make for the number of core atoms within a subsystem. The benefit of this is that the computational cost is lowered by a factor related to the number of core atoms per subsystem. In the limit where serial diagonalisation dominates, the cost will be reduced by the third power of the number of atoms combined per core (assuming all have the same number of basis functions per atom). The disadvantage is that in a system with an evolving geometric structure there is greater risk of discontinuities in the potential energy surface should a functional group dissociate and the subsystems are dynamically updated. If the membership of the subsystems remains fixed then the quality of the electronic structure would be a non-uniform function of the nuclear configuration. Section 4.1 details the implementation of a switching function applied to each subsystem Hamiltonian that smooths the interactions between atoms in the central region of the subsystem and atoms near the boundary of the subsystem. The application of the switching function tends to reduce the effects of the energy discontinuities. Having a subsystem centred on each atom represents the conservative option that minimises such errors, at an increased computational cost.

consequence of the matrix elements arising from the non-local projectors of the pseudopotential.

3.3 Divide-and-Conquer Formulation

The original D&C formulation [21, 22] uses the electron density as the basic variable. With the electron density defined as;

$$\rho(\boldsymbol{r}) = 2\sum_{m}^{N/2} \psi_m^*(\boldsymbol{r})\psi_m(\boldsymbol{r}) \tag{3.1}$$

where m is summed over occupied states, $N/2$ (accounting for spin). The premise of the D&C method is to forego calculating the density from all $N/2$ orbital solutions, $\psi_m(\boldsymbol{r})$, of the Kohn-Sham equations (2.50). Instead, the electron density is computed directly from a local approximation.

Beginning with a reformulation of (3.1);

$$\rho(\boldsymbol{r}) = 2\langle \boldsymbol{r}|\eta(\epsilon_F - \hat{H})|\boldsymbol{r}\rangle \tag{3.2}$$

where $\eta(x)$ is the Heaviside step function ($\eta(x) = 1$ for $x > 0, \eta(x) = 0$ for $x \leq 0$), $\hat{H}$ is the Kohn-Sham Hamiltonian and ϵ_F is the Fermi level. Through the use of a smooth and normalised partition;

$$\sum_{\alpha} \mathbf{P}^{\alpha}(\boldsymbol{r}) = 1 \tag{3.3}$$

the system is physically divided into overlapping subsystems, α. The partition weight functions, $\mathbf{P}^{\alpha}(\boldsymbol{r})$, are large in the subspace of α and have small values in locations far from α [21]. Summing up all subsystem contributions, the global density can be constructed as;

$$\rho(\boldsymbol{r}) = 2\sum_{\alpha} \mathbf{P}^{\alpha}(\boldsymbol{r})\langle \boldsymbol{r}|\eta(\epsilon_F - \hat{H})|\boldsymbol{r}\rangle = \sum_{\alpha} \rho^{\alpha}(\boldsymbol{r}) \tag{3.4}$$

where the subsystem density matrix is $\rho^{\alpha}(\boldsymbol{r}) = \mathbf{P}^{\alpha}(\boldsymbol{r})\rho(\boldsymbol{r})$. The following local approximation is now introduced for $\rho^{\alpha}(\boldsymbol{r})$;

$$\tilde{\rho}^{\alpha}(\boldsymbol{r}) = 2\mathbf{P}^{\alpha}(\boldsymbol{r})\langle \boldsymbol{r}|f_{\beta}(\epsilon_F - \hat{H}^{\alpha})|\boldsymbol{r}\rangle \tag{3.5}$$

where f_β is the Fermi function ($f_\beta = [1 + \exp(-\beta x)]^{-1}$) approximating an occupation number, β is the inverse electronic temperature ($\beta = 1/k_B T$, where k_B is the Boltzmann constant and T is the temperature), ϵ_F is the Fermi level common to all subsystems and $\hat{H}^\alpha$ is the local approximation to the Kohn-Sham Hamiltonian for subsystem α. An expression for the direct calculation of the global density (from (3.5)) is given by:

$$\tilde{\rho}^\alpha(\boldsymbol{r}) = 2\mathbf{P}^\alpha(\boldsymbol{r}) \sum_m f_\beta(\epsilon_F - \epsilon_m^\alpha)|\psi_m^\alpha(\boldsymbol{r})|^2 \tag{3.6}$$

The eigenfunctions, ψ_m^α, are obtained within the LCAO expansion;

$$\psi_m^\alpha(\boldsymbol{r}) = \sum_j C_{jm}^\alpha \phi_j^\alpha(\boldsymbol{r}) \tag{3.7}$$

and are the solutions to the following generalised eigenvalue equation:

$$(\mathbf{H}^\alpha - \epsilon_m^\alpha \mathbf{S}^\alpha)\mathbf{C}_m^\alpha = 0 \tag{3.8}$$

The subsystem Hamiltonian is defined as $\mathbf{H}_{ij}^\alpha = \langle \phi_i^\alpha | \hat{H} | \phi_j^\alpha \rangle$, the subsystem overlap matrix is defined as $\mathbf{S}_{ij}^\alpha = \langle \phi_i^\alpha | \phi_j^\alpha \rangle$ and the subsystem coefficient matrix is defined as $\mathbf{C}_m^\alpha$. The Fermi level, ϵ_F, is determined by the electron density normalisation condition:

$$N = \int \rho(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} = 2 \sum_\alpha \sum_m f_\beta(\epsilon_F - \epsilon_m^\alpha) \langle \psi_m^\alpha | \mathbf{P}^\alpha(\boldsymbol{r}) | \psi_m^\alpha \rangle \tag{3.9}$$

The D&C method can also be formulated using the density matrix [20]. With this version, the expensive integral calculations associated with the partition functions are avoided. The density matrix version is much more efficient. It is also applicable to other *ab initio* methods, such as the Hartree-Fock and semiempirical methods. The density matrix version is used in this thesis and its formulation follows.

3.4 Density Matrix Divide-and-Conquer Formulation

The formulation described here is based on the density matrix version of the D&C method [20]. Here, the density matrix is the primary entity in the formulation; the focus of D&C is to estimate the global density matrix from the sum of contributions from all subsystem density matrices.

Within D&C, the global density matrix is divided up into individual subsystem density-matrices weighted by a normalised partition function;

$$\sum_{\alpha} \mathbf{P}_{ij}^{\alpha} = 1 \tag{3.10}$$

where α is the subsystem index, i and j are orbital indices. The partition function, $\mathbf{P}_{ij}^{\alpha}$ is defined by a Mulliken-type [129] weight matrix (suitable for subsystems consisting of one core atom):

$$\mathbf{P}_{ij}^{\alpha} = \begin{cases} 1 & \text{if } i \in \alpha \textit{ and } j \in \alpha \\ 1/2 & \text{if } i \in \alpha \textit{ and } j \notin \alpha \\ 0 & \text{if } i \notin \alpha \textit{ and } j \notin \alpha \end{cases} \tag{3.11}$$

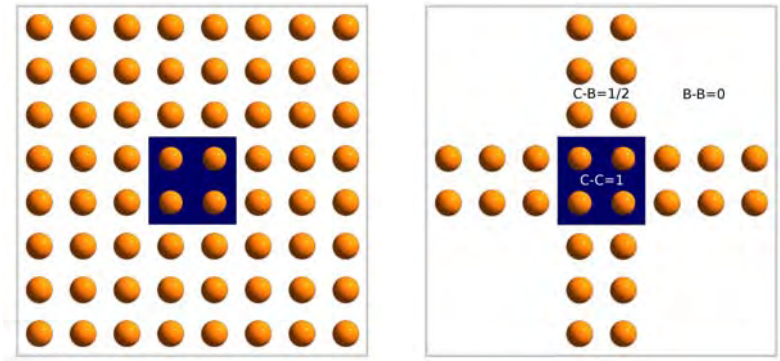
Defining the Kohn-Sham electron density;

$$\rho(\mathbf{r}, \mathbf{r}') = 2 \sum_{m}^{N/2} \psi_m(\mathbf{r}) \psi_m(\mathbf{r}') = \sum_{ij} \rho_{ij} \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \tag{3.12}$$

where electron density is defined in the space of the Kohn-Sham orbitals, $\{\psi_m(\mathbf{r})\}$. The density matrix, ρ_{ij}, is defined in the atomic orbital space, $\{\phi_i(\mathbf{r})\}$, and is given by the linear coefficients, $\{C_{im}\}$, as follows:

$$\rho_{ij} = 2 \sum_{m}^{N/2} C_{im} C_{jm} \tag{3.13}$$

The density matrix can be divided into subsystem contributions. The density matrix is then a sum of contributions from all subsystems, weighted by the partition



(a) (b)

Figure 3.3: The localisation of the density matrix specific to a subsystem. a) The subsystem density matrix. b) Subsystem density matrix elements contributing to the global density matrix. The partition function reduces the full subsystem density matrix to a cross-shaped arrangement of the elements. c-c represents the core-core orbital elements with a partition function weight of 1 (background shaded in blue). c-b represents the core-buffer orbital elements with a partition function weight of 0.5. b-b represents the buffer-buffer orbital elements with a partition function weight of 0. Because the matrix elements of the b-b section don't contribute to the global density matrix they are not drawn in.

matrix:

$$\rho_{ij} \equiv \sum_{\alpha} \mathbf{P}^{\alpha}_{ij} \rho_{ij} = \sum_{\alpha} \rho^{\alpha}_{ij} \tag{3.14}$$

The local nature of the density matrix allows each subsystem density matrix contribution to be approximated by;

$$\rho^{\alpha}_{ij} = 2\mathbf{P}^{\alpha}_{ij} \sum_{m} f_{\beta}(\epsilon_F - \epsilon^{\alpha}_{m}) C^{\alpha}_{im} C^{\alpha}_{jm} \tag{3.15}$$

where f_β is the Fermi function ($f_\beta = [1 + \exp(-\beta x)]^{-1}$) approximating an occupation number, β is the inverse electronic temperature ($\beta = 1/k_B T$, where k_B is the Boltzmann constant and T is the temperature), ϵ_F is the Fermi level common to all subsystems and ϵ^{α}_{m} is the orbital energy of the m th state in subsystem α.

The Fermi level needs to be found iteratively. This is achieved by using the midway energy between the maximum eigenvalue and minimum eigenvalue found from all subsystems as the initial guess. This Fermi level then is iteratively changed until it is narrowed down to a specified tolerance in the number of

electrons i.e. the Fermi level is determined by the normalisation of the correct number of electrons, N, given by:

$$N = \sum_{ij} \rho_{ij}\mathbf{S}_{ij} = \sum_{ij} \left(2 \sum_{\alpha} \mathbf{P}_{ij}^{\alpha} \sum_{m} f_{\beta}(\epsilon_F - \epsilon_m^{\alpha}) \times C_{im}^{\alpha} C_{jm}^{\alpha} \right) \mathbf{S}_{ij} \tag{3.16}$$

An example of subsystem density matrix in shown in Figure 3.3. The figure shows all the elements in the subsystem density matrix and also only the elements which will contribute to the global density matrix according to (3.11).

3.4.1 Density of States

As reported by Lee and Yang [130], the density of states (DOS) can be found using the D&C method by recognising that the DOS and is given by;

$$\text{DOS} = \frac{dN}{d\epsilon} \tag{3.17}$$

where N is the number of electrons and ϵ is the eigenstate energy. Using (3.16), equation (3.17) is expanded out as:

$$\frac{dN}{d\epsilon} = \sum_{ij} \left(2 \sum_{\alpha} \mathbf{P}_{ij}^{\alpha} \sum_{m} f_{\beta}^{'}(\epsilon_F - \epsilon_m^{\alpha}) \times C_{im}^{\alpha} C_{jm}^{\alpha} \right) \mathbf{S}_{ij} \tag{3.18}$$

The derivative of the Fermi function, $f_{\beta}^{'}$, is found by the chain rule:

$$f_{\beta}^{'}(\epsilon) = \frac{\beta e^{-\beta\epsilon}}{(1 + e^{-\beta\epsilon})^2} \tag{3.19}$$

The role of the Fermi function derivative is to extract out the eigenvalues from the possible spectrum. In practice, the Fermi function derivative is approximated by a Gaussian, eliminating any problems with states above the Fermi level in regards to the denominator. The β variable determines the width of the Gaussian, where the higher the value, the narrower the peak.

3.5 Implementation

In the present thesis the density matrix D&C scheme is combined with the SIESTA methodology [2] for the linear-scaling construction of the Hamiltonian and overlap matrices. Given the use of localised PAOs as basis functions within the SIESTA methodology, this is a natural combination to achieve full linear-scaling for large systems with relatively modest resources. The following sections contain a description of the key aspects of the present methodology.

3.5.1 Algorithm

Before explaining the algorithm of the current D&C implementation, it is worthwhile to describe the sparse memory model and the parallelisation scheme within the standard SIESTA package. Both features of SIESTA will be used within the D&C implementation.

Sparse arrays are arrays where the majority of the elements have a value of zero. The storage of the sparse arrays can be accomplished in order-N scaling with sparse-matrix memory models. Within the SIESTA methodology the overlap matrix, Hamiltonian matrix, and the density matrix are all sparse due to the strict confinement imposed on the PAOs. A naive approach would be to store these matrices in a 2-dimensional array, which would scale as $\mathrm{O}\left(N^2\right)$, where N is the number of orbitals, and any operations are performed in non-linear scaling. Using sparse-matrix techniques the storage and certain matrix operations can be accomplished in order-N scaling. The sparse-matrix format employed in SIESTA stores all non-zero elements of the sparse matrix in a 1-dimensional array, **A**. An integer array, **numA**, of size N, stores the number of non-zero elements of each row. An integer array, **listAptr**, of size N, stores the pointers to the start of each row in a 1-D packed array of non-zero values within **A**. The last integer array, **listA**, has the same size as **A** and stores the index pointers to each column. An example that prints each element of **A** is shown using pseudocode in Algorithm 1.

There are two parallelisation schemes within SIESTA. The first handles the distribution of dense matrices when conventional diagonalisation routines are

Algorithm 1 Print each element of sparse matrix, **A**

```
for io = 1 to N do
  for j = 1 to numA(io) do
    ind ← listAptr(io) + j
    jo ← listA(ind)
    print Element (jo, io) → A(ind)
  end for
end for
```

needed. Specifically, a block-cyclic orbital decomposition (either 1-D or 2-D) scheme is used to enable compatibility with the ScaLAPACK [131] parallel eigensolvers. When the order-N orbital minimisation approach in SIESTA is used, a uniform grid spatial decomposition algorithm is used to distribute the atoms amongst the compute nodes. The domain decomposition algorithm divides the unit cell into right-angled sections of side lengths as close to being equal while remaining commensurate with the lattice vectors. It then allocates each section with a non-zero atom count to a node. The allocation is conducted in a way so as to try to achieve a balanced number of atoms per node. This process could be further refined by accounting for the neighbour density in order to achieve improved load balancing. The contributions to the Hamiltonian, overlap and density matrices from each atom are then stored on the corresponding compute nodes. In short, each compute node will be responsible for a subset of orbitals localised in a region of space and all the corresponding electronic information pertaining to those orbitals. Each node then generates the elements of the Hamiltonian and overlap matrices that it is uniquely responsible for. The spatial decomposition algorithm is used as the preferred parallelisation scheme with the D&C implementation.

The general overview of the D&C implementation within the SIESTA code is shown in a flowchart in Figure 3.4. The flowchart has been appropriately marked to indicate which parts of the code involve the original SIESTA routines (solid box), parallel communication (dashed box) and the present D&C module (dotted line). The algorithm begins by reading the spatial locations of all atoms and options to perform the DFT run. Once the atom specifics have been read into SIESTA it will distribute the atom information across the compute nodes

according to the domain decomposition algorithm.

The D&C section of the code then begins from this point. If it is the first SCF cycle, the system will be divided into subsystems. This entails creating a list structure to store the orbital information for each subsystem with distinguishing lists for the core and buffer atoms. If running in parallel, the matrix elements belonging to buffer orbitals that reside on other compute nodes need to be communicated to the nodes with ownership of subsystems requiring that data. Because of the spatial locality of the domain decomposition, the number of compute nodes to be communicated with should remain constant or decrease as the system size increases, according to whether the number of processors employed scales with the system size or remains fixed, respectively.

The solution for the global density matrix proceeds by first solving the generalised eigenvalue problem for each subsystem, calculating the partition weights (equation 3.11) and other values that will benefit from caching. Once the eigenvalues of all subsystems are known, the Fermi level is found by iterative variation until equation 3.16 is satisfied to within a specified tolerance (e.g. $1\text{x}10^{-10}$). Having determined the Fermi level, the global density matrix is found by calculating the density matrices for each subsystem and then combining the contributions multiplied by the previously calculated partition weights.

3.5.2 Memory Considerations

When using D&C for large systems, the amount of memory used by the process must be manageable and scale linearly with system size. For D&C to be practical for very large systems only the information that is absolutely required should be stored. A large part of the task is already accomplished within SIESTA since all matrices that represent orbital based information (such as the Hamiltonian, overlap and density matrices) are stored in a sparse matrix representation as a 1-D array of non-zero valued elements. Because of the strict spatial locality of basis functions, the sparsity patterns for the Hamiltonian and overlap matrix are known *a priori* and fixed for any given nuclear configuration, while the density matrix is assumed to adopt the same sparsity pattern as the Hamiltonian. This use of sparse arrays ensures that the SIESTA methodology, by default, is linear-scaling

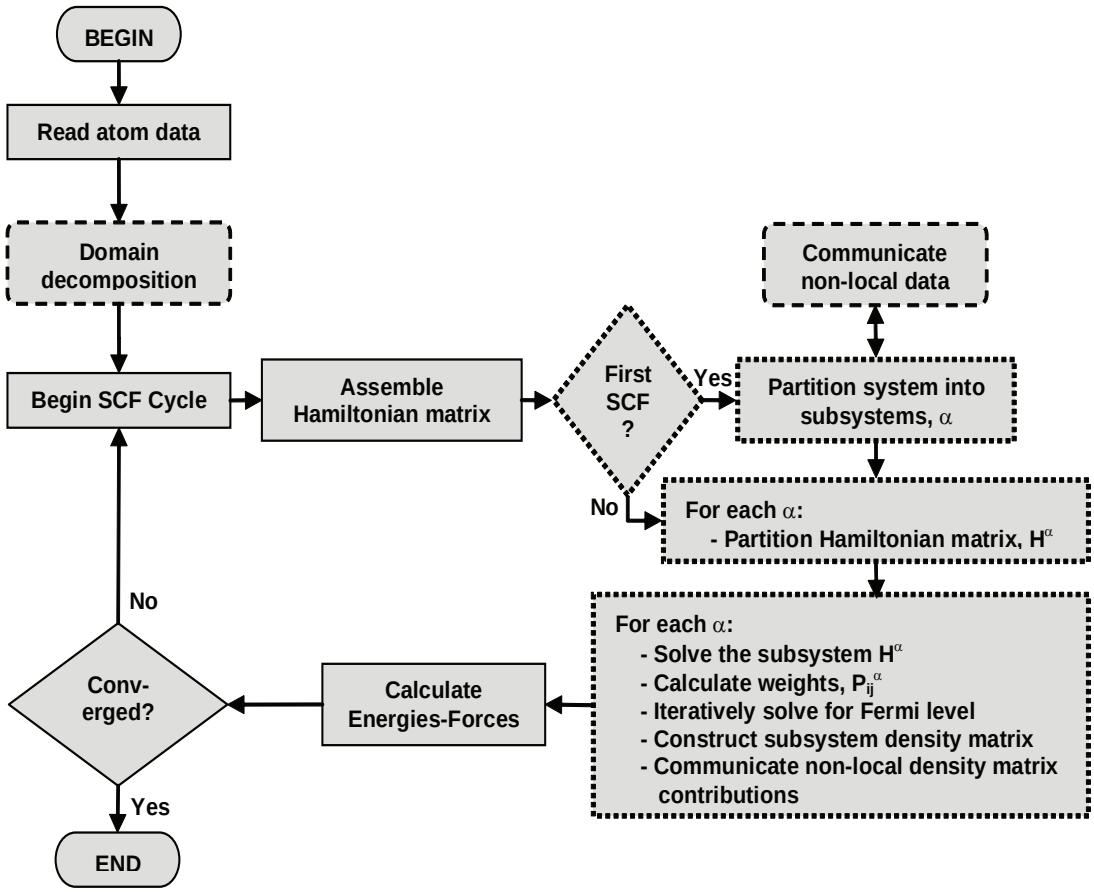


Figure 3.4: Schematic outlining the major implementation sections and process flow for the implementation of D&C within the SIESTA code. The original SIESTA routines are represented with solid boxes, the newly implemented D&C modules are shown in boxes with dotted lines and any parallel communication modules are shown in boxes with dashed lines.

in memory usage, except when diagonalisation is employed. Here dense matrix algebra is used locally for compatibility with standard eigensolution routines. Diagonalisation is typically used in cases where the system size is below the cross-over point at which linear-scaling becomes advantageous.

The D&C implementation, as has been described in section 3.5.1, can consume large amounts of memory for large systems. This is due to the fact that each subsystem must store 2-dimensional arrays for the subsystem Hamiltonian matrix, the subsystem overlap matrix, the subsystem eigenvector solutions and the subsystem density matrix. However, the subsystem Hamiltonian and overlap

matrices are not in use by the time it comes to constructing the density matrix, reducing the peak memory use. In the algorithm where the computational effort is minimised, the eigenstates of all subsystems must be stored simultaneously since they cannot be used in the construction of the local density matrix until the global Fermi level is known. When the number of subsystems is large and the subsystem sizes are considerable this can lead to a prohibitive amount of memory usage.

To overcome this issue, an alternate algorithm has been implemented that counters this problem, if so desired. It is accomplished by using a single allocation of memory for each matrix (Hamiltonian, overlap, eigenvectors and density matrix) that is large enough to store the information for the largest subsystem. That is, instead of storing matrices for each subsystem, only one set of matrices are stored and reused for each subsystem. This reduces the memory usage from $\sum_i^{N_p}(N_{orb}^H)^2+(N_{orb}^S)^2+(N_{orb}^{eig})^2$ to $(N_{maxorb}^H)^2+(N_{maxorb}^S)^2+(N_{maxorb}^{eig})^2$, where N_p is the number of subsystems, N_{orb}^H is the number of orbitals (basis functions) in the subsystem Hamiltonian matrix, N_{orb}^S is the number of orbitals in the subsystem overlap matrix, the N_{orb}^{eig} is the number of orbitals in the calculated subsystem eigenvectors; the subscript *maxorb* denotes the use of the maximum number of orbitals found within any of the subsystems. Using this memory conserving option leads to the memory usage scaling in a sub-linear fashion, but does increase the computing time required for each SCF iteration, since the subsystem Hamiltonian and overlap matrices will need to be diagonalised twice (the first time just requiring determination of the eigenvalues) if no caching of eigenvectors for later use can be performed. Depending on whether the calculation time is dominated by the diagonalisation step, this can have a significant influence on the time required for the SCF cycle. On average there is 50% increase in computing time and the worst case scenario will yield a doubling of the prefactor.

If memory usage is the key bottleneck, then it can be reduced to the absolute minimum required by computing all eigenvalues for the subsystems on the fly as required. Given that the eigenvalues are needed at each iteration of the Fermi level solution, this likely to make this algorithm uncompetitive as it would increase the prefactor by at least an order of magnitude, if not more. Memory

reduction can also be achieved by grouping atoms together to form subsystems (i.e. multiple core atoms per subsystem), since this reduces the total number of eigenstates to be stored by eliminating some duplication. An alternative and more viable approach would be to spread the information relating to each subsystem across many compute nodes (i.e. the diagonalisation of the subsystem Hamiltonian will occur on more than one compute node). This will help rectify the problem of memory usage on a symmetric multiprocessor machine, where each compute node has access to only a specific amount of memory.

3.5.3 Parallelisation

The parallel version makes use of the load balancing scheme included within the SIESTA package for the KMG order-N method, namely a domain decomposition algorithm to distribute the atoms amongst the compute nodes.

Because of the use of spatial locality during the parallel construction and solution for each subsystem, the only global communication occurs during the determination of the Fermi level. Here the eigenvalues and weights are stored on the node responsible for that particular subsystem. For every trial value of the chemical potential, the occupancy of each subsystem must be determined and a global summation performed to determine the total number of electrons before iteratively refining the Fermi level. Once the Fermi level is converged then each subsystem density matrix is calculated. The overall density matrix is then constructed through local communication between neighbouring nodes.

By taking advantage of symmetric multiprocessor machines, the diagonalisation of the subsystem Hamiltonians can be distributed across many compute nodes with the use of block-cyclic orbital decomposition for each subsystem. This approach will drastically reduce the prefactor of the D&C method and allow for calculations of very large systems, as has been accomplished in [23]. The memory required for the storage of a subsystem is distributed amongst many compute nodes reducing the overall memory per compute node.

3.6 Results

Calculations have been performed on a range of different systems in order to examine the performance of the present combination of the D&C implementation with the SIESTA methodology. The examples chosen include insulating, semi-conducting and near-metallic systems in order to demonstrate the varied application of D&C. The specific test cases are a linear alkane chain, C_nH_{2n+2}, for the insulating system, previously studied by Warschkow *et al* [112] using their D&C implementation, bulk silicon for the semi-conducting system, and a single walled (5,5) armchair carbon nanotube for the near-metallic system. The linear-scaling and the rate of convergence of the total energy to the Kohn-Sham energy when increasing the subsystem radius are studied. By increasing the subsystem radius, this implies increasing the number of buffer atoms in the buffer region. This is reported as an increase in the buffer region radius surrounding the core atom (subsystem centre). As with all tests in this study, each subsystem contains a single core atom surrounded by a buffer region. With this type of partitioning the number of subsystems equals the number of atoms within the system.

The scaling of the calculation time is shown by plots of the time required to complete the first SCF cycle and the section of the first SCF cycle only relevant to the D&C module. The first SCF cycle incorporates the building of the Hamiltonian and overlap matrices (handled by the SIESTA code) and the diagonalisation and building of the global density matrix (handled by the D&C module). For comparison, the performance of the KMG order-N solver already implemented within SIESTA is examined for the polymer and bulk silicon. Due to the inherent difficulties of achieving convergence when working at fixed chemical potential, the KMG algorithm was not examined for the near-metallic nanotube.

Calculations for semi-conducting bulk silicon were performed using the memory conservation scheme, as described in section 3.5.2. The remaining calculations were performed using the algorithm in which the eigenvectors for each subsystem are stored during the computation of the Fermi level.

Calculations were performed on a 32 processor SGI Altix machine (1.5 GHz) with 64 GB of RAM. All calculations were run on a single processor, except

those in Section 3.6.4 where the parallel performance of the code for a bulk silicon system consisting of 21,952 atoms is examined.

The systems used in the following calculations are all periodic either in 1-dimension or 3-dimensions. Order-N methods are designed to work with large systems. In supercell calculations the Order-N method will find the eigensolutions explicitly at the gamma point. Hence, when working with periodic systems it is important the supercell be large enough to effectively sample all the required k-points in the Brillouin zone. The size of the subsystems will play a factor when the subsystem size is larger than the supercell, in that the k-point sampling will improve as the subsystem size increases. One way to improve the sampling would be to use a partition function that is geared towards the periodic nature of the system, as found by Zhu *et al* [110].

3.6.1 Insulating System

The example of an insulating system studied here is the 1-D periodic linear alkane chain, C_nH_{2n+2}, where the number of formula units per unit cell, n, has been varied. This system should provide a favourable case for all linear-scaling methods as a closed-shell, wide gap, material with low dimensionality. The calculations were carried out using a 150 Rydberg cut-off for the real-space integration grid used to represent the density, an energy shift of 0.02 Rydberg for the PAO orbital confinement, and a density matrix convergence criteria of $1\text{x}10^{-4}$ for self-consistency. The Perdew-Burke-Ernzerhof (PBE) [54] form of the GGA was used for the XC functional.

The dependence of the D&C method on the basis set and the buffer region size is examined for various length alkane chains in Table 3.1. The table shows the energy difference per atom between the D&C calculated total energy and the conventional SIESTA calculated total energy, $(E_{dc} - E_{siesta})/n$, computed by diagonalisation. The errors found for all basis sets and buffer region sizes are relatively small. Given that the numbers quoted are the absolute differences in energy, any relative energies would exhibit even smaller discrepancies. Furthermore, even for the smallest buffer region size any error is likely to be small at the level of the accuracy of DFT. As the quality of the basis set is improved from

Table 3.1: Energy differences per formula unit (eV) between diagonalisation and D&C as a function of buffer region size and basis set for the C_nH_{2n+2} alkane chain.

Number of Atoms	Buffer Region (Å)	Basis Set			
		SZ[a]	SZP[b]	DZ[c]	DZP[d]
192	5.0	4.285E-03	2.705E-03	-1.661E-02	4.170E-03
	7.5	6.0765E-04	3.164E-04	-9.237E-04	8.031E-05
	10.0	-7.074E-07	6.057E-06	-4.656E-05	4.705E-05
384	5.0	4.288E-03	2.705E-03	-1.661E-02	4.167E-03
	7.5	6.076E-04	3.164E-04	-9.237E-04	8.030E-05
	10.0	-7.075E-07	6.063E-06	-4.656E-05	4.705E-05
768	5.0	4.286E-03	2.705E-03	-1.661E-02	5.258E-03
	7.5	6.074E-04	3.164E-04	-9.151E-04	1.026E-04
	10.0	-7.075E-07	6.061E-06	-4.656E-05	4.705E-05

[a] Single-zeta. [b] Single-zeta + polarisation. [c] Double-zeta. [d] Double-zeta + polarisation.

SZ to DZ, the discrepancy in the energy increases, while inclusion of polarisation functions actually leads to a reduction in error, at least for smaller buffer regions. While such variations will be sensitive to the details of the construction of the basis functions, such as the split-norm for radial degrees of freedom, the important conclusion is that there is unlikely to be a strong influence on the convergence behaviour of the D&C method.

As is to be expected, the errors decrease in size as the buffer region radius is increased. Table 3.1 shows that even a small buffer region radius of 5.0 Å is adequate for this system, regardless of basis set size, even though the buffer region is smaller than the maximum interaction range in the Hamiltonian of 7.3030 Å (for SZ) to 7.4416 Å (for DZP). The 0.2619 Å difference in the Hamiltonian interaction range is due to the extra KB projectors generated for the DZP basis set. The largest new KB projector for the DZP basis set has a radius of 1.1435 Å, which is 0.1310 Å larger than the largest KB projector found with the SZ basis set. When considering the Hamiltonian interaction range, the diameter is used so this value is doubled to give the difference in the ranges for both basis sets. The errors in the calculated forces are shown in Table 3.2. The errors in the

Table 3.2: Force differences per formula unit (eV/Å) between diagonalisation and D&C as a function of buffer region size and basis set for the C_nH_{2n+2} alkane chain.

Number of Atoms	Buffer Region (Å)	Basis Set			
		SZ[a]	SZP[b]	DZ[c]	DZP[d]
192	5.0	4.62E-02	-7.97E-03	-8.15E-02	-1.03E-01
	7.5	-1.24E-03	-1.77E-03	2.20E-03	-4.74E-03
	10.0	3.50E-05	5.00E-05	6.10E-05	-9.91E-04
384	5.0	4.67E-02	-8.00E-03	-8.15E-02	-1.03E-01
	7.5	-1.24E-03	-1.77E-03	2.20E-03	-4.74E-03
	10.0	3.50E-05	5.00E-05	6.10E-05	-9.91E-04
768	5.0	4.65E-02	-7.92E-03	-8.15E-02	-1.02E-01
	7.5	-2.62E-03	-1.59E-03	2.02E-03	-4.73E-03
	10.0	3.50E-05	5.00E-05	6.10E-05	-9.91E-04

[a] Single-zeta. [b] Single-zeta + polarisation. [c] Double-zeta. [d] Double-zeta + polarisation.

forces are larger than the total energy errors. As with the total energy errors, the errors in the force decrease as the buffer region is increased. The size of the errors for the 10.0 Å buffer region indicate that MD simulations are a possibility with the D&C scheme, as long as the buffer region is an adequate size. For geometry optimisation calculations, the error in the forces is expected to increase the number of steps required for convergence. With this in mind, and convergence criteria that is typically set at a maximum of 0.04 eV/Å, the error in the forces has to be reduced for efficiently found and accurate geometries. For the linear-alkane chain a buffer region of 7.5 Å is required for all basis set sizes, to efficiently be able to optimise the geometry at a 0.04eV/Å tolerance.

For comparison to the present D&C results, calculations were performed on this model system using the KMG order-N functional. The same localisation radius has been applied to the Wannier functions within the KMG approach as for the subsystem radius in the D&C technique. Consequently, both methods are attempting to find localised solutions with the same confinement constraint. The methods differ though in that the KMG approach contains a further approximation in that inverse of the overlap matrix is represented by a series expansion,

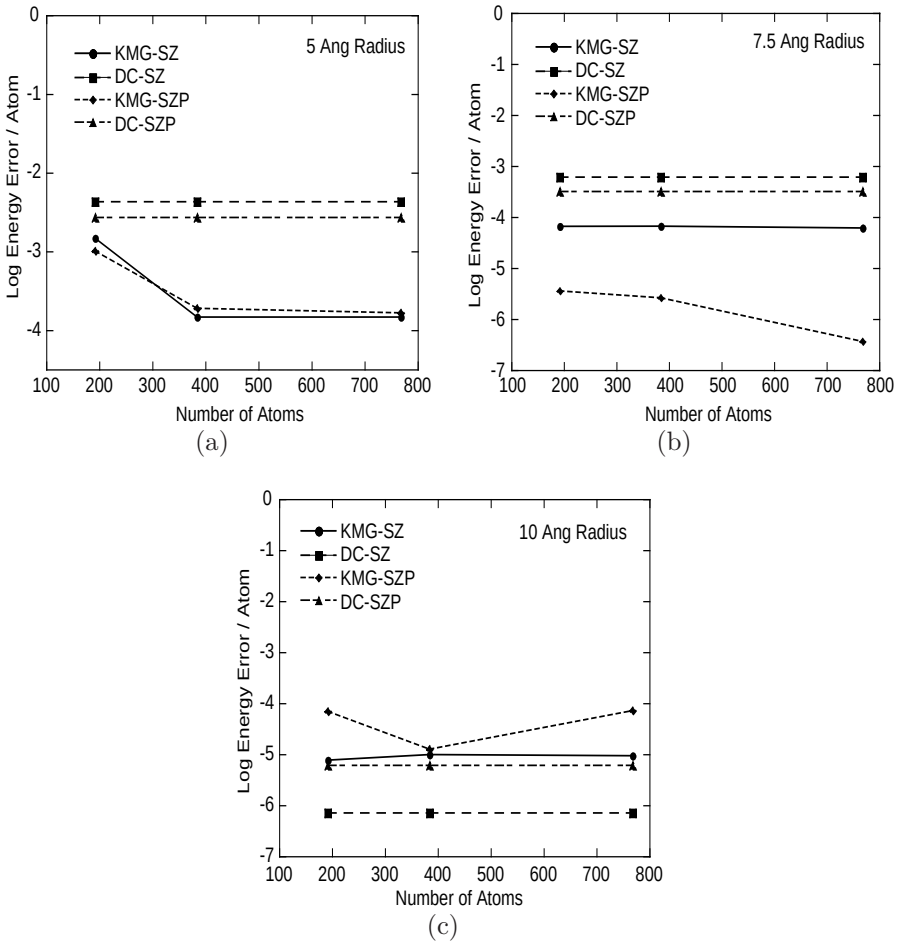


Figure 3.5: Comparisons of the errors per atom (eV) in the total energy between the D&C method and the KMG method for the C_nH_{2n+2} alkane chain with buffer radii. The D&C method exhibits a constant error as a function of the system size, while for the KMG method, the error becomes constant as the system size is increased. a) 5.0 Å buffer radius (subsystem for D&C and Wannier function radius for KMG). b) 7.5 Å buffer radius. c) 10.0 Å buffer radius.

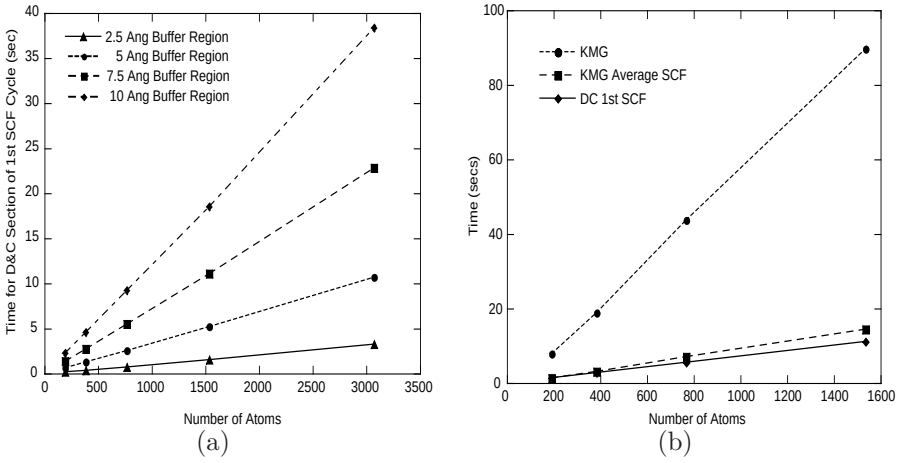


Figure 3.6: The CPU time scaling as a function of the number of atoms per supercell for a linear alkane chain, C_nH_{2n+2}. a) The D&C contribution to the first SCF iteration. b) A comparison between the KMG method and the D&C method for a buffer radius of 10.0 Å. The KMG method's first SCF and average SCF iteration calculation times are shown.

usually truncated at first order. The errors in the total energy relative to full diagonalisation are shown as their logarithms in Figure 3.5 for both KMG and D&C. For D&C the order of magnitude of the error is relatively constant as a function of increasing system size, while that for KMG decreases. This behaviour is likely to be, at least in part, a consequence of the increased sparsity of the overlap matrix leading to the additional approximation within the KMG scheme improving. Interestingly, for the smaller radii of confinement for the eigenstates the KMG yields a lower error in the total energy than the D&C scheme, which is somewhat unexpected, though the situation reverses for a radius of 10.0 Å. The error in the 10.0 Å KMG SZP calculation is larger than the error for the 7.5 Å KMG SZP calculation. This can be explained due to an oscillatory behaviour of the total energy convergence as the localisation region is increased.

The scaling of the calculation time of SZ basis set calculations for increasing supercell dimensions of the C_nH_{2n+2} alkane chain is shown in Figure 3.6(a). The graph shows the timing contribution of the D&C module section to the first SCF cycle. The graph clearly exhibits linear-scaling of the calculation time as the system size is increased for all buffer region sizes (i.e. the diagonalisation of the Hamiltonian matrix and the assembly of the global density matrix are all linear-

scaling processes). Although not shown here, the scaling is found to be linear regardless of basis set size, as expected. It is also possible to analyse the prefactor associated with the buffer region radius for this simple case. For radii of 5.0 Å, 7.5 Å, and 10.0 Å, the number of orbitals within the subsystem centred on a carbon atom is 42, 66 and 90, respectively, for a single-zeta basis set. When the slopes of the lines in Figure 3.6 are compared against these numbers, it appears that the prefactor scales approximately as the second power of the number of orbitals in the subsystem, as opposed to the theoretical maximum of a cubic scaling. Figure 3.6(b) shows a comparison of calculation time with the KMG order-N method. A direct comparison is not appropriate in this case as the KMG method generally has differing times for each SCF iteration, due to the CG minimisation process involved in the method (see section 2.4.3.1). As the LWFs are transformed to the form of the final states at each SCF iteration, it requires less computational effort for the minimisation process. In general, the first few SCF iterations take the longest time and as the calculation progresses through the SCF steps the number of iterations in the CG optimisation decreases. Figure 3.6(b) displays the timings for the contribution to the KMG order-N method for the first SCF iteration and the average time for all SCF iterations compared with the calculation time for the D&C section of the first SCF cycle. The prefactor of the first SCF time for KMG is much larger than the prefactor of the D&C method for the average SCF time[c]. The average SCF times of both methods are more in line with each other; where the higher number of SCF iterations required (in this case) for the KMG method and the fact that each SCF iteration takes less time has reduced the gradient of the average SCF time.

3.6.2 Semiconducting System

Bulk silicon has been chosen as the test case for the semiconducting system, having been previously widely studied using linear-scaling methods. The calculation was performed using a 40 Rydberg cut-off for the real space integration grid used to represent the density, an energy shift of 0.01 Rydberg for the PAO

[c]This comparison is possible as the first SCF time should be very close to the average SCF time, due to the constant SCF times for the D&C method

Table 3.3: Energy differences (eV/atom) between D&C and diagonalisation for a bulk silicon supercell consisting of 512 atoms as a function of buffer radius and basis set size.

Number of Atoms	Buffer Region (Å)	Basis Set			
		SZ[a]	SZP[b]	DZ[c]	DZP[d]
512	6.0	-4.879E-02	9.306E-03	5.570E-02	-7.512E-02
	7.0	1.751E-02	-9.124E-03	9.001E-02	-2.960E-02
	8.0	1.320E-02	-4.685E-03	3.115E-02	-1.346E-01

[a] Single-zeta. [b] Single-zeta + polarisation. [c] Double-zeta. [d] Double-zeta + polarisation.

orbital confinement, and a density matrix convergence criteria of $1x10^{-3}$. The maximum interaction range within the Hamiltonian matrix is 9.3843 Å for both the SZ and DZP basis sets. Again the PBE functional was used for the XC energy and potential. As in the insulating case, the energy difference per atom between the D&C total energy and that obtained via full system diagonalisation is calculated, see Table 3.3, as a function of basis set and buffer region size for a supercell consisting of 512 atoms.

As before, a small dependence was found on the basis set used and that by increasing the subsystem size (i.e. the buffer region) the error in the total energy is reduced, with one exception discussed below. Due to the smaller band gap and higher dimensionality of this system, the errors in the total energy (for a given subsystem size) are larger than in the insulating polymer case. Consequently, larger buffer regions are required to capture the decay length of the eigenfunctions accurately. However, the use of subsystems shorter than the interaction range of the Hamiltonian is still acceptable for at least qualitative results. Because the sparsity pattern of the density matrix in SIESTA is determined by that of the Hamiltonian, the computational penalty for using a large buffer radius only becomes particularly pronounced once the Hamiltonian interaction length scale is exceeded.

There is one discrepancy in the results; for the 8.0 Å buffer region size and DZP basis set the error in the total energy, -1.34620E-01 eV, is larger than errors found for decreasing buffer region sizes. In changing the radius from 7 to 8 Å two

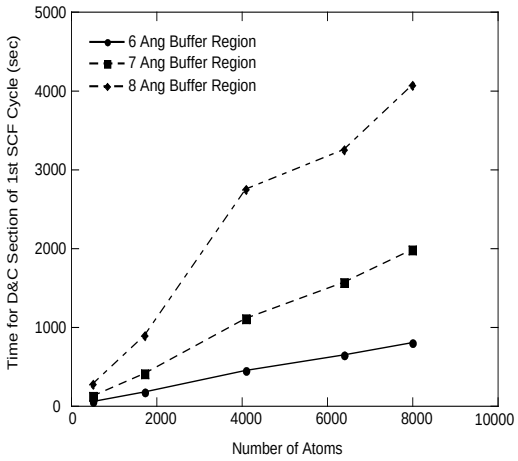


Figure 3.7: The CPU time scaling of a series of varying sized bulk silicon supercells. The contribution of the D&C section of the code to the first SCF iteration is shown.

extra shells of silicon atoms are included within the buffer region, comprising 28 atoms, as opposed to a single shell for the first transition. This demonstrates that the convergence with respect to buffer region is not guaranteed to be smooth and fluctuations are likely to be particularly pronounced when all atoms are symmetry equivalent due to the extent of mixing in the bands on the system.

The scaling performance of this system (with increasing atom numbers) is shown in Figure 3.7. The graph shows the calculation time for the D&C section of the first SCF cycle. The calculations examine the scaling from 512 atoms to 8,000 atoms using the SZ basis set. For the 6.0 Å buffer region size, linear-scaling is evident with increasing system size. The 7.0 Å buffer region size calculations show linear-scaling beyond 4,096 atoms, but deviate below this. This behaviour is even more evident with the 8.0 Å buffer region size calculations, where there is approximately $O(N^3)$ scaling for the system sizes examined up to 4,096 atoms and near linear-scaling for larger supercells. There is a discrepancy between 6,400 atoms and 8,000 atoms which is not currently resolved. It is assumed that it was due to hardware issues and not the D&C method itself, as there is no indication from the other results that linear-scaling should not occur. The absence of linear-scaling for small system sizes is due to the larger buffer region radii being greater than half the supercell length, based on a lattice constant of 5.43 Å for a single

unit cell of silicon. Within this regime, each subsystem includes nearly all the atoms of the supercell and so the cubic scaling of the diagonalisation for the subsystems dominates. Once the unit cell length becomes greater than the buffer region diameter there is a progressive transition to the expected linear-scaling until the crossover point is reached at which D&C becomes more efficient.

3.6.3 Near-Metallic System

This last test case was chosen to demonstrate the applicability of the D&C method for (near)metals. A near-metallic (5,5) armchair single walled carbon nanotube (SWNT) is chosen for this purpose. The calculations were performed using the PBE functional with a 100 Rydberg cut-off for the density integration mesh, 0.02 Rydberg for the PAO energy shift and a density matrix convergence criteria of $1x10^{-4}$. The resulting interaction ranges within the Hamiltonian vary from 7.3030 Å for the SZ basis set to 7.4416 Å for the DZP basis set. Once again, the difference in the interaction range is due to the extra KB projectors.

As in the previous two cases, calculations are performed of the variation of the error in the total energy with respect to different basis sets and buffer region sizes. The test system consisted of a 1,000 atoms within the one-dimensional supercell. The results are summarised in Table 3.4. The trends in the total energy with subsystem radius are less well defined for the present system, as would be expected due to the longer decay length. For the DZ and DZP basis sets the error does consistently decrease with increasing radius, though slowly, while for the SZ basis set the absolute magnitude decreases, but with the sign oscillating. For the SZP there is no apparent convergence within the range of radii examined and a more extensive exploration of larger radii is required. Despite the lack of a clear and rapid decay in error with radius, the magnitude of the difference from the full diagonalisation results, per atom, is comparable to that of thermal energy at ambient conditions and so higher levels of convergence may not be required for all calculations.

The error in the total energy increases as the basis set size becomes larger for any given subsystem radius. In regards to the basis set size, as the basis set size increases (improves), the band gap will generally decrease. The decrease

Table 3.4: Energy difference (eV/atom) between D&C and diagonalisation as a function of buffer region radius and basis set quality for a single walled (5,5) near-metallic carbon nanotube.

Number of Atoms	Buffer Region (Å)	Basis Set			
		SZ[a]	SZP[b]	DZ[c]	DZP[d]
1000	5.1121	1.194E-02	1.100E-03	-3.409E-02	-7.250E-02
	5.8424	-8.730E-03	-3.894E-03	-2.499E-02	-3.111E-02
	7.3030	2.272E-03	-1.335E-03	-1.315E-02	-1.225E-02

[a] Single-zeta. [b] Single-zeta + polarisation. [c] Double-zeta. [d] Double-zeta + polarisation.

in the size of the band gap is due to the bands becoming broader as the basis set improves, and hence the states tend to be delocalised. Yang *et al* [132] has found, for the case of a SWNT and a SWNT doped with adsorbed titanium chains, that the states near the Fermi level are more delocalised than states deeper in the valence band. The delocalised states near the Fermi level require larger subsystem sizes as the the basis set improves.

Figure 3.8 shows the scaling of the calculation times of the D&C section which contributes to the first SCF cycle with increasing system size. The SZ basis set was used for all the timing calculations. For all buffer region sizes the scaling is indeed found to be linear. To reduce the error in the total energy larger buffer region sizes are required. The timing results show that by increasing the buffer region slightly, as shown by the transition from a radius of 5.8 Å to 7.3 Å, this will increase the prefactor considerably. This requirement of a larger buffer region will inhibit the use of the D&C method for small metallic systems. The so-called crossover point, where it is computational beneficial to use the D&C method rather than conventional techniques, is pushed out to larger problems, which makes the use of the D&C method really only applicable to fairly large near-metallic systems. Using different partition schemes that produce smaller numbers of subsystems can help reduce the prefactor. Currently, there is a serial version of a partitioning scheme that allows for non-overlapping core regions with multiple atoms. This type of partitioning reduces the number of subsystems, but also increases the subsystem sizes. There should be a cross-over point when this

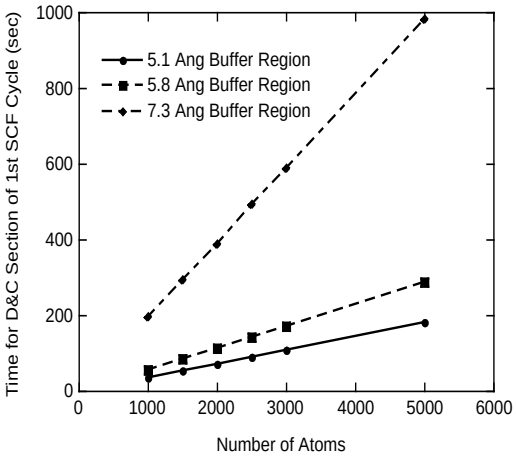


Figure 3.8: The CPU time scaling of a series of varying length (5,5) single walled carbon nanotube with a SZ basis set. The contribution of the D&C section implemented within the code to the first SCF iteration is shown.

type of partitioning is compared to the standard single core atom partitioning scheme. Further work is required to extend this to work in a parallel scheme.

3.6.4 Parallelisation

The parallel performance of the D&C implementation was tested on the bulk silicon system for a supercell containing 21,902 atoms. Using a SZ basis set, 40 Rydberg mesh cut-off for the integration grid, a PAO energy shift of 0.02 Rydberg and a buffer region radius of 6.08 Å, the test examined the parallel performance in going from 1 to 32 processors.

All calculations were executed using the memory conservation option (see section 3.5.2). Figure 3.9 shows that the speedup gained from using larger numbers of processors is nearly perfect relative to the calculation time for a single processor. For 32 processors, the speed up of 31.78 times is very close to the ideal value of 32. This indicates that the computational effort is indeed dominated by the diagonalisation of the subsystems, which is embarrassingly parallel, while the computation of the Fermi level and build of the Hamiltonian matrices, where communication is required, represents a small overhead. Similar results were obtained by Pan *et al* [113] with their parallel implementation of the D&C method.

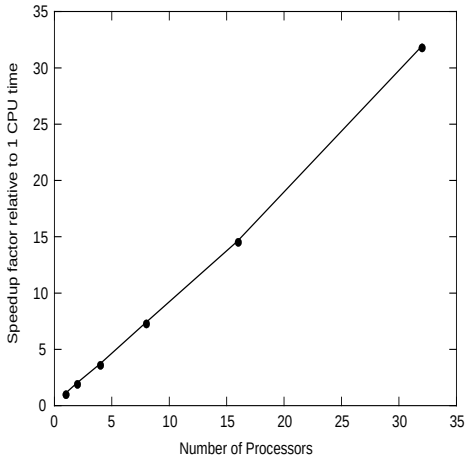


Figure 3.9: Parallel performance of the D&C implementation when studying a bulk silicon supercell containing 21,952 atoms with a SZ basis set. Shown here is the speedup when increasing the number of processors relative to a single processor calculation.

It should be noted that for this specific case the load balancing is perfect, i.e. in all cases each compute node has an equal number of subsystems of equal size due to the high symmetry of the problem. This is an important factor in contributing to the near perfect speedup. However, perfect load balancing will not always occur in practice with the present scheme for systems with inhomogeneous density or atom type distributions. Further refinement of the implementation is required to handle cases when the current distribution scheme does not perform well. The use of octrees [133, 134] or kd-trees [135–137], instead of a uniform grid, to distribute the atoms can improve the load-balancing; where the octree or Kd-tree can be balanced with knowledge of the size of the subsystems.

3.7 Hybrid Divide and Conquer - Orbital Minimisation Method

For all the benefits of ease of use, stability and no prior knowledge requirement of the chemical potential, the D&C method suffers from the problem of duplication of effort. That is, any given matrix element will be part of the Hamiltonian of multiple localised states, leading to the generation of many similar eigenstates

from different subsystems. The overlap of the subsystems leads to replication that increases the prefactor of the method and in turn, the cross-over point. On the contrary, orbital minimisation linear-scaling methods eliminate the duplication of effort present in the D&C method. In the current implementation of the SIESTA package, the KMG functional (2.91) [32] is the energy functional generally employed to determine the electronic states in a linear-scaling fashion. The KMG method is subject to difficulties of its own. Because the algorithm works at a constant chemical potential, rather than fixed number of electrons, it is necessary to *a priori* specify the Fermi level to lie within the band gap. If this is not the case, then the method diverges. For wide gap insulators this is rarely an issue since there is considerable margin for error when guessing the chemical potential to use, whereas for a semiconductor or small gap system it becomes a matter of trial and error. To complicate things further, the Fermi level is a function of the density matrix and therefore will change during the SCF iterations, leading to the potential need to adjust the chemical potential at each cycle during the early stages of SCF convergence.

For all the issues inherent with orbital minimisation methods, they do have one important benefit, in that the computational cost of the minimisation process lessens as the SCF procedure heads towards convergence. The calculation time to complete the minimisation within the first few SCF iterations is typically much greater than in the later stages. The D&C method, on the other hand, has a constant SCF calculation time. It would be beneficial to forego the unstable and lengthy early stages of the KMG method with an alternate faster and more stable method, but then take advantage of the final *faster* stages within the functional minimisation method. What is proposed here is to use the D&C method, with its constant SCF calculation time for the initial calculation method that will seed the KMG method with a Fermi level, density matrix and initial LWFs.

The proposed method is possible due to the localisation region scheme shared by both methods. The D&C subsystems correlate to the localisation regions used to describe the LWFs of the KMG method. The D&C subsystem eigensolutions are localised within the subsystem boundaries, making them a possible candidate to be transformed into a LWF. By supplying a Fermi level and a set of LWFs that

correspond to the supplied density matrix it is expected that the KMG method will be close to functional convergence and hence, the final SCF iterations using the KMG method will be faster than a D&C SCF iteration, reducing the prefactor of both methods. Unfortunately, the hybrid method (D&C-KMG), does not fulfil this aim and the following work is considered a first attempt at such a method. Further work is required to improve the method.

3.7.1 Divide and Conquer Seeding

The general procedure followed by the D&C-KMG method is to run a D&C calculation until the Fermi level is stable. For typical systems, this generally occurs between 2-6 SCF iterations. Once the Fermi level is stable, the information gained by the D&C is used to seed the KMG method in the next SCF iteration. For each atom the KMG method creates N_{WS} (from (2.92)) LWFs, where the number is dependent on the atomic species. The Mulliken-type partitioning used by the D&C implementation generally has $\geq N_{WS}$ eigensolutions for each atom centred subsystem. Here, the first N_{WS} subsystem eigensolutions are used to create the LWFs. For large basis set sizes, the number of subsystem states will always be larger than the required number of LWFs. It is not necessary to use the first N_{WS} eigenstates of the D&C subsystem, as it might be beneficial to use some linear combination or to specifically use the polarisation orbitals. Even though the KMG functional itself does not require orthonormal LWFs[d], an option to *further* orthonormalise the D&C subsystem eigensolutions using the modified Gram-Schmidt orthogonalisation algorithm [138] has been implemented to see if there is an improvement with the KMG minimisation process.

The modified Gram-Schmidt orthogonalisation process takes a set of k vectors, $\mathbf{S} = \{\mathbf{v}_1, \mathbf{v}_2 \dots \mathbf{v}_k\}$, and transforms them into an orthogonal set, $\mathbf{S}' = \{\mathbf{u}_1, \mathbf{u}_2 \dots \mathbf{u}_k\}$, that spans the same k-dimensional subset of the inner product

[d]The KMG functional penalises non-orthogonality, whereby the minimised states will eventually be orthogonalised.

space of **S**. The orthogonal vectors are found as follows;

$$\begin{aligned}
\mathbf{u}_k^{(1)} &= \mathbf{v}_k - \mathrm{proj}_{\mathbf{u}_1}(\mathbf{v}_k), \\
\mathbf{u}_k^{(2)} &= \mathbf{u}_k^{(1)} - \mathrm{proj}_{\mathbf{u}_2}(\mathbf{u}_k^{(1)}), \\
&\vdots \\
\mathbf{u}_k^{(k-2)} &= \mathbf{u}_k^{(k-3)} - \mathrm{proj}_{\mathbf{u}_{k-2}}(\mathbf{u}_k^{(k-3)}), \\
\mathbf{u}_k^{(k-1)} &= \mathbf{u}_k^{(k-2)} - \mathrm{proj}_{\mathbf{u}_{k-1}}(\mathbf{u}_k^{(k-2)}) \\
\mathbf{u}_k &= \mathbf{u}_k^{(k-1)} - \mathrm{proj}_{\mathbf{u}_k}(\mathbf{u}_k^{(k-1)})
\end{aligned}$$

where the projection operator, $\mathrm{proj}_{\mathbf{u}}$, is defined as:

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\langle \mathbf{v}|\mathbf{u}\rangle}{\langle \mathbf{u}|\mathbf{u}\rangle}\mathbf{u}$$

Unfortunately, the hybrid method (D&C-KMG), does not successfully function as intended and the following work is considered a first attempt at such a method. Further work is required to improve the method.

3.7.2 Results

Calculations have been performed in order to examine the performance of the present implementation of the D&C-KMG method. All calculations were performed using a single processor of a Intel Centrino 2 duo core system (2.0 GHz) with 2 GB of RAM. The code was compiled with no compiler optimisations to ensure that any timing information is purely based on the actual implementation of the method.

Firstly, the convergence behaviour of the D&C and KMG method is examined for a CO_2 molecule. This will give an indication of convergence issues that will play a part in determining the efficiency of the hybrid method. Table 3.5 shows a summary of the number of SCF iterations required for the CO_2 density matrix to converge to a tolerance of $1x10^{-4}$ for varying configurations based on the real-space integration grid fineness, Broyden mixing weight and initial Fermi level specification for the KMG method. For both the KMG and D&C method the

Table 3.5: CO_2 molecule self-consistency dependence on initial Fermi level, mixing weight and integration grid resolution. The number of SCF cycles required for convergence is shown for the KMG method and D&C method. The KMG method seems to be sensitive to these conditions. The D&C method has a constant convergence rate. It must be noted that for larger systems the results will tend to vary, generally with larger numbers of SCF iterations.

Broyden Mixing Weight	KMG Method with Initial Fermi Level						D&C
	-3 eV	-4 eV	-5 eV	-6 eV	-7 eV	-8 eV	
350 Rydberg Integration Grid							
0.1	58	79	23	21	14	16	10
0.15	16	36	38	24	24	42	10
0.2	16	33	62	10	13	11	10
500 Rydberg Integration Grid							
0.1	DNC[a]	23	18	81	16	32	10
0.15	28	38	21	20	51	55	10
0.2	16	27	11	10	12	11	10

[a] DNC - did not converge.

localisation regions encompass the complete system.

In this case, the D&C method converges in 10 SCF iterations for all mixing weights and integration grids. Hence, there isn't a dependence on the quality of the calculation. This is not necessarily true for larger systems, as the quality of the calculation and the mixing weight will play a large role in the convergence rate. On the other hand, even for this small system, the KMG method is dependent on the mixing weight and the fineness of the integration grid. With the integration computed on a real-space grid, the symmetry of the system is broken. This in turn breaks symmetry in the orbital solutions and hence the density, leading to issues with SCF convergence. The issues manifest primarily as oscillations in the density during the mixing process, that in turn make it difficult to converge the density. The initial Fermi level guess also plays a role in the SCF convergence. Even though all initial Fermi levels are within the "band gap" of the system, there are only a few cases for when convergence is reached in 10 steps. For a Fermi level of -3 eV and a 500 Rydberg integration grid the KMG method failed to converge.

It must be noted, that high SCF iterations occur because of oscillations in the SCF procedure when near convergence. That is, the convergence behaviour is regular until near self-consistency, when the oscillations occur. For larger systems the convergence rate of the KMG method will play a major role in lowering the prefactor of the complete calculation time.

The convergence behaviour of the KMG and the D&C-KMG method for the CO_2 molecule with a 500 Rydberg integration grid is shown in Table 3.6. The KMG calculations are run with the Broyden mixing scheme at a mixing weight of 0.2 with an inclusion of a linear-mixing kick at every 6th SCF iteration for different initial Fermi levels. The linear-mixing kick was found to help with SCF convergence with both methods, especially for the KMG method which has much lower SCF iterations than without the linear-mixing kick (except for the KMG case at a -4 eV Fermi level). For the D&C-KMG method there are three separate calculations for a mixing weight of 0.2 and varying initial Fermi levels. The transition from the D&C method to the KMG method occurs on the 6th SCF iteration. The first set of D&C-KMG based calculations use only the calculated density matrix to seed the KMG method (labelled SDM), the second set uses the density matrix and the subsystem eigenfunction coefficients to seed the KMG method (labelled SLWF) and the third set of calculations uses the density matrix and orthonormalised subsystem eigenfunction coefficients to seed the KMG method (labelled SOLWF).

The number of SCF iterations for the D&C-KMG cases is generally lower than the KMG based method, with the smallest number of iterations occurring at 11 steps. The number of CG iterations required to minimise the KMG functional for the first two SCF steps, when the KMG method is running, will indicate if the seeding of the D&C KS orbitals benefit the KMG method. The best case is shown for the -7 eV Fermi level run where the KMG method requires 67 CG iterations. The SDM case requires only 33 iterations, the SLWF case requires 26 iterations and the SOLWF case requires 27 iterations. There is a similar decrease in the number of CG iterations for the second SCF step as well. For these settings, the seeding of the KMG method from the D&C method is proven to be beneficial in reducing the computational cost of the KMG method. In all other cases with

Table 3.6: Self-consistent field convergence of the CO_2 molecule. The number of SCF cycles required for convergence is shown for the standard KMG method and the hybrid D&C-KMG method with a localisation region and subsystem radius of 3.0 Å. The calculations are run at varying initial Fermi levels. The transition from the D&C method to the KMG method (for the D&C-KMG method) occurs on the *sixth* SCF iteration. Refer to text for meanings of the different schemes.

Scheme	SCF Transition	Mixing Kick	Initial Fermi Level (eV)						
			-3	-4	-5	-6	-7	-8	-7.8326[a]
Number of SCF iterations									
KMG	-	6	14	DNC[b]	14	13	18	18	18
SDM	6	6	11	11	17	11	15	11	DNC
SLWF	6	6	11	20	21	15	16	DNC	16
SOLWF	6	6	11	DNC	DNC	16	16	16	16
Number of CG iterations - 1st SCF iteration after transition									
KMG	-	6	21	>1000	44	37	67	34	14
SDM	6	6	58	37	33	34	33	54	CGD[c]
SLWF	6	6	106	39	28	30	26	CGD	25
SOLWF	6	6	102	37	33	37	27	40	36
Number of CG iterations - 2nd SCF iteration after transition									
KMG	-	6	16	14	15	17	41	13	9
KMG	-	6	7	7	6	7	10	6	CGD
SLWF	6	6	7	12	18	12	14	CGD	13
SOLWF	6	6	7	12	12	12	12	14	11
Total Calculation Time (s)									
KMG	-	6	156	DNC	156	151	195	194	197
SDM	6	6	126	128	184	128	209	136	DNC
SLWF	6	6	133	213	226	165	174	DNC	174
SOLWF	6	6	126	DNC	DNC	174	174	176	174

[a] The Fermi level found by the D&C method and used to seed the KMG based orbital minimisation method. [b] DNC - Did not converge. [c] CGD - Conjugate-gradients minimisation process diverged.

varying Fermi levels there are similar trends, expect for the case of the -3 eV Fermi level. Here, the number of CG iterations is actually much higher than the KMG method at 58 iterations for the SDM case, 106 iterations for both the SWLF and SOLWF cases compared to the 21 iterations of the KMG case.

The D&C-KMG method is now tested on a 98-atom linear-alkane chain with an integration grid of 150 Rydberg, using a SZ basis set with 0.02 Rydberg energy shift. The convergence of standard diagonalisation and the D&C method are shown in Table 3.7. For this larger system, the D&C method also exhibits the oscillatory behaviour in the SCF procedure as did the KMG method for the CO_2 molecule. To overcome the oscillatory behaviour, a linear-mixing kick is introduced into the SCF process. Where for every specified number of SCF iterations linear-mixing is performed instead of Broyden mixing. The standard diagonalisation converges in 7 SCF iterations for all mixing weights. For a subsystem radius of 7.5 Å the D&C method converges in 9 or 8 iterations depending on the mixing weight, this is reduced to 7 iterations when a linear-mixing kick on the sixth SCF iteration is applied. For the larger 10 Å subsystem radius, the convergence occurs at 7 iterations, the same as the standard diagonalisation. This is expected because as the subsystem radius is increased, the approximation in the D&C method is reduced. The total calculation time, total energy and the maximum constrained force are also listed for reference.

For the same 98-atom linear-alkane chain system as above, the next set of calculations as shown in Table 3.8 are carried out using the KMG and D&C-KMG methods, using a LWF localisation region and subsystem radius of 7.5 Å. The KMG calculations are run with the Broyden mixing scheme at a mixing weight of 0.2 and with a linear-mixing kick at every 4th SCF iteration for different initial Fermi levels. For the D&C-KMG method there are three separate calculations for each mixing weight and initial Fermi level. The transition from the D&C method to the KMG method occurs on the 4th SCF iteration. The first set of D&C-KMG based calculations use only the calculated density matrix to seed the KMG method (SDM), the second set uses the density matrix and the subsystem eigenfunction coefficients to seed the KMG method (SLWF) and the third set of calculations uses the density matrix and orthonormalised subsystem

Table 3.7: Self-consistent field convergence of the 98-atom linear-alkane chain system. The number of SCF cycles required for convergence is shown for standard diagonalisation and the D&C method with a subsystem radius of 7.5 Å and 10.0 Å. A linear-mixing kick has been introduced to help with convergence. The total energy is included for reference.

Mixing Weight	Mixing Kick	SCF Iterations	Total Time (s)	Total Energy (eV)
Conventional diagonalisation				
0.1	-	7	286	-5951.9807
0.15	-	7	282	-5951.9808
0.2	-	7	286	-5951.9799
D&C with 7.5 Å subsystem radius				
0.1	-	9	351	-5951.9493
0.15	-	9	350	-5951.9496
0.2	-	8	316	-5951.9494
0.1	6	9	351	-5951.9497
0.15	6	9	351	-5951.9498
0.2	6	7	284	-5951.9482
D&C with 10 Å subsystem radius				
0.1	-	7	301	-5951.9948
0.15	-	7	300	-5951.9949
0.2	-	7	302	-5951.9949

eigenfunction coefficients to seed the KMG method (SOLWF).

The KMG based methods all converge in 7 SCF iterations, except for the case with an initial Fermi level of -4 eV, which converges in 11 iterations. The linear-mixing kick at the fourth iteration helps quicken the convergence. The number of CG optimisation steps for the first SCF iteration after the transition ranges from 109 steps to 126 steps. The number of CG optimisations for the second SCF iteration after the transition ranges from 12 to 13 steps. For the D&C-KMG method to be successful, it must have approximately the same number of the SCF iterations and the number of CG steps must be lower than the KMG method. For the SDM case, the initial LWFs coefficients are randomly chosen. Hence, only the initial Hamiltonian is close to convergence.

For the initial Fermi levels of -5 eV, -6 eV and the D&C found Fermi level of -7.858 eV convergence occurs at 7 SCF iterations. The number of CG steps for the first and second transitioned SCF iteration are slightly lower than the KMG method, hence the total calculation time is slightly lower, for example the -6 eV Fermi level run finished in 313 secs while the KMG method for the same Fermi level finished in 318 secs. The D&C-KMG method in this case is faster than the KMG method. However, it is still slower than the 284 secs using the D&C method with a linear-mixing kick at the sixth SCF iteration. The reason is due to the number of CG steps in the first SCF iteration after the transition. The number of CG iterations has to be reduced to improve upon the D&C-KMG method efficiency.

The next set (SLWF) of calculations uses the density matrix and the D&C subsystem eigensolutions to seed the KMG functional. In this case the number of CG iterations increased dramatically, just below three times on average, accompanied with a large increase in the total calculation time. The reason for the increase in CG steps is not fully understood. When compared to using random LWF coefficients, as in the case of the SDM calculations, the CG minimisation for the D&C subsystem eigensolution based LWFs would be expected occur much more quickly as the LWFs *should* resemble the final states more accurately. Trying to overcome this problem, the D&C subsystem eigensolutions are further orthonormalised and shown in the set of SOLWF calculations. Unfortunately,

Table 3.8: Self-consistent field convergence of the 98-atom linear-alkane chain system. The number of SCF cycles required for convergence is shown for the standard KMG method and the hybrid D&C-KMG method with a localisation region and subsystem radius of 7.5 Å. The calculations are run at varying initial Fermi levels. The transition from the D&C method to the KMG method (for the D&C-KMG method) occurs on the *fourth* SCF iteration. Refer to text for meanings of the different schemes.

Scheme	SCF Transition	Mixing Kick	Initial Fermi Level (eV)						
			-3	-4	-5	-6	-7	-8	-7.8580[a]
Number of SCF iterations									
KMG	-	4	7	11	7	7	7	7	7
SDM	4	4	15	15	7	7	15	19	7
SLWF	4	4	15	15	15	7	15	15	15
SOLWF	4	4	15	15	15	7	7	15	15
Number of CG iterations - 1st SCF iteration after transition									
KMG	-	4	119	118	116	110	109	126	115
SDM	4	4	120	123	121	108	110	138	127
SLWF	4	4	304	304	304	302	304	306	305
SOLWF	4	4	304	304	304	302	302	305	302
Number of CG iterations - 2nd SCF iteration after transition									
KMG	-	4	13	13	13	12	12	12	12
SDM	4	4	5	5	4	4	4	4	4
SLWF	4	4	5	4	4	4	4	4	4
SOLWF	4	4	5	5	4	3	3	4	4
Total Calculation Time (s)									
KMG	-	4	331	442	323	318	318	339	317
SDM	4	4	566	561	319	313	556	683	319
SLWF-	4	4	615	616	617	375	619	618	610
SOLWF	4	4	614	621	616	375	373	615	624

[a] The Fermi level found by the D&C method and used to seed the KMG based orbital minimisation method.

the orthonormalisation process produced similar results to the SLWF set, not improving the situation.

The next step is to try the same set of calculations with the transition from the D&C method to the KMG method occurring at a point closer to self-consistency. In this case, the transition occurs on the sixth SCF iteration. Table 3.9 shows a summary of the results. The linear-mixing kick is also moved to the sixth SCF iteration. Findings show that applying the linear-mixing kick at the same time of the transition helps with convergence. The KMG calculations all converged in 8 SCF iterations, while 11 iterations is the lowest number of iterations for the D&C-KMG method. Prolonging the transition between the two methods did not greatly effect the number of CG steps in the first SCF iteration after the transition. This suggests that even though the seeded density matrix assists with the convergence, the seeded LWFs are the factor that hinder the process. Surprisingly, the CG minimisation in the second transitioned SCF iteration for all D&C-KMG calculations completed in 1 step, as the density matrix was not close to convergence.

Because delaying the transition did not improve the CG minimisation process, the next set of calculations are set back to a transition at the fourth SCF iteration with the linear-mixing kick also set back to the fourth SCF iteration. In this case, the D&C subsystem radius and KMG LWF localisation region is set to 10 Å to see if a more accurate calculation assists the method. For the SDM cases, the number of SCF iterations ranged from 7 to 15 iterations and once again the number of CG minimisation steps in the first transitioned SCF iteration is still comparable to the KMG method. For the SLWF and SOWLF cases the there is a similar pattern when compared to the calculations with a localisation region of 7.5 Å. Even though the number of SCF iterations can be kept low, the number of CG steps in the first SCF iteration after the transition is comparable to the number of the CG steps with the standard KMG method. For an initial Fermi level of -8.0 eV the SLWF and SOLWF calculations failed to converge since the CG minimisation diverged in the second SCF iteration after the transition.

In summary, the primary issue for all sets of calculations is that the number of CG steps for the first SCF iteration after the transition from the D&C method

Table 3.9: Same as Table 3.8 using a 7.5 Å D&C subsystem radius and KMG localisation region. The difference this time is the transition from the D&C method to the KMG method (for the D&C-KMG method) occurs on the *sixth* SCF iteration.

Scheme	SCF Transition	Mixing Kick	Initial Fermi Level (eV)						
			-3	-4	-5	-6	-7	-8	-7.7764[a]
Number of SCF iterations									
KMG	-	6	8	8	8	8	8	8	9
SDM	6	6	16	16	29	16	27	13	13
SLW	6	6	14	11	29	29	11	11	15
SOLWF	6	6	13	11	29	23	21	11	15
Number of CG iterations - 1st SCF iteration after transition									
KMG	-	6	119	118	116	110	109	126	114
SDM	6	6	120	123	121	108	110	139	127
SLWF	6	6	304	304	302	302	304	306	302
SOLWF	6	6	304	304	302	302	302	306	302
Number of CG iterations - 2nd SCF iteration after transition									
KMG	-	6	13	13	13	12	12	12	12
SDM	6	6	1	1	1	1	1	1	1
SLWF	6	6	1	1	1	1	1	1	1
SOLWF	6	6	1	1	1	1	1	1	1
Total Calculation Time (s)									
KMG	-	6	355	355	353	347	347	353	377
SDM	6	6	600	595	979	590	922	506	505
SLWF	6	6	591	516	1116	1043	515	497	692
SOLWF	6	6	163	143	319	260	240	142	182

[a] The Fermi level found by the D&C method and used to seed the KMG based orbital minimisation method.

Table 3.10: Same as Table 3.8 except with a 10.0 Å D&C subsystem radius and KMG localisation region. The transition from the D&C method to the KMG method (for the D&C-KMG method) occurs on the *fourth* SCF iteration.

Scheme	SCF Transition	Mixing Kick	Initial Fermi Level (eV)						
			-3	-4	-5	-6	-7	-8	-7.5084[a]
Number of SCF iterations									
KMG	-	4	11	7	7	11	7	7	7
SDM	4	4	15	7	7	7	7	11	7
SLWF	4	4	11	15	7	7	7	DNC[b]	7
SOLWF	4	4	11	15	15	15	7	DNC	7
Number of CG iterations - 1st SCF iteration after transition									
KMG	-	4	67	64	59	57	56	75	61
SDM	4	4	67	62	58	56	59	105	66
SLWF	4	4	302	302	302	302	302	312	302
SOLWF	4	4	302	302	302	302	302	309	297
Number of CG iterations - 2nd SCF iteration after transition									
KMG	-	4	13	13	12	12	12	12	12
SDM	4	4	1	4	4	4	4	15	4
SLWF	4	4	1	1	3	3	3	CGD[c]	3
SOLWF	4	4	1	1	1	1	3	CGD	4
Total Calculation Time (s)									
KMG	-	4	443	315	315	433	313	317	313
SDM	4	4	563	320	314	314	315	464	318
SLWF	4	4	574	671	424	429	428	DNC	428
SOLWF	4	4	548	668	675	670	439	DNC	432

[a] The Fermi level found by the D&C method and used to seed the KMG based orbital minimisation method. [b] DNC - Did not converge. [c] CGD - Conjugate-gradients minimisation process diverged.

to the KMG method is either comparable or much larger (depending on if the LWF coefficients are seeded from the D&C method) than the number of CG steps required for the standard KMG method. This particular issue has been shown in all sets of calculations and is the primary cause of the failure of the method. Possible steps that can be taken to reduce the time spent in the CG minimisation are as follows. The form of the seeded LWFs can be made to be more representative of a Wannier-type function with transformation procedures, such as orbital localisation techniques [139–143]. The final suggestion involves modifying the KMG band structure energy functional to accommodate the D&C seeded LWFs. Because the KMG functional naturally transforms any LWF form into the final states, transforming the seeded LWFs becomes more of a viable option.

3.8 Concluding Remarks

This implementation successfully combined the density matrix D&C scheme with the SIESTA methodology for computing the Hamiltonian and overlap matrices. The implementation exhibits linear-scaling within the D&C scheme, provided the dimensions of the physical system exceed those of the allowed range for the localised states. The applicability to a variety of systems with varying band gaps has been demonstrated, including a near-metallic carbon nanotube. This scheme will allow practical electronic structure calculations of very large systems, consisting of thousands to tens of thousands of atoms, with relatively modest computational resources. While the results of the D&C scheme are comparable to those currently obtained with the Kim-Mauri-Galli algorithm in SIESTA, the robustness of the approach leads to it being advantageous for systems with small band gaps, and therefore a valuable alternative approach to achieving linear-scaling within the SIESTA methodology. When executed in parallel for large systems the D&C approach exhibits near perfect speedup, providing there is appropriate load balancing.

No need to ask.
He's a smooth operator.
- Sade

4

Divide-and-Conquer Dynamics

As shown in the previous chapter, the D&C linear-scaling method is very efficient and effective at calculating the energetics and forces of large quantum systems. A study involving the dynamics of atoms with regard to the D&C method follows. Firstly, it is shown that discontinuities in the potential energy surface occur when atoms enter or leave a subsystem. A method to alleviate this problem by smoothing(tapering) the interactions between all atoms in the subsystem and with atoms near the boundary of the subsystem is investigated. The use of an auxiliary outer buffer region coupled with the tapering method is also investigated. The second part of the chapter investigates the Frozen Density Matrix (FDM) method developed by Lee & Yang for the D&C method. The D&C FDM method improves the efficiency of the D&C method in MD simulations. The premise of the FDM method is to only calculate portions of the density matrix from the subsystem eigenvectors in areas which are deemed interesting at each molecular dynamics (MD) step. The rest of the density matrix is calculated from eigenvectors cached from a previous MD step. In this way only a relatively few subsystems are diagonalised while the rest are already stored in memory, reducing the calculation time per MD step. The standard FDM method is enhanced to handle multiple regions instead of only the dynamic and frozen regions. All methods are implemented within the current D&C implementation

4.1 Combating Energy Discontinuities

When a system is solved in a particular way, as in the D&C method, that provides an alternate and somewhat easier avenue for finding a solution, an approximation is generally made. As with all approximations, there is an inherent error in the accuracy of the results that usually cannot be avoided, though the error can be reduced. With the D&C method a common indicator of the level of approximation is the difference found in the total energy of the Kohn-Sham system when compared to conventional diagonalisation techniques. A simple way to reduce this error is to increase the size of the subsystems used to partition the system, although, as shown in 3.6, an increase in the prefactor of the method also occurs. This in turn increases the cross-over point at which the D&C method becomes beneficial to use in terms of the computational time required for a given level of accuracy. Generally, the size of the subsystem will be of the order of twice the largest orbital radius found in the system, or more commonly, the largest interaction length between any two elements in the subsystem Hamiltonian matrix. Regardless of the size of the subsystem, unless the subsystems encapsulate the complete system, there will always be an approximation to the total energy with the D&C method.

Within a simulation where atomic positions are altered, whether it be an MD simulation or a geometry optimisation, the D&C approximation provides another source of error related to the total energy that must be addressed. Illustrated in Figure 4.1, this error comes in the form of discontinuities in the potential energy surface as atoms enter and/or leave subsystems. In this scenario, the total energy of a water dimer is plotted as the distance separating the monomers is varied. Comparing the smooth and continuous energy landscape of the conventional diagonalisation method found in SIESTA (thick black line) to the discontinuities found in the D&C runs (dotted line and lines with symbols) it is easy to foresee the build up of possible errors in a dynamics simulation.

SIESTA, using localised PAOs, requires the consideration of the overlap of orbitals as an important aspect of this error. The error will be most prominent when the core atom orbitals couple with buffer atom orbitals near the boundary of

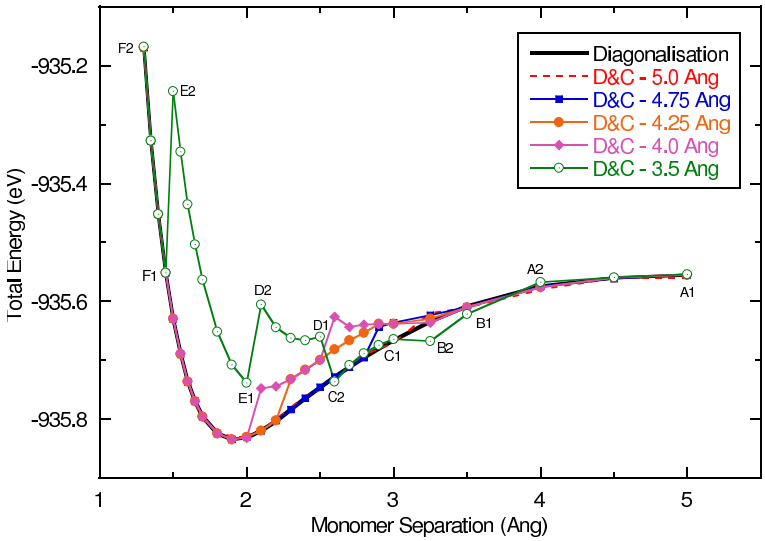


Figure 4.1: Energy curves of a water dimer, calculated by varying the distance between the two water monomers. Conventional diagonalisation and the D&C method with a variety of subsystem sizes are used to solve for the total energy. Discontinuities caused by the D&C method are shown to occur as the atom membership of the subsystems changes as the distance between the water monomers is decreased.

the subsystem. The size of the discontinuity will be proportional to the strength of the orbital coupling. The increase or decrease in energy as atoms enter or leave a subsystem is also proportional to the number of atoms already within the subsystem relative to the number of atoms entering or leaving the subsystem. The larger the number of atoms within a subsystem, the larger the number of atoms that must change subsystem memberships to produce a change in energy which is fractionally different from the trend. Typically, the number of atoms entering or leaving a subsystem will be much smaller than the number of atoms within the subsystem, which can lead to small changes in energy. Regardless of the size of the changes in energy, these must be handled correctly to produce smooth energy landscapes.

A closer look at Figure 4.1 shows that as the subsystem size is reduced, the energy discontinuities become larger. Starting with a subsystem radius of 5 Å (red dashed line), it is more or less the same result as the conventional diagonalisation (thick black line) found in SIESTA. There is a slight difference in the energy between a monomer separation of 5 Å to 3.25 Å, where the number of buffer atoms

Table 4.1: Listing of the number of buffer atoms found within three subsystems of one of the water monomers as the monomer separation is varied. The case of a subsystem with radius of 3.5 Å is shown. The monomer separation is segmented into regions labelled alpha-numerically as shown in Figure 4.1. Changes in the number of buffer atoms in all subsystems correspond to an energy discontinuity. The regions have been chosen specifically to highlight the energy discontinuities.

Subsystem Core Atom	Number of buffer atoms in region					
	A1 - A2	B1 - B2	C1 - C2	D1 - D2	E1 - E2	F1 - F2
O	2	3	3	4	5	5
H_1	2	2	3	3	4	5
H_2	2	2	3	3	4	5

within the subsystems varies between four and five, with a small discontinuity at 3.25 Å. Once the monomer separation reaches 3 Å each subsystem holds five buffer atoms, which is the maximum number of buffer atoms for this system. In the limit of each subsystem containing the maximum number of buffer atoms, the D&C method is equivalent to the conventional diagonalisation method. For a subsystem radius of 4.75 Å (blue line + squares) and 4.25 Å (orange line + circles) the discontinuities are larger than the 5 Å subsystem case. The discontinuities in the energy curve are detrimental to any dynamics simulation based on sampling a potential energy surface. An extreme case is shown with a subsystem radius of 3.5 Å (green line + diamonds) where there are many large discontinuities. The data for the 3.5 Å subsystem radius is marked with labels that indicate different monomer separation ranges. Within these ranges the number of buffer atoms in all subsystems remains relatively constant. A change in the number of the buffer atoms can result in a discontinuity. The number of buffer atoms within these ranges for a single water monomer are listed in Table 4.1.

The discontinuities in the total energy suggest the use of a large subsystem is necessary. However, this option will not always be available due to constraints on resources. Large subsystems will certainly not be desirable for quick qualitative simulations. When this is the case, a procedure to alleviate the discontinuities is needed. What is opted for here is to use a switching function which will

taper(smooth) the interactions of all atoms in a subsystem, dependent on the radial distance of the atom from the centre of the subsystem. Interactions will slowly be introduced or diminished near the boundary of the subsystem depending on whether an atom is entering or leaving the subsystem. As the tapering is always applied, the tracking of subsystem atom memberships between dynamics steps is not required.

The use of switching functions to taper interactions between particles is found within potentials using in MD simulations, in particular for non-bonded, long-range electrostatic interactions [144–151]. Switching functions are also found in various implementations and derivations of the more complex many-body *conductor like-screening model* (COSMO) [152], implemented in quantum and molecular mechanical codes. Senn *et al* [153] used a product of switching functions in their modification of the COSMO scheme within *ab initio* molecular dynamics. Similarly Delley [154] used switching functions in his periodic version of COSMO, while Gale and Rohl used a product of switching functions in a derivative of COSMO called COSMIC [155], which also deals with periodic systems within their molecular mechanics code GULP [7]. No work was found in the literature regarding the smoothing of interactions in the D&C method, though there is related work with the use of a double buffer partitioning scheme to help with SCF convergence, developed by Dixon and Merz Jr. [25] in their semi-empirical D&C implementation [24]. Dixon and Merz Jr. favour the use of an outer buffer region (surrounding the first buffer region) to dampen any interactions with core atoms and buffer atoms near the boundary of the subsystem, which they call *edge contributions*. The outer buffer region is used when diagonalising the subsystem Hamiltonian, though the calculated eigenvectors from the outer buffer region do not contribute to the global density matrix. While Dixon and Merz Jr. developed this technique to help with the SCF convergence rate, the technique can be considered equally valid within the context of the work done here.

It is also expected that the SCF convergence rate will be improved upon with the use of the switching function, as the coupling between the core and buffer orbitals at the boundary is damped in an analogous fashion to when using an outer buffer region. In regards to the SCF convergence, Akama *et al* [156] kept the

number of SCF iterations low, as well as the error in the total energy, by reducing the electronic temperature during the D&C SCF process. That particular group also developed a D&C DIIS method [28] to help with SCF convergence. For systems that are difficult to converge Shaw *et al* [114] recommend turning off DIIS to help reach self-consistency when near convergence. In general, it has been found that a very low mixing weight can help with convergence for systems that prove to be difficult to converge. Use of a low mixing weight is not specific to the D&C method. The group of Vashishta and Shimojo used a multi-level embedding scheme within their hierarchical real-space D&C implementation [23] for MD simulations [115–119]. They also use two buffer regions that are embedded in a global Kohn-Sham self-consistent field.

The outer buffer scheme has also been implemented within the current D&C SIESTA implementation. Tests on the the outer buffer region and combining the switching function with the outer buffer region are presented in section 4.1.3.

4.1.1 Formalism

The switching function is applied to both the subsystem Hamiltonian and overlap matrices. The function smoothly changes from 1 to 0 within a specified range. A cosine-based switching function is used here, and defined as follows:

$$\Theta(r) = \begin{cases} 1 & \text{if } r \leq r_{min} \\ 0 & \text{if } r \geq r_{max} \\ \frac{1}{2}\left(1 + \cos\left(\frac{\pi(r-r_{min})}{r_{max}-r_{min}}\right)\right) & \text{if } r_{min} < r < r_{max} \end{cases} \tag{4.1}$$

The constants r_{min} and r_{max} define the range in which the function smoothly goes from 1 to 0 and the variable, r, is the distance at which it is to be evaluated. The switching function is then applied to each subsystem Hamiltonian;

$$H_{ij}^{\alpha} = H_{ij}^{\alpha}\Theta(r_i)\Theta(r_j) \tag{4.2}$$

where H_{ij}^{α} is the Hamiltonian element belonging to a subsystem, α, with corresponding basis functions, i and j. Similarly for the subsystem overlap matrix[a];

$$S_{ij}^{\alpha} = \begin{cases} S_{ij}^{\alpha} & \text{if } i = j \\ S_{ij}^{\alpha}\Theta(r_i)\Theta(r_j) & \text{if } i \neq j \end{cases} \tag{4.3}$$

where in this case only the off-diagonal matrix elements are effected, as the diagonal elements must remain equal to 1. This is to avoid numerical instability issues when the taper function approaches zero. When this occurs, the on-diagonal elements of overlap matrix, S, will go to zero and the matrix can become singular as its determinant will tend to zero. A singular S matrix will cause numerical issues in the evaluation of S^{-1} due to singularities.

It must be noted that any number of switching functions can be used to taper the interactions, provided that the function goes smoothly to zero. In practice, only the core - buffer and buffer - buffer orbital interactions are considered. The core - core interactions typically occur far from the subsystem boundary.

4.1.2 Water Dimer

The switching function is now applied to the water dimer example considered previously, see Figure 4.1. A comparison of a variety of tapering parameters is carried out for the 4.75 Å and 3.5 Å subsystem radii data sets, shown in Figure 4.2. The results are presented for the case where the minimum taper radius, r_{min}, is varied and the maximum taper radius, r_{max}, is set to the subsystem radius (i.e. the switching function will go to zero at the subsystem boundary).

The calculations were carried out using a 350 Rydberg cut-off for the real-space integration grid used to represent the density, a DZP basis set on all atoms with an energy shift of 0.005 Rydberg for the PAO orbital confinement and with a density matrix convergence criteria of $1\text{x}10^{-4}$ for self-consistency. The PBE [54] form of the GGA was used for the XC functional. Norm-conserving Troullier-Martins pseudopotentials [88, 91] in the Kleinman-Bylander factorised form [89, 90] were used. The employed pseudopotential reference configurations

[a]Note that no weighting of the density matrix is performed.

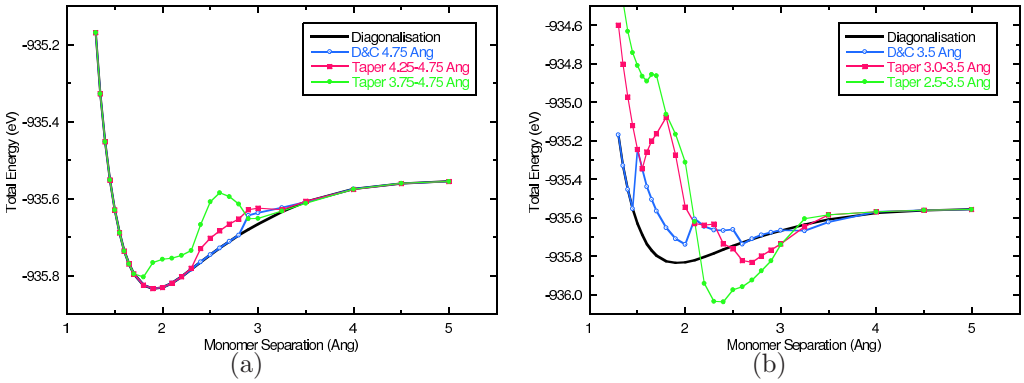


Figure 4.2: The effects of applying a switching function to smooth the discontinuities in the total energy curve for the water dimer system. a) subsystem size of 4.75 Å. b) subsystem size of 3.5 Å.

were $2s^2 2p^4 3d^0 4f^0$ and $1s^1 2p^0 3d^0 4f^0$ for O and H, respectively. The cut-off radii for all angular momentum channels, s, p, d and f, were specified as follows; for O, $1.14a_0$ and for H, $1.25a_0$.

The energy curve for a subsystem radius of 4.75 Å has a single steep change in the total energy at a monomer separation between 2.8 Å and 3.25 Å. It is unclear if this change in energy is actually a discontinuity or just a steep but smooth change; in either case the tapering mechanism is applicable. Figure 4.2(a) shows the effects of the tapering.

The first set of data uses a taper region of 0.5 Å (pink line + squares), where r_{min} is set to 4.25 Å. In this case, the energy value at a monomer separation of 2.8 Å is slightly reduced due to the tapering, although the change energy has not been eliminated. Overall, the form of energy curve deviates from the energy curve calculated from the standard D&C calculation (blue line + empty circles) and the standard diagonalisation (black line). A larger taper radius of 1.0 Å (green line + circles), where r_{min} is set to 3.75 Å, is also shown. Here the change in energy of the standard D&C calculation is slightly smoothed, although the form of the energy curve deviates greatly from both standard D&C and diagonalisation curves.

Shown in Figure 4.2(b), the second energy curve with a subsystem radius of 3.5 Å (blue line + empty circles) has many discontinuities in its energy landscape. The application of the switching function with a tapering region of 0.5 Å (pink

line + squares) does little to smooth the curve. A larger taper region of 1.0 Å (green line + circles) does smooth the curve. However, the form of the curve deviates from the standard diagonalisation (black line) curve and has a large difference in the minimum energy.

The ineffective nature of the switching function, when applied in this case, is due to the small number of atoms and relatively large changes in energy. The next section will deal with a system with a larger atom count and proportionally smaller discontinuities.

4.1.3 Interaction Between Two Silicon Clusters

The switching function was able to produce relatively smooth energy curves for the water dimer case, though the proportionally large changes in the energy made it difficult to demonstrate any elimination of discontinuities and the resulting form of the curves were vastly changed. In this section, the elimination of discontinuities resulting in a smooth and continuous energy landscape is shown for a larger system of two interacting hydrogenated 20 Å silicon clusters, shown in Figure 4.3. More precisely, the interaction energy of two 20 Å clusters is investigated as the distance between them is varied. The separation of the clusters is measured between the closest two atoms of each cluster. Only the second cluster is moved in the super cell, while the first cluster is kept at the same position. The tapering mechanism is applied to standard D&C and D&C which incorporates an outer buffer region in its partitioning scheme.

The outer buffer partitioning scheme [25] was originally developed to help accelerate SCF convergence rates. It is used here first to determine if the partitioning scheme can also smooth an energy landscape. Secondly, it is also examined if the use of a switching function within the outer buffer region offers any benefits when compared with using just a taper region. Finally, it is anticipated that the tapering will also accelerate the SCF convergence rate, as the tapering effectively dampens subsystem boundary contributions (called *edge contributions* by Dixon and Merz Jr. [25]) in an analogous fashion to the outer buffer partition scheme. A comparison between the tapered and outer buffer region scheme is made in this respect.

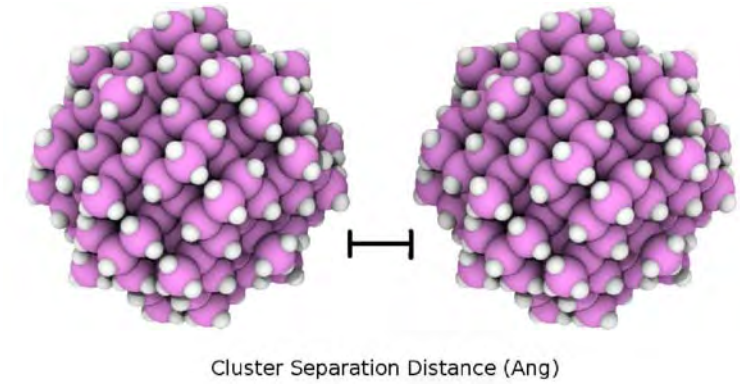


Figure 4.3: Rendering of the two hydrogenated silicon cluster system. Each cluster is 20 Å in diameter with the cluster separation distance measured between the two closest atoms of each particle.

The calculations were carried out using a 500 Rydberg cut-off for the real-space integration grid used to represent the density. A very fine grid was required to reduce the egg-box effect[b] when using real-space methods. A SZ basis set was used on all atoms with an energy shift of 0.02 Rydberg for the PAO orbital confinement and with a density matrix convergence criteria of $1\text{x}10^{-5}$ for self-consistency. Self-consistency is also reached when the total energy is within a tolerance of $1\text{x}10^{-5}$ eV. The PBE [54] form of the GGA was used for the XC functional. Norm-conserving Troullier-Martins pseudopotentials [88, 91] in the Kleinman-Bylander factorised form [89, 90] were used. The employed pseudopotential reference configurations were $3s^2 3p^2 3d^0 4f^0$ and $1s^1 2p^0 3d^0 4f^0$ for Si and H, respectively. The cut-off radii for each angular momentum channel, s, p, d and f, were specified as follows; for Si, $1.89a_0$ and for H, $1.25a_0$ for all channels.

The interaction energy at a specific cluster separation, r, was calculated using $E_{Int}(r) = E_{1,2} - (E_1 + E_2(r))$; where $E_{1,2}$ is total energy of the two cluster system, E_1 is the total energy of the first isolated cluster and $E_2(r)$ is the total energy of the second isolated cluster located at a monomer separation of r. $(E_1 + E_2(r))$ is used instead of just $2E_1$ as to remove the egg-box effect when positioning the second cluster.

[b]The egg-box effect describes the oscillation in the total energy due the integration grid breaking translational symmetry. The total energy oscillates with the grid periodicity and is likened to the shape of an egg-box. It is clearly noticeable as atoms move around within the simulation cell.

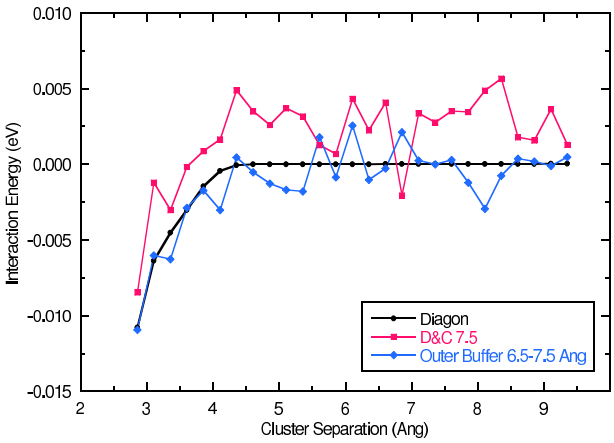


Figure 4.4: The interaction energy of two hydrogenated 20 Å silicon clusters, using standard diagonalisation and the D&C method. The D&C method based curves are discontinuous. i) Standard diagonalisation ii) Standard D&C using a 7.5 Å subsystem iii) D&C using a 6.5 Å subsystem plus a 1.0 Å outer buffer region.

Figure 4.4 shows the interaction energy of the dimer using the D&C method with a subsystem radius of 7.5 Å (pink line + squares) and a subsystem radius of 6.5 Å including a 1.0 Å outer buffer region (blue line + diamonds). These curves are compared to the interaction energy curve generated from standard diagonalisation (black line + circles). Both the standard D&C and D&C with an outer buffer region produce curves with discontinuities. In regards to this system, the outer buffer region did not eliminate any discontinuities.

4.1.3.1 Tapering Mechanism

The switching function is now used to smooth the interaction energy curves of both standard D&C and D&C including an outer buffer region. Six sets of calculations are investigated with varying taper region sizes (for both schemes). All taper regions go to zero at the subsystem boundary (the boundary of the outer buffer region, when this scheme is employed). The taper region sizes range from 0.5 Å to 1.75 Å.

The interaction energy curves for when the tapering mechanism is applied are shown in Figure 4.5. The tapered curves for the standard D&C (pink line + filled squares) and the tapered curves for D&C including an outer buffer region (blue

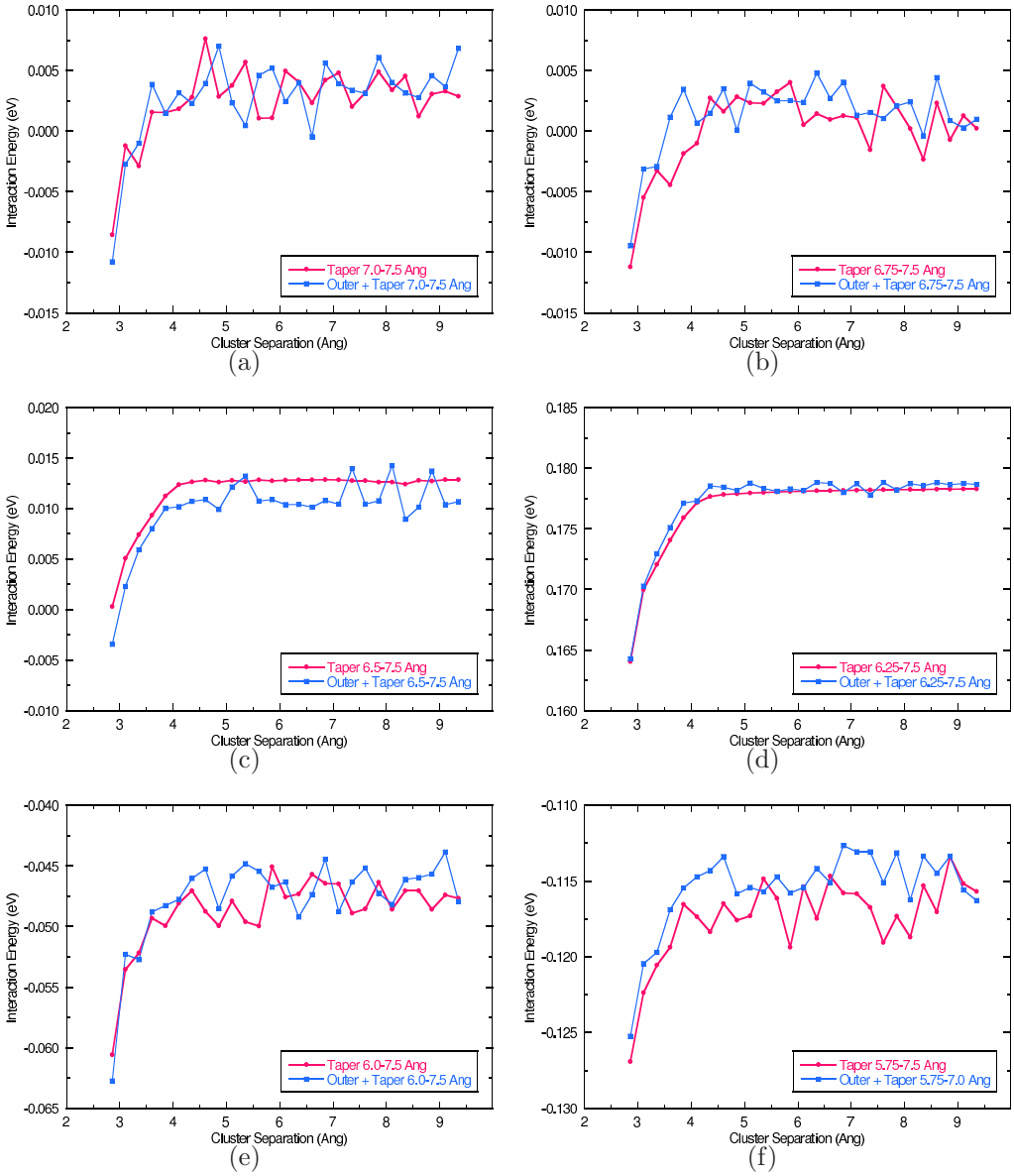


Figure 4.5: The interaction energy of two hydrogenated silicon clusters when the tapering mechanism is applied. The taper region size varies for each plot. The tapering mechanism is applied for both standard D&C and D&C outer buffer schemes with taper sizes of a) 0.5 Å, b) 0.75 Å, c) 1.0 Å, d) 1.25 Å, e) 1.5 Å and f) 1.75 Å.

line + filled diamonds) are compared to the ideal case of standard diagonalisation (black line + filled circles).

A taper region of 0.5Å has little effect on the energy curves, in that the energy curve is still discontinuous (see Figure 4.5(a)). Increasing the taper region to 0.75Å produces curves that are still discontinuous, although there are now regions in the curve that are becoming smoother (see Figure 4.5(b)). For example, between a monomer separation of 2.75Å and 4.25Å for the standard D&C curve and between 4.25Å and 5.25Å for the D&C plus outer buffer region curve. It is obvious that the taper regions are too narrow for the taper mechanism to be effective. With a 1.0Å taper region, the taper mechanism produces a smooth and continuous energy curve for the standard D&C calculation (see Figure 4.5(c)). The D&C with outer buffer region curve is mostly smooth with a few discontinuities. In this regime the taper mechanism is working as expected. Increasing the taper region size up to 1.25Å (see Figure 4.5(d)) produces a smoother curve, and in this instance the D&C with outer buffer scheme also produces a relatively smooth curve, albeit more noisy than the standard tapered D&C curve. For this particular system, a taper region of 1.25Å seems to produce the most desirable results. This result also confirms the effectiveness and applicability of the proposed tapering mechanism. It was found that further increases to the size of the taper region produced discontinuities in the energy once again, as shown in Figures 4.5(e) and 4.5(f). Possible reasons as to why this has occurred will be investigated in section 4.1.3.3.

The tapered interaction energy does not go to zero as the clusters are separated. The standard D&C interaction energy (see Figure 4.4) also does not seem to go to zero. It is expected that the interaction energy would go to zero when the clusters are far enough apart that the D&C subsystems from one cluster do not overlap with the other cluster. To help understand why this occurs it will be worthwhile to examine the individual contributions to the interaction energy.

Table 4.2 lists a summary of the relative contributions to the interaction energy for the two cluster configuration that has the clusters separated the furthest apart at 9.3559 Å. At this separation distance the subsystems of one cluster do not overlap with the other cluster. Except for the long-range Coulomb potential

Table 4.2: Energies (eV) of the two cluster and single cluster hydrogenated silicon systems relative to the energy calculated using standard diagonalisation. The second cluster is located at separation distance of 9.3559 Å.

System	D&C	Taper Region			
	7.5 Å	0.5 Å	0.75 Å	1.0 Å	1.25 Å
1st Cluster	3.4303	3.3470	4.7125	2.8399	-1.8082
2nd Cluster	3.4312	3.3470	4.7123	2.8399	-1.8082
Difference	0.0009	0.0000	-0.0002	0.0000	0.0000
Combined Two Clusters	6.8628	6.6968	9.4249	5.6925	-3.4382
1st + 2nd Clusters	6.8615	6.6940	9.4248	5.6797	-3.6164
Difference	0.0013	0.0028	0.0002	0.0128	0.1782

the system should be treated as two non-interacting clusters within D&C. The top of the table shows the total energies of the single clusters relative to the total energy calculated using standard diagonalisation. The standard D&C and Tapered D&C energies are basically equivalent for each cluster, as can be seen by the energy difference.

The bottom half of Table 4.2 lists the relative total energies of the combined two cluster system and the summation of the single cluster systems. The difference between these energies is equivalent to the interaction energy and at a cluster separation of 9.3559 Å the energies should tend towards zero (see the standard diagonalisation curve in Figure 4.4). As can been seen from Table 4.2 there are non-zero differences for all schemes.

It is expected that the summation of the relative two isolated cluster total energies be close or equal to the relative combined two cluster total energy. For the 1.25 Å taper region case, the sum of the relative isolated cluster energies gives -3.6164 eV while the relative combined two cluster energy is -3.4382 eV which is a difference (interaction energy) of 0.1782 eV. Currently, it is unclear why the D&C method with the tapering mechanism sees an isolated cluster system different to a system consisting of two isolated clusters. Care must be taken when

considering these energies.

4.1.3.2 SCF Convergence

The outer buffer region scheme of Dixon and Merz Jr. [25] was devised to reduce the number of SCF iterations required for convergence when using the D&C method. Akama *et al* [28, 156] have also devised schemes to keep the number of SCF iterations low. It is anticipated that the tapering mechanism will also help accelerate the SCF convergence rate.

The buffer orbitals closest to the subsystem boundary are the least well described orbitals within the subsystem, due to those buffer orbitals' environment not being sufficiently represented. The contributions to the global density matrix from the boundary buffer orbitals can lead to small variations in the global density. The density may then oscillate about the ground state configuration as the SCF procedure heads towards self-consistency. This oscillation, when near convergence, can be the main cause of large SCF iteration counts with the D&C method. The tapering mechanism is a scheme that is proposed to reduce (eliminate) the oscillations and produce low SCF iterations counts.

The number of SCF iterations required to reach self-consistency for each of the silicon cluster curves is shown in Figure 4.6 with Figure 4.6(a) containing the plots for the standard D&C scheme and Figure 4.6(b) containing the plots for the D&C with outer buffer scheme. The standard diagonalisation (black line + filled circles) has a steady convergence rate at 27 iterations for each cluster separation distance. On the other hand, the iteration count of the standard D&C method oscillates about ≈ 40 steps. For the two silicon cluster system, there seems to be no clear benefit of including an outer buffer region with the number of the SCF iterations oscillating about ≈ 45 iterations.

For the taper regions which produce smooth energy curves, the number of SCF iterations is greatly reduced and remains steady. The iteration count remains high and oscillates for cases when the energy curves are not smooth. For example, the 1.25Å taper region calculation (dark green line + empty circles in both figures) has a low and constant SCF iteration count. When the 1.25Å taper region is coupled with the outer buffer region, the number of SCF iterations required for

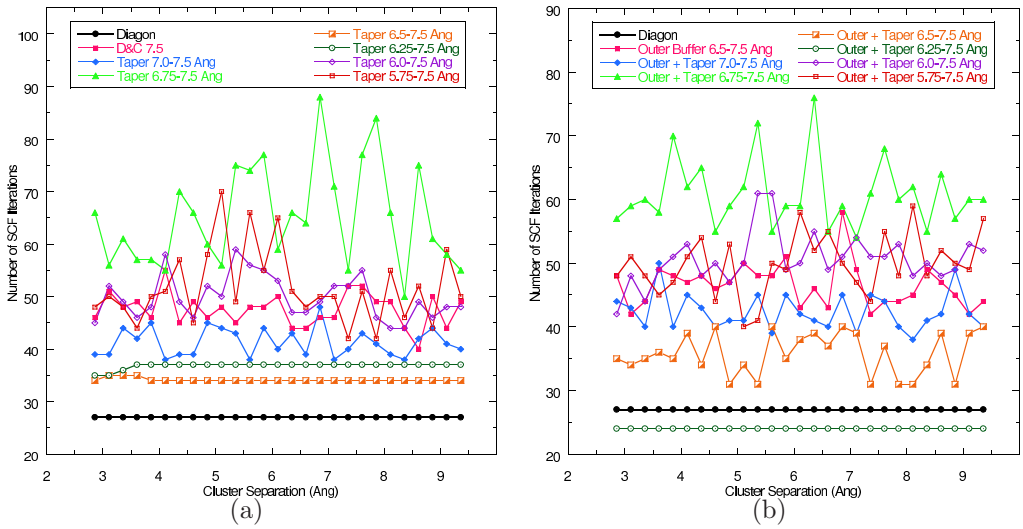


Figure 4.6: The number of SCF iterations required for self-consistency in the density for standard D&C, D&C plus an outer buffer and with the tapering mechanism applied to both schemes for the interacting silicon cluster calculations. For curves which are smooth the number of SCF iterations is also smooth and low in count. a) Tapering mechanism applied to standard D&C b) Tapering mechanism applied to D&C with the outer buffer scheme.

each cluster separation is a constant 24 iterations; surprisingly, the approximate D&C method requires less iterations than the standard diagonalisation method. The tapering mechanism has been found to require less SCF iterations than the standard D&C method (and including an outer buffer region) when the taper mechanism produces smooth energy curves.

4.1.3.3 Tapering Mechanism Issues

The reason for the reintroduction of discontinuities for taper regions larger than 1.25 Å is unclear. To help understand why this occurs, a gradual increase of the taper region from 1.25 Å to 1.4 Å is made, as shown in Figure 4.7(a). The curves between and including taper regions 1.25 Å to 1.375 Å are fairly smooth, although after a cluster separation of 3.5 Å there are variations in the energy. It is expected that these curves will have the same form as the 1.25 Å taper region curve. These variations could be leading up to the discontinuities found in the 1.4 Å taper region curve. As the taper region increases, the energy curves become

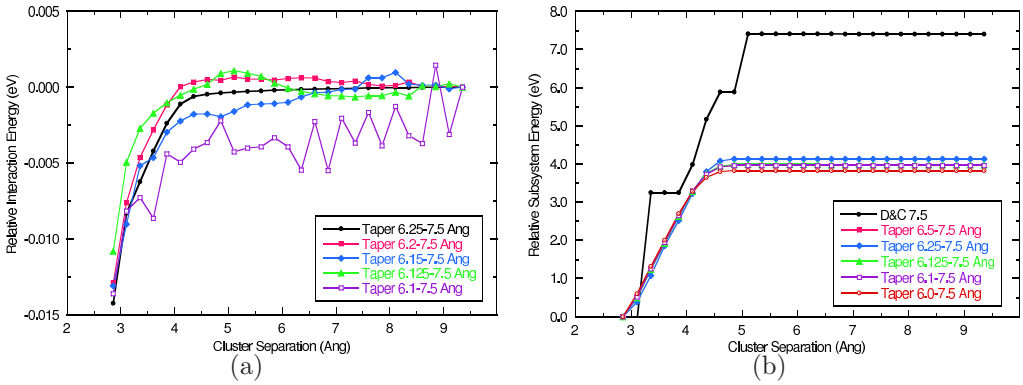


Figure 4.7: Investigations into the reintroduction of discontinuities for taper regions greater than 1.25 Å. a) A gradual increase in the taper region from 1.25 Å to 1.4 Å. b) The energy curves produced with the tapering mechanism for a single subsystem as the separation of the two silicon clusters is varied.

less smooth and their form differs more from the curve of the 1.25 Å taper region, albeit the change in form from a taper region of 1.375 Å to 1.4 Å is more drastic than the others. The 1.4 Å taper region could also be interpreted as a threshold at which the discontinuities are reintroduced. This could mean that there could be fundamental problems with the tapering mechanism itself.

To determine if the tapering mechanism has fundamental problems with the current D&C implementation, the effects of the tapering mechanism on the energy of a single subsystem is shown in Figure 4.7(b). The subsystem consists of a silicon core atom near the edge of the first cluster. This particular subsystem has many atom membership changes and is a good indicator to demonstrate the effectiveness of the tapering mechanism. As the total energy is a weighted sum of the subsystem energies, the form of the subsystem energy is representative of the interaction energy curves. If there are discontinuities found in the subsystem energy after the application of the tapering mechanism, then there will definitely be a fundamental problem with the mechanism.

The standard D&C curve (black line + circles) has discontinuities, as expected; however, the shape and form of the curve is not representative of the interaction energy curves of Figure 4.5. All other curves have the taper mechanism applied to this system using taper regions between 1.0 Å to 1.5 Å. The

energy curves for all taper regions are smooth and have the form of the curve in Figure 4.5(d). Even the 1.5Å taper region produces a smooth curve in the subsystem energy, where the interaction energy curve is discontinuous in this range.

The smooth subsystem energy curves suggest that the tapering mechanism does successfully taper the energy of the subsystem. As the total energy is the weighted sum of the subsystem energies, it would be expected that the total energy (and any derived energy) would also be smooth. There must be another factor that is causing the reintroduction of the discontinuities for taper regions larger than 1.25 Å. One proposed reason could be due to charge fluctuations between subsystems, which could be handled by using the so-called positive and negative fragments partitioning scheme by Zhao *et al* [120]. They typically passivate each subsystem, which turns the subsystem into an insulator and then combines the subsystems in a way so as to a cancel out artificial boundary effects.

4.1.4 Concluding Remarks

It has been shown that discontinuities in the potential energy surface occur when atoms enter or leave a subsystem during atomic displacement. To produce smooth energy surfaces, a switching function has been applied to each subsystem Hamiltonian and overlap matrix to taper the interactions between the core atoms and buffer atoms near the boundary of the subsystem. There were issues with the reintroduction of the discontinuities for large taper regions which requires further work to discover the reason behind this behaviour. The application of the switching function has also been shown to reduce the number of the SCF iterations for cases when the taper produces smooth energy curves.

Finally, a new method is proposed that is an alternative to the tapering mechanism. A potential is added to each subsystem that is the sum of the background charge found surrounding the subsystem. The addition of the potential to each subsystem will be equivalent to using larger subsystems (i.e. indirectly expanding the subsystem boundary), hence reducing the boundary effects on the orbitals central to the subsystem. The potential can be calculated using the particle mesh

Ewald method [157] (or one of its many improvements [158]) or more appropriately the fast multipole method [101, 159]. This proposed method can smooth the energy landscape and also improve the accuracy of the D&C method at the expense of larger computational cost than the tapering mechanism. This new method will be investigated in the near future.

4.2 Frozen Density Matrix Approach

The decrease in computational effort with the use of the D&C method is significant when compared to conventional DFT diagonalisation methods for large systems. The D&C method breaks the $\mathrm{O}(N^3)$ bottleneck of conventional diagonalisation and is proven to be an accurate method for electronic structure problems. However, when considering dynamics simulations of large systems where thousands of dynamics steps could be required to reach the desired outcome, the relatively short time required to converge a D&C SCF calculation can still be prohibitive. It is then crucial that the computational cost for each dynamics step be as small as possible.

Fortunately a large majority of dynamics simulations are only concerned with a certain aspect of a system i.e. a region of interest (active region). Local geometry optimisations are a primary example of this, as one would do so with the active region of a protein. There are a large variety of simulations where there can be deemed an active region to be studied via a dynamical simulation. With these types of problems, it makes sense to apply a high quality, fully quantum calculation in the active region and approximate the rest of the system with a faster, preferably quantum based, method.

One common approach is to use a QM level of theory in the active region and to use a molecular mechanical (MM) approach for the rest of the system [160–169]. The hybrid QM/MM strategy has been successful, but has some shortcomings. Important physical processes and phenomena, like charge transfer between the QM and MM regions, are difficult to approximate using MM techniques. Link atoms [163, 170] are a necessity to represent the QM - MM boundary accurately, and this requires extra input from the user. Some atoms in the MM formalism

do not possess a charge, which makes it difficult for these atoms to interact with the QM handled region. In some situations, the only difference between MM atoms sharing the same charge is in the van der Waal's potential. Primarily these are difficulties in linking a classical atomic theory with a quantum electronic theory [171, 172]. One last issue is the parameterisation of the MM atoms and the link atoms being an undesirable extra level of complication in the calculation setup stage. The drawbacks with the use of MM to accurately represent the physics and interact correctly with the QM region hinders the use of this method. What is needed is a method based on first principles, where the physical processes and phenomena are inherently incorporated into the theory and the input from the user is limited, especially in the sense of parameterising interactions.

One strategy to overcome the above problem is to use the D&C method with the frozen density matrix (FDM) approach [173–175]. In the FDM method, the electron density of the active region is calculated with a high update frequency (generally every dynamics step) while the density of the rest of the system is in essence frozen [176], where the density for the frozen region is calculated from eigenstates cached from previous dynamics steps. New eigenstates for the frozen region are calculated at a much lower frequency than for the active region. This saves on the computational cost per dynamics step. Lee and Wang were the first to adapt the FDM approach to work with their D&C method [27]. Using the D&C subsystems allows for an easy implementation of the FDM method, where by each subsystem is designated to either be an active or frozen region. Ermolaeva *et al* [33] have also developed a FDM method in their semi-empirical based D&C implementation and were able to achieve large gains in computational speed for MD simulations. Not much work has gone on to further the work on D&C FDM, probably due to the large cost in memory required for the technique. The eigenstates of each frozen subsystem must be stored for the calculation of the global density matrix at each dynamics step. Nevertheless, as computing resources expand over time the D&C FDM method will be begin to play a major role in large-scale dynamical simulations.

The following outlines the implementation of a FDM method within the present D&C implementation. The implementation extends the ideas of Lee

and Yang by allowing each subsystem to have its own update rate for making a new contribution to the global density matrix. In effect, the use of multiple regions with varying density matrix update rates is possible. This will allow for a more effective charge transfer between the active dynamical region and the rest of the system. The consequences of working with a more accurate density is that the number of SCF iterations at every dynamics step is reduced and there could be a reduction in the number of relaxation steps in a local geometry relaxation calculation. Ermolaeva *et al* [33] proposed a similar scheme in which the update rates were a function of distance from the active region. The difference between the current scheme and the scheme from Ermolaeva *et al* is one of implementation. The current scheme allows for finer control of the update rates and can allow for systems with multiple active or even semi-active regions.

The present implementation supports parallel computation. The method is tested on a linear alkane chain, C_nH_{2n+2}, consisting of 194 atoms. A local geometry optimisation is performed on one end of the polymer and the results are compared with standard SIESTA and the present D&C implementation.

4.2.1 Implementation

Standard D&C FDM implementations keep the non-active region of the density matrix frozen between consecutive dynamics steps and allow the active region of the density matrix to spatially evolve. This results in the reduction of the number of computed eigensolutions at every SCF iteration. Each subsystem is assigned to either region, where the subsystems in the active region will usually make a new contribution to the density matrix at each dynamics step while the subsystems in the frozen region will make a new contribution to the density matrix at a much lower rate (e.g every 30 dynamics steps). At each SCF step the Fermi energy is computed, using (3.16), from the newly calculated subsystem eigenstates in the active region and the stored subsystem eigenstates from the frozen region. This requires storage of the frozen subsystem eigenvectors and eigenvalues. The dense eigenvector solution coefficients can not be stored in a sparse matrix representation; hence there is a large requirement for memory. The present implementation follows the standard scheme with one difference, in that

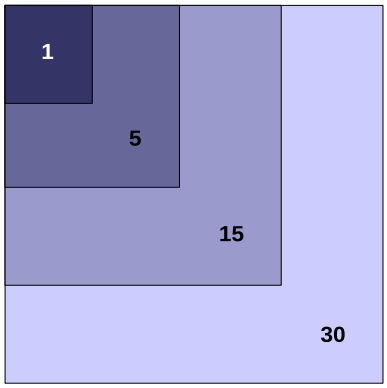


Figure 4.8: A multi-region representation of the present frozen density matrix scheme. The numbers indicate the density matrix update frequency for the subsystems in the region to make a new contribution to the density matrix.

the two regions are generalised to possibly many regions.

The current FDM scheme generalises the designation of the active and frozen regions by allowing each subsystem to update its own density matrix contribution at an assigned frequency. In effect, each subsystem is decoupled from belonging to either *just* an active or frozen region, allowing the designation of many regions with different density matrix update frequencies. One can imagine using many regions with decreasing update frequencies for regions further away from a region of interest (there may be more than one active region). This would allow for more effective density flow between an active region and the rest of the system. An example of the FDM scheme is shown in Figure 4.8 where an active region is located at the top left hand corner. The numbers represent the number of dynamics steps after which the particular region will add its new contribution to the density matrix. Regions close to the area deemed interesting have a high update rate, while regions further away from the active region (top-left corner) do not need to make a new contribution to the density matrix so regularly.

The specification of the update rates are either explicitly defined or taken from the constrained atom specification. Within the SIESTA code, the constrained atom specification defines all atoms which remain stationary during a dynamical simulation. If the frequencies are defined based on the constrained atom specification then there can be only two update frequencies. One frequency

for the active region (the unconstrained atoms), which will generally be updated at each dynamics step, and one frequency for the frozen region (the constrained atoms) updated at a user defined frequency. An option to automatically update the subsystems depending on if they are currently interacting with any active subsystems is also available. This option allows users to only specify the region that is part of the dynamics and not the surrounding environment, thus reducing the complexity of the setup stage. At the first SCF cycle of each dynamics step the eigensolutions of all subsystems are computed. This is to update the eigenstate data of each active subsystem in case the atom membership of the subsystems have changed during the dynamics move. Ideally, only the active subsystems and subsystems interacting with active subsystems should be considered as this would slightly decrease the total calculation time. When using the manual specification of the subsystem density matrix update rates there is no dependence on (un)constrained atoms and update frequencies. It is advisable that atoms which are unconstrained should update the density matrix at each dynamics step, but for all other atoms there are no predetermined requirements and should be defined so that the electron density can easily flow across the active and non-active boundary. Finally, an option to recalculate all eigensolutions for all subsystems at a specified SCF step is made available. This option will allow for a better convergence rate if issues arise with converging the SCF density for a given atomic configuration.

4.2.2 Memory Considerations

Storage of the subsystem eigenvectors and eigenvalues limit the use of the D&C FDM method to only moderately sized systems. An attempt to overcome the memory problem by using a different D&C FDM scheme has been implemented. The new scheme keeps the Fermi level constant throughout the calculation. This scheme does not require the storage of the frozen eigenstates. The global density matrix is calculated by summing the previously calculated non-active region only of the global density matrix with new contributions to the density matrix coming from the active region. Due to the constant Fermi level, there is always a constant orbital occupation within the non-active region. Only the sparse density matrix

from the previous SCF step is required to be stored.

The above scheme failed to converge during the SCF process. The issue is with the flow of electron density between subsystems, especially between the active and frozen regions, is severely limited when using a constant Fermi energy. Electron density will be built up or depleted near the boundary of the active and frozen regions creating unphysical densities, eventually leading to failure in the convergence. Calculating a new Fermi level by solving for the complete global density matrix once every specified number of SCF cycles would allow charge to flow between the active and frozen parts of the density. However, allowing a complete recalculation of the density matrix is still not enough to alleviate the convergence problem.

Another proposed scheme to help reduce the memory cost involves storing the subsystem eigenvectors and eigenvalues onto disk and then accessing them when required. Using this scheme the number of subsystems loaded into memory can be set at runtime. This is yet to be tested and validated, where the efficiency of reading the data from disk will be hardware dependent i.e. the competitiveness of this method is in question. Finally, an alternative scheme would be to use a partitioning scheme in which there are many core atoms. In this type of partitioning scheme the duplication of buffer atoms in adjacent subsystems is reduced leading to less memory usage.

4.2.3 Results

The accuracy and performance of the D&C FDM implementation when performing a local geometry optimisation is investigated. The local geometry optimisation is of 10 atoms at one end of a 194-atom linear alkane molecule. Comparisons to standard SIESTA diagonalisation and the present D&C implementation are made.

All following calculations were carried out using a 350 Rydberg cut-off for the real-space integration grid used to represent the density, a DZP basis set on all atoms with an energy shift of 0.01 Rydberg for the PAO orbital confinement and with a density matrix convergence criteria of $1\text{x}10^{-4}$ for self-consistency. An electronic temperature of 100 K was used. The PBE [54] form of the GGA

was used for the XC functional. Norm-conserving Troullier-Martins pseudopotentials [88, 91] in the Kleinman-Bylander factorised form [89, 90] were used. The employed pseudopotential reference configurations were $2s^2 2p^2 3d^0 4f^0$ and $1s^1 2p^0 3d^0 4f^0$ for C and H, respectively. The cut-off radii for each angular momentum channel, s, p, d and f, were specified as follows; for C, $1.33a_0$ for all channels; and for H, $1.25a_0$ for all channels. Calculations were performed in parallel using 4 processors of a 11900-processor SUN Constellation machine (2.93GHz Intel Nehalem cpus).

The system used in the local geometry optimisation is first set up. A relaxation of the complete linear alkane molecule is performed, where the atoms were allowed to relax to a convergence criteria of 0.01 eV/Å. Then using the standard Verlet MD simulation, the atomic positions of the first 10 atoms at one end of the molecule were allowed to evolve for a duration of 13 fs using a timestep of 0.1 fs at an initial temperature of 5,000 K. The high temperature introduced random velocities for each of the 10 atoms, sampled from a Maxwell-Boltzmann distribution with the corresponding temperature. A short timestep was used to ensure that the atom trajectories were stable especially at such a high temperature. The resultant atomic configuration (of the 10 atoms) strayed greatly from their relaxed positions; with a maximum force of 15.39 eV/Å. It is from this configuration that the local geometry optimisations will take place.

The local geometry optimisation is performed on the 10 atoms that were allowed to evolve during the above mentioned MD simulation. A comparison of five sets of relaxations are carried out, where summaries of the results are shown in Tables 4.3 and 4.4. The first calculation uses the standard SIESTA diagonalisation method (labelled Diagon), the second uses the present D&C implementation with a subsystem radius of 7.5Å (labelled D&C). The last three calculations use the D&C FDM implementation, as detailed earlier, all using a subsystem radius of 7.5Å. The first of the D&C FDM (labelled FDM$_A$) calculations incorporates two regions[c], where the first region encompasses the 10 unconstrained atoms and updates its contribution to the density matrix at each relaxation step. The sec-

[c]A reminder that the notion of a region does not technically exist in this implementation; only density matrix update rates for individual subsystems are understood. It is the grouping of subsystems with a common density matrix update rate which designates a region.

Table 4.3: The performance of the FDM D&C method compared with standard SIESTA diagonalisation and the present standard D&C implementation.

Scheme	Number of Relaxation Steps	Average Diagonalisation Time (secs)
Diagon	60	1.27
D&C	64	2.10
FDM_A	224	1.01
FDM_B	124	1.03
FDM_C	62	0.97

ond region encompasses all the constrained atoms and updates the density matrix at every 8 dynamics steps. The second set of D&C FDM calculations (labelled FDM_B) takes advantage of the different density matrix update rates available to each subsystem by allocating three update regions. The first region encompasses the 10 unconstrained atoms and updates its density matrix at every relaxation step. The next 10 atoms along the alkane molecule make up the second region, where updates to the density matrix are made at every 4 steps. The rest of the other atoms make up the final region and the density matrix update rate for this region occurs at every eight relaxation steps.[d] It is advisable that the update rates for the regions be a factor of the second lowest rate (4 in this case) so that a full system update to the density matrix is possible at some stage during the relaxation. The final D&C FDM calculation (labelled FDM_C) uses the same allocation of density matrix update rates as does FDM_B with the application of the tapering mechanism, as detailed in section 4.1, to provide a faster SCF convergence rate. The tapering range is set between 6.5Å and 7.5Å (the subsystem boundary).

Table 4.3 lists a summary of the performance of all schemes. The number of relaxation steps indicates the number of steps required for the relaxation process to complete. As expected, the Diagon scheme has the lowest number of steps at 60 followed closely by the D&C method at 64 steps. The D&C FDM scheme,

[d]Due to the fact the partitioning scheme is the one used by Yang *et al* [20] where each atom is the core of a subsystem, the allocation of atoms to so-called regions is in fact the allocation of subsystems to regions.

FDM_A, has the highest number of relaxation steps at 224. The basic FDM scheme needed 3.5 times the number of relaxation steps than the standard D&C method. Even for this simple system the FDM method requires a large number of steps to converge, emphasising the fact that improvements to the method are needed. The multi-region scheme, FDM_B, improves on the performance of FDM_A with 124 steps, nearly half the number of steps. The use of multiple update rates leads to a faster path to convergence in the relaxation process, providing evidence that such a scheme is worthwhile. This indicates that larger approximations will decrease the geometry convergence rate. Where the FDM_B scheme is less of an approximation than FDM_A, and where the D&C FDM methods are a greater approximation than the Diagon and D&C methods. FDM_C has the least amount of relaxation steps of the D&C FDM schemes at a count of 62 (comparable to the Diagon and D&C runs). This is what is expected when tapering the interactions within each subsystem, as this not only smooths the potential energy surface but also accelerates the SCF process.

The average time required to calculate the eigensolutions is also shown in Table 4.3. The Diagon scheme diagonalises the Hamiltonian in 1.27 secs on average. The standard D&C method runs in the longest time at an average of 2.10 secs. The reason why the D&C method is not faster than the Diagon method is because the size of this system is below the cross-over point where it becomes beneficial to use the D&C method. For a larger alkane chain the D&C method will be faster than the Diagon method. All D&C FDM schemes operate in a shorter time than the D&C schemes, at ≈ 1.00 secs. The use of extra regions has not increased the calculation in this case. It is expected that for larger systems that the proposed FDM scheme will take longer to process than the standard FDM, due to the extra subsystem diagonalisations required at each SCF step. The D&C FDM schemes, as expected, perform more efficiently that the standard D&C scheme and Diagon scheme. The minimum time to assemble the Hamiltonian and overlap matrices was found to be 31.29 secs. The FDM method does offer such a larger performance increase over the standard D&C method for this alkane system. The most important finding for the proposed scheme is the reduction in the number of relaxation steps.

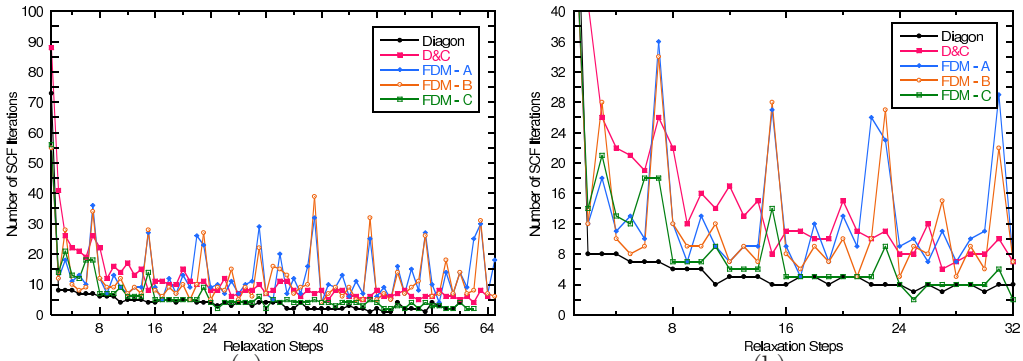


Figure 4.9: The number of SCF iterations per relaxation step required for the 194-atom linear alkane chain geometry optimisation. a) the first 64 SCF iterations b) a close up view of the first 32 SCF iterations.

The number of SCF iterations per relaxation step is also another important factor to consider, as it is desirable to keep the SCF iterations to a minimum. Using the FDM approximation should also increase the number of SCF iterations due to the limited charge transfer between subsystems. The use of the standard D&C FDM scheme does not always accurately represent the electron density. With charge depletion and build up near the boundary between the active and frozen regions, the electron density might be slightly unphysical. This will complicate the SCF convergence, increasing the number required iterations. The use of the proposed D&C FDM scheme, with individual subsystem density matrix update rates, should help alleviate unphysical densities by allowing more frequent charge transfer between regions and help keep the SCF iteration count low.

Figure 4.9 shows the first 64 SCF iteration counts for each scheme. The average number of SCF iterations for FDM_A (blue line + diamonds) is comparable to the average number of SCF iterations for FDM_B (orange line + empty circles). In this case the use of extra regions has not improved upon the standard FDM scheme. This isn't indicative of the proposed scheme not being effective, as it already has been shown that the proposed scheme improved on the relaxation process. In both D&C FDM cases, for relaxation steps before every 8th relaxation step, the number of SCF iterations increases (non-equilibrium atomic configurations also contribute to the fluctuations). On relaxation step number 8, there is a complete global density matrix calculation for both schemes. Electron

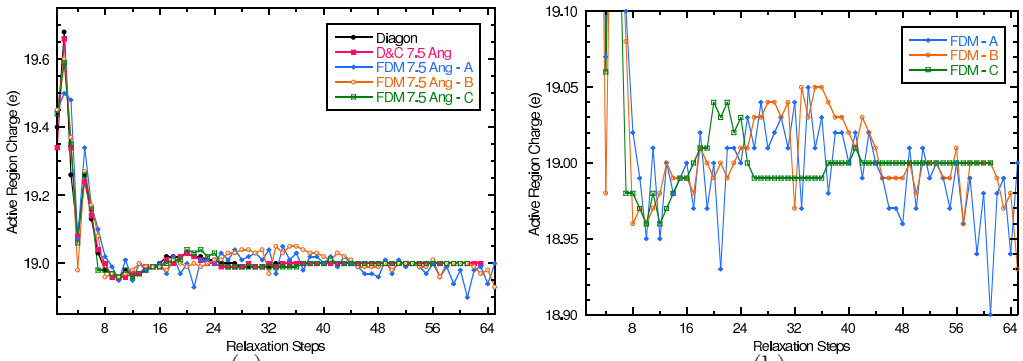


Figure 4.10: The charge of the active region calculated using a Mulliken population analysis a) for the first 64 relaxation steps with b) a close up view of a).

density is allowed to flow across the active-frozen region boundary, reducing and steadying the number of SCF iterations for the subsequent relaxation steps. For FDM_C (green line + empty squares), the tapering mechanism reduces the number of SCF iterations at all steps and is comparable to the Diagon scheme. It too also has a rise in the SCF count before every 8th relaxation step, although it is smaller than for the other FDM schemes. In between the peaks, the tapered FDM scheme has a near constant SCF iteration count, which even the Diagon scheme does not have. The tapering mechanism has helped with accelerating the relaxation process and the SCF process.

It has been found that the proposed scheme has improved upon the standard FDM scheme in terms of performance. To prove that the benefits were due to an improvement in charge transfer between the active region and frozen regions, the charge of the active region (the 10 unconstrained atoms) is plotted at each relaxation step for the first 64 steps in Figure 4.10(a). The charge of the active region was calculated using Mulliken population analysis [177]. All schemes follow the trend set by the Diagon scheme. The differences in the schemes are more easily discerned when examining the close up view shown in Figure 4.10(b). With the FDM_A scheme, large quantities of charge are transferred back and forth between the active and frozen regions at each relaxation step. This translates to unphysical charge at the boundary of the active and frozen regions. This is a major contributor to the lengthy relaxation process of the standard FDM

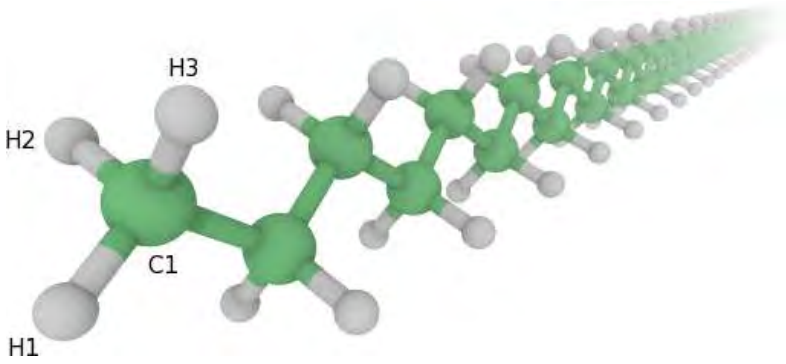


Figure 4.11: Local geometry optimisation of 194-atom linear alkane chain. The first 10 atoms from the labelled end are the only atoms allowed to relax. The labelled atoms are the atoms that are reported in the bond properties section. The green atoms are carbon and the grey atoms are hydrogen.

scheme. The proposed FDM scheme produces a more smooth and realistic charge transfer process. Using the proposed scheme, FDM_B, with multiple regions, the charge flow between relaxation steps is smoother than FDM_A. There is a definite improvement with using multiple regions, which in turn accelerates the relaxation process. The smoothest charge transfer occurs with the tapered FDM scheme, FDM_C. In fact, there is little or no charge transfer between relaxation steps 26 and 62. This illustrates the effectiveness of the tapering mechanism in eliminating spurious charge build up or depletion at the subsystem boundaries. It is not surprising that the average number of SCF iterations and the relaxation step count was the lowest out of all FDM schemes present here.

A summary of the bond properties of the FDM method is shown in Table 4.4. The results show the bond lengths and bond angles of the first 4 atoms located in the active region (refer to Figure 4.11 for the atom labelling scheme). All bond lengths produced by the D&C schemes were more or less equal to the bond lengths produced by the Diagon scheme. Larger differences are found in the bond angles. FDM_B and FDM_C improves upon the standard FDM scheme, FDM_A , and even the D&C scheme for the H1-C1-H1 and H1-C1-H3 angles. The FDM_B has larger deviations for the H2-C2-H3 and H1-C1-C2 bonds compared to the FDM_A and D&C schemes. The tapered FDM_C improves upon the FDM_B scheme, but is still not as precise as the FDM_A or D&C schemes.

The cause of the small differences in the bond angles for the D&C and FDM

Table 4.4: Comparison of geometries of the 194 atom C_nH_{2n+2} alkane chain. The D&C and D&C FDM schemes are relative to the Diagon scheme.

Bond lengths between active atoms (Å)						
Atom 1	Atom 2	Diagon	D&C	FDM_A	FDM_B	FDM_C
C1	H1	1.106	0.000	0.000	0.000	0.000
C1	H2	1.108	0.001	0.001	0.000	0.001
C1	H3	1.107	0.000	0.000	0.000	0.000
C1	C2	1.540	0.000	0.000	0.000	0.000

Bond angles between active atoms (degrees)							
Atom 1	Atom 2	Atom 3	Diagon	D&C	FDM_A	FDM_B	FDM_C
H1	C1	H2	107.708	0.014	0.050	0.003	0.010
H1	C1	H3	107.701	0.017	0.054	-0.005	-0.010
H2	C1	H3	107.485	-0.008	-0.030	-0.029	-0.024
H1	C1	C2	111.585	0.029	0.066	0.142	0.091

methods needs to be determined. The difference could arise due to the error in the D&C and FDM approximations or the bond angles could all be equally valid differences within the specified force tolerance of 0.01 eV/Å. To determine the cause, single-point standard diagonalisation calculations of the D&C and FDM optimised geometries were carried out. Table 4.5 shows the maximum constrained forces of the single-point calculations. The D&C and FDM_A geometries are found to be converged with a maximum constrained force equal to or below the force tolerance of 0.01 eV/Å. The FDM_B and FDM_C geometries are found to be slightly above the force tolerance. As there are no considerable differences in the bond lengths and angles and that the maximum constrained forces of the optimised FDM geometries are below or close to the specified force tolerance, it is concluded that the differences in the optimised geometries are primarily due to scatter associated with the magnitude of the force convergence criterion. The error associated with the D&C and FDM approximations is found not to be significant. Further verification is given by a D&C single-point calculation using the Diagon optimised geometry. The D&C method produced a maximum constrained force tolerance of 0.009 eV/Å, which is a converged geometry. As this is below the specified force tolerance, the error within the D&C method approximation is not

Table 4.5: Comparison of the maximum forces (eV/Å) of the 194 atom C_nH_{2n+2} alkane chain. The forces at the D&C and FDM optimised geometries were calculated using standard diagonalisation.

Diagon	D&C	FDM_A	FDM_B	FDM_C
0.009	0.010	0.008	0.013	0.014

significant in this simulation. This result is also applicable to the FDM schemes because the first relaxation step of the FDM scheme is equivalent to D&C.

4.3 Concluding Remarks

The standard D&C FDM method has been generalised by allowing each subsystem to have its own density matrix update rate. In effect, the designation of more than just the active and frozen regions is possible. This scheme can decrease the number of SCF iterations and relaxations steps during a simulation by allowing more effective electron density flow between active and non-active regions. Application of the tapering mechanism accelerated the SCF convergence and relaxation convergence with comparable accuracy with the FDM scheme. The issue with large memory requirements is still present, limiting this scheme to moderately sized systems.

5

Real-Time Divide-and-Conquer Time-Dependent Density Functional Theory

TDDFT has become a promising technique for the calculation of excited state properties of finite systems. Polarisability, hyperpolarisabilities, Raman intensities and other response properties can be obtained within the TDDFT formalism. The majority of TDDFT implementations, and in particular the real-time propagation method [34], can formally scale from $O(N^3)$ upwards, where N is some measure of the system size. The non-linear scaling is computationally prohibitive for the calculation of large systems, thus it is desirable to obtain a linear-scaling method for TDDFT calculations. The following work takes advantage of the locality inferred by the D&C method to solve the time-dependent KS equations in a linear-scaling fashion. The current state of the method produces divergences in the dipole moment, caused by anomalous subsystem boundary effects. Further work is required to eliminate the divergence.

5.1 Introduction

There has been a significant effort devoted to the development of order-N methods for ground state calculations. On the other hand, linear-scaling methods for the computation of response properties has attracted less attention. In this chapter

a new method is proposed to solve the time-dependent KS equations in real-time using a D&C paradigm. Currently the method works to some extent and with further developments based on the findings of this thesis it is anticipated the method will execute successfully.

With all physically intuitive linear-scaling methods, the one common aspect amongst these methods is to invoke the "near-sightedness" [18] principle. In the case of DFT, locality within the ground-state density is observed. That is, the density at a particular point does not greatly depend on the density at another other point some distance away. As found by the work of Chen and Mukamel [178], this principle also applies to the time-dependent density. Chen and Mukamel found locality within the first order induced density, $\delta\rho$; more specifically the off-diagonal elements in $\delta\rho_{ij}$ go to zero as the distance between the orbitals, i and j, increases for a π-conjugated polyene system. They also discovered that higher-orders of the induced density[a] have greater electronic coherence, which in turn make it more difficult to invoke any sense of locality.

The group of Liang, Yam, Yokojima and Chen have taken advantage of the locality of the first order induced density matrix to derive a linear-scaling technique which they have termed the Localised Density Matrix (LDM) method [179–182] and additionally for non-orthogonal basis sets have derived the Generalised LDM method [183]. They have also been able to apply the LDM method to non-linear response properties [184]. The premise of the LDM method lies with formulating the time-dependent KS equations directly in terms of the density, specifically the time-dependent KS equations are transformed into the von Neumann equation (analogous to the Liouville equation). Directly applying sparsity to the density matrix by setting to zero off-diagonal elements beyond a specified length, the main LDM equation can be propagated in a linear-scaling fashion.

The group of Niklasson, Weber and Challacombe derived a linear-scaling method based on an orbital-free quantum perturbation theory of the density matrix [185]. A generalised version utilising a perturbation-dependent non-orthogonal basis [186] was also developed. The group was able to calculate the static electric polarisability of a series of water clusters in a linear-scaling fashion [187]. These

[a]For this thesis, interest is only concerned with the first order response of the system.

methods are similar to the DMM methods [37, 38, 105–107] where the perturbation theory is based on the purification of a perturbed Hamiltonian. Kussmann and Ochsenfeld [188] present an alternative approach which allows for determining the response to a dynamic perturbation. They use a density matrix based reformulation and directly solve for the transition density matrices.

Linear-scaling fragment molecular orbital calculations have also been devised by Mochizuki and Ishikawa *et al* [189, 190]. A similar method to the fragment molecular orbital scheme was devised by Coriani *et al* [191]. Another known linear-scaling method is by Walker *et al* [192], which is a linear-scaling perturbation theory in the formalism of Casida's linear-response theory [193, 194].

At the time of writing this thesis, only one D&C based method is available for time-dependent quantum systems, namely time-dependent Hartree-Fock. Touma *et al* [195], calculate the frequency-dependent polarisability from the coupled-perturbed Hartree-Fock equation using the D&C paradigm. Their implementation does not scale linearly due to the construction of the Hamiltonian, although they mention that the construction can also be made to scale linearly. Their method can also be used with TDDFT formalism with necessary adjustments to handle the XC potential.

In this thesis, the approach taken is to solve the time-dependent KS equations in real-time using the D&C paradigm. This type of linear-scaling method has not been attempted, although similarities exist between the real-time LDM method in regards to localisation regions. Because this method is not based on quantum perturbation theory, it will allow the extraction of both linear and non-linear response properties more easily. The simultaneous calculation of the response to all frequencies is also available, which is not possible with the frequency-dependent perturbation theorems. A D&C approach will also allow the calculation of dynamic polarisabilities in regard to the electronic excitations of the complete system, where the fragment molecular orbital methods can only handle each subsystem separately. Fragment molecular orbital methods differ from the D&C method in that individual fragments are not coupled with each other, as the subsystems in D&C method are. The perturbation of an individual fragment is only possible with fragment molecular orbital methods. The success of the

LDM method and the D&C time-dependent Hartree-Fock method provides some justification that this method should also be successful.

Within the real-time TDDFT method there are certain aspects which can be made to scale linearly without applying the D&C paradigm. In particular, computing the inverse of the overlap matrix is a common operation which can be made to scale linearly [196, 197]. The use of non-D&C linear-scaling techniques might not justify the need for a D&C approach. However, these techniques can only be applied to certain approaches and aspects (specifically different propagators, see section 5.2.3) within the real-time method. When considering all approaches within the real-time method and considering that, more than likely, the matrices in question will be dense, a D&C approach becomes a viable option to produce a truly linear-scaling computation. When not considering linear-scaling methods, the dense matrices become an issue with matrix-vector and matrix-matrix multiplications. However, it is yet to be determined if the D&C method can provide a good approximation to these operations.

The following sections outline the details of the dynamics of electrons. The proposed DCTDDFT method will be described followed by the implementation details. The method is only partially successful, with attempts at eliminating possible subsystem boundary effects and the investigation of stability in the method being performed. The optical response is found and compared to the standard TDDFT method. Finally, a 2-dimensional partitioning scheme is proposed to alleviate the divergence in the dipole moment.

5.2 Electron Dynamics

The time evolution of a physical system is governed by the time-dependent Schrödinger equation, shown in (5.2). Finding a solution to this equation depends primarily on whether the Hamiltonian depends explicitly on time or not. For Hamiltonians which do not depend explicitly on time, the system is said to be conservative and obeys the conservation of energy law. The path to a solution in this case is simpler than those of the time-dependent Hamiltonian. The following formulation derives the dynamics for both types of Hamiltonians within

the Schrödinger picture[b].

5.2.1 Time-Independent Hamiltonians

When the physical system is conservative, and hence the Hamiltonian does not depend explicitly on time, the time evolution of the system can be found from the following.

Starting with the expectation value of any observable, $\hat{A}$, given by:

$$\langle \hat{A}(t) \rangle \equiv \langle \psi(t) | \hat{A} | \psi(t) \rangle = \int \psi^*(\boldsymbol{r}, t) \hat{A} \psi(\boldsymbol{r}, t) \mathrm{d}\boldsymbol{r} \tag{5.1}$$

The eigenstate, $|\psi(t)\rangle$, can be found by solving the time-dependent Schrödinger equation:

$$i \frac{\partial \Psi(\mathbf{r}, t)}{\partial t} = \hat{H} \Psi(\mathbf{r}, t) \tag{5.2}$$

Solutions of equation (5.2) may be represented by the eigensolutions of the time-independent Schrödinger equation:

$$\hat{H} \left| \phi_i \right\rangle = E_i \left| \phi_i \right\rangle \tag{5.3}$$

The eigenvectors, $|\phi_i\rangle$, and eigenvalues, E_i, form a complete orthonormal basis set in Hilbert space:

$$\langle \phi_i | \phi_j \rangle = \int \phi_i^*(\mathbf{x}) \phi_j(\mathbf{x}) d\mathbf{x} \tag{5.4}$$

$$= \delta_{ij} \tag{5.5}$$

Expanding the eigenvectors, $|\psi(t)\rangle$, in the set of basis functions, $|\phi_i\rangle$, leads to:

$$\left| \psi(t) \right\rangle = \sum_i \left| \phi_i \right\rangle \left\langle \phi_i | \psi(t) \right\rangle \tag{5.6}$$

Equation (5.6) is then substituted into (5.2) and multiplying by $\langle \phi_i |$ from the left

[b]The methodology used here is based primarily on the methodology presented by Mukamel [198], secondly on the work of Castro and Marques [199, 200] and thirdly on the work of Tsolakidis, Sánchez-Portal and Martin [201].

leads to the following equation;

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\phi_i|\psi(t)\rangle = -iE_i\langle\phi_i|\psi(t)\rangle \tag{5.7}$$

the solution to this equation is;

$$\langle\phi_i|\psi(t)\rangle = \exp\left[-i(t-t_0)E_i\right]\langle\phi_i|\psi(t_0)\rangle \tag{5.8}$$

where $\langle\phi_i|\psi(t_0)\rangle$ are the *initial* coefficients of the eigenfunction, which are known from the ground state solution. Multiplying by $\sum_i |\phi_i\rangle$, equation (5.8) becomes:

$$|\psi(t)\rangle = \sum_i \exp\left[-i(t-t_0)E_i\right]|\phi_i\rangle\langle\phi_i|\psi(t_0)\rangle \tag{5.9}$$

Equation (5.9) evolves the quantum state $|\psi(t_0)\rangle$ to a state $|\psi(t)\rangle$ within the specific representation of eigenstates, E_i, of the Hamiltonian, $\hat{H}$. As the initial conditions are varied, the Schrödinger equation must be re-solved. A more general time evolution scheme can be acquired by introducing the time evolution operator, $\hat{U}(t,t_0)$. The purpose of the time evolution operator is to transform the eigenstate at time, t_0, to an eigenstate at time, t, without having to solve the Schrödinger equation whenever the initial conditions are varied; it is defined as:

$$|\psi(t)\rangle = \hat{U}(t,t_0)|\psi(t_0)\rangle \tag{5.10}$$

Comparing (5.9) and (5.10) a definition of the time evolution operator can be found:

$$\hat{U}(t,t_0) = \sum_i |\phi_i\rangle\exp\left[-i(t-t_0)E_i\right]\langle\phi_i| \tag{5.11}$$

Equation (5.11) is in a representation consisting of the spectrum of eigenvalues of the Hamiltonian, $\hat{H}$. This representation is only useful for systems in which the eigenvalues are known. Ideally, one would prefer eigenvalues from analytic solutions. Unfortunately these are limited to a small selection of systems. Therefore, it would be beneficial to recast the evolution operator in terms of a more general form. By formulating the evolution operator in terms of the Hamiltonian, $\hat{H}$, a

more flexible evolution operator can be used to act on more complex systems, namely:

$$\hat{U}(t,t_0) = \sum_i |\phi_i\rangle \exp\left[-i(t-t_0)\hat{H}\right] \langle\phi_i| \tag{5.12}$$

5.2.2 Time-Dependent Hamiltonians

The preceding formulation applies only to time-independent Hamiltonians. This is only applicable if all the degrees of freedom are included within the Hamiltonian e.g. including the radiation field degrees of freedom as well as the material systems' degrees of freedom in the Hamiltonian. This situation is not always the case and it is more useful to replace some degrees of freedom with the addition of external forces into the Hamiltonian. When this happens, the Hamiltonian becomes time-dependent and a more complex approach to evolving the system is required.

Within the the context of TDDFT, the time-dependent KS Hamiltonian, (2.65), is used as the Hamiltonian in the following formulation.

The first step is to substitute (5.10) into (5.2):

$$\frac{\partial}{\partial t} U(t,t_0)\,|\psi(t_0)\rangle = -i\hat{H}(t)\hat{U}(t,t_0)\,|\psi(t_0)\rangle \tag{5.13}$$

This equation holds for any initial vector, $|\psi(t_0)\rangle$, resulting in:

$$\frac{\partial}{\partial t} U(t,t_0) = -i\hat{H}(t)\hat{U}(t,t_0) \tag{5.14}$$

Recasting (5.14) into integral form by integrating both sides in the interval from t_0 to t and using $\hat{U}(t_0,t_0) = 1$, results in:

$$\hat{U}(t,t_0) = 1 - i\int_{t_0}^{t} \hat{H}(\tau)U(\tau,t_0)\mathrm{d}\tau \tag{5.15}$$

One can solve (5.15) by iteratively substituting it into itself, forming a Dyson

series:

$$\hat{U}(t,t_0) = 1 + \sum_{n=1}^{\infty} (-i)^n \int_{t_0}^{t} \mathrm{d}\tau_n \int_{t_0}^{t_n} \mathrm{d}\tau_{n-1} \ldots \int_{t_0}^{t_2} \mathrm{d}\tau_1 \hat{H}(\tau_n)\hat{H}(\tau_{n-1}) \ldots \hat{H}(\tau_1) \tag{5.16}$$

Introducing the Dyson time-ordering operator, $\hat{\mathcal{T}}$;

$$\hat{\mathcal{T}}\hat{A}(t_1)\hat{B}(t_2) = \begin{cases} \hat{A}(t_1)\hat{B}(t_2) & \text{if } t_1 < t_2 \\ \hat{B}(t_2)\hat{A}(t_1) & \text{if } t_1 > t_2 \end{cases} \tag{5.17}$$

and operating on (5.16), the series takes on the form [199, 200]:

$$\hat{U}(t,t_0) = 1 + \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} \int_{t_0}^{t} \mathrm{d}\tau_n \int_{t_0}^{t} \mathrm{d}\tau_{n-1} \ldots \int_{t_0}^{t} \mathrm{d}\tau_1 \hat{\mathcal{T}} \left[\hat{H}(\tau_n)\hat{H}(\tau_{n-1}) \ldots \hat{H}(\tau_1) \right] \tag{5.18}$$

This new series resembles an exponential and due to this, it is recast in a simplified form, defining the time-ordered exponential:

$$\hat{U}(t,t_0) = \hat{\mathcal{T}} exp \left[-i \int_{t_0}^{t} \hat{H}(\tau)\mathrm{d}\tau \right] \tag{5.19}$$

The time-ordered exponential is only a cleaner and abbreviated form of (5.18). The difficult task of calculating the complete series, order by order, is still required. Of course exact calculations are intractable, requiring the use of certain approximations.

5.2.3 Time Evolution Operator

There are certain properties that any worthwhile approximation to the time evolution operator must adhere to and preserve. The first property states that for a Hermitian Hamiltonian, the time evolution operator is unitary, that is:

$$\hat{U}^{\dagger}(t + \Delta t, t) = \hat{U}^{-1}(t + \Delta t, t) \tag{5.20}$$

This property assures that the norm of the probability of the eigenstate is conserved. More precisely, as the eigenstate evolves, the probability amplitude of

the eigenstate remains equal to unity, maintaining conservation of energy within the system. The second property is time-reversal symmetry, as this is indicative of the stability and accuracy of the approximation:

$$\hat{U}(t+\Delta t,t)=\hat{U}^{-1}(t,t+\Delta t) \tag{5.21}$$

The final property breaks up the time evolution operator into smaller pieces providing an avenue for the actual calculation of the wavefunction evolution. Using the following sequence of time, with a timestep of Δt;

$$\Omega_{\Delta t}\left[t_0,t\right]=\left\{t_0,t_{k+1}=t_k+\Delta t,\left(k=0,1,2,\ldots,n\right),t_n=t\right\} \tag{5.22}$$

within the time interval, $[t_0,t]$, the time evolution operator can be split into smaller time intervals:

$$\hat{U}(t,t_0)=\hat{U}(t,t_n)\hat{U}(t_n,t_{n-1})\ldots\hat{U}(t_1,t_0) \tag{5.23}$$

Any approximation to the time evolution operator must strive to fulfil and maintain these properties throughout the evolution. The approximate time evolution operator used in this thesis is known as the Crank-Nicholson (CN) [202, 203] propagator [200]:

$$\hat{U}_{CN}(t+\Delta t,t)=\frac{1-\frac{i}{2}\Delta t\hat{H}(t+\Delta t/2)}{1+\frac{i}{2}\Delta t\hat{H}(t+\Delta t/2)} \tag{5.24}$$

The stability of the solution can be increased by including more terms in the expansion of the propagator. The third-order CN propagator [201] is defined as:

$$\hat{U}^{3rd}_{CN}(t+\Delta t,t)=\frac{1-\frac{i\Delta t}{2}\hat{H}(t+\frac{\Delta t}{2})-\frac{1}{2}(\frac{\Delta t}{2}\hat{H}(t+\frac{\Delta t}{2}))^2+\frac{i}{6}(\frac{\Delta t}{2}\hat{H}(t+\frac{\Delta t}{2}))^3}{1+\frac{i\Delta t}{2}\hat{H}(t+\frac{\Delta t}{2})-\frac{1}{2}(\frac{\Delta t}{2}\hat{H}(t+\frac{\Delta t}{2}))^2-\frac{i}{6}(\frac{\Delta t}{2}\hat{H}(t+\frac{\Delta t}{2}))^3} \tag{5.25}$$

It is an implicit method and belongs to the family of classical propagators. For time-independent Hamiltonians it is unitary, preserves time-reversal symmetry and exactly conserves energy. Stability of the propagator occurs for cases when $\Delta t\Delta \mathrm{E}_{max}\ll 1$, where E_{max} is the range of the eigenvalue spectrum of the cor-

responding Hamiltonian. The propagator has a scaling of $O(N^3)$, predominately due to the matrix inversion.

The actual propagation of the eigenstates is achieved using the real-time approach developed by Yabana and Bertsch [34], that allows the calculation of response of the system for all frequencies.

5.3 Real-Time Propagation

The real-time propagation method is a physically intuitive way to solve the time-dependent KS equations. It works by perturbing the ground-state with an external potential, $\hat{H}_{ext} = \mathbf{E} \cdot \boldsymbol{r}$, and then explicitly follows the evolution of the system by solving the time-dependent KS equations in real-time. That is, the eigenstates are evolved in time (using a propagator that adheres to the properties mentioned earlier), a new density matrix is constructed from the newly propagated eigenfunctions, which in turn is used to create a new corresponding Hamiltonian. Using the new Hamiltonian, the eigenfunctions are propagated to a future time and the whole process is repeated. Once the propagation of the eigenstates is complete, the induced density, $\delta\rho(\boldsymbol{r}, t)$, and the induced dipole moment, $\mathbf{D}(t)$, are found. By taking the Fourier transform of $\mathbf{D}(t)$, the response of the system can be calculated for all frequencies simultaneously, provided the propagation time can account for all frequencies i.e. to retrieve longer wavelengths, longer propagation times are required. The maximum frequency, ω_{max}, attainable using this method is determined by the timestep, Δt, not being larger than $\approx 1/\omega_{max}$.

The induced density can only directly provide so much information about the final state. Linear response theory is required to extract the appropriate properties. One of the most important properties to examine is the polarisability. It is a response function that describes the distortion of the electron cloud due to an applied external electric field and provides a direct correlation between the response of the electrons and optical properties. The frequency-dependent linear

polarisability, $\alpha(\omega)$, is defined as;

$$\alpha(\omega) = \frac{e^2\hbar}{m}\int_0^\infty \frac{S(\omega')d\omega'}{\omega'^2 - \omega^2} \tag{5.26}$$

where e is the electric charge, $\hbar$ is the reduced Planck constant, m is the mass of the electron and $S(\omega)$ is the dipole strength function, The dipole strength function is defined as;

$$S(\omega) = \frac{2m}{\pi e^2\hbar}\omega Im\left\{\alpha(\omega)\right\} \tag{5.27}$$

and is normalised to the number of electrons, N_e, according to the sum rule, $\int_0^\infty \mathrm{d}E\, S = N_e$. Because $S(\omega)$ is proportional to the photoabsorption cross-section, $\sigma(\omega)$, a direct comparison with experiment is possible, defined as;

$$\sigma(\omega) = \frac{2\pi}{c}\omega Im\left\{\alpha(\omega)\right\} \tag{5.28}$$

with c being the speed of light. To calculate $\sigma(\omega)$, at every time step the induced dipole moment, $\mathbf{D}(t)$, is calculated by summing over the occupied orbitals:

$$\mathbf{D}(t) = \sum_{i=1}^{occ}\langle\psi_i(t)|\boldsymbol{r}|\psi_i(t)\rangle \tag{5.29}$$

Then by taking the Fourier transform of $\mathbf{D}(t)$;

$$\mathbf{D}(\omega) = \int dt\, e^{i\omega t - \delta t}\,\mathbf{D}(t) \tag{5.30}$$

with δ being a damping factor to introduce peak broadening of the imaginary part of the response, $\mathbf{D}(\omega)$ can be used with the linear polarisability given by $\mathbf{D}(\omega) = \alpha(\omega)\mathbf{E}(\omega)$ (where $\mathbf{E}$ is the electric field), leading to the final relationship;

$$Im\left\{\alpha(\omega)\right\} = \omega\,\frac{Re\left\{D(\omega)\right\}}{E} \tag{5.31}$$

Once the frequency-dependent polarisability tensor is calculated, the calculation is repeated with the electric field aligned to different axes. The average linear

polarisability is then given by:

$$\langle \alpha(\omega) \rangle = \frac{1}{3} Tr \, \{\alpha_{ij}(\omega)\} \tag{5.32}$$

5.3.1 Real-Time Propagation in SIESTA

The real-time method implementation within SIESTA [201] follows the approach described in section 5.3. The only slight modifications come from employing an atomic orbital expansion, which in turn transforms the time-dependent KS equations into the following form;

$$i\frac{\partial C}{\partial t} = S^{-1}\hat{H}C \tag{5.33}$$

where C are the coefficients of the orbitals and S is the overlap matrix. Using the expansion of the inverse of the overlap matrix [100];

$$\mathbf{S}^{-1} = [\mathbf{I} - (\mathbf{I} - \mathbf{S})]^{-1} \approx \sum_{n=0}^{k} (\mathbf{I} - \mathbf{S})^n \tag{5.34}$$

where in the first-order, $(k = 1)$;

$$\mathbf{S} \approx 2\mathbf{I} - \mathbf{S} \tag{5.35}$$

the CN propagator then takes the form:

$$\hat{U}_{CN}(t + \Delta t, t) = \frac{S - \frac{i}{2}\Delta t \hat{H}(t + \Delta t/2)}{S + \frac{i}{2}\Delta t \hat{H}(t + \Delta t/2)} \tag{5.36}$$

5.4 Real-Time Divide-and-Conquer TDDFT

The proposed real-time D&C TDDFT (DCTDDFT) method uses the real-time method of Yabana and Bertsch [34], where instead of propagating the complete set of eigenfunctions, the system is partitioned into overlapping subsystems and the eigenfunctions of each subsystem are separately propagated. The electronic information obtained for each subsystem is then combined in a specific way as

to provide an approximation to the global density matrix at each propagation step. The speedup in calculation time occurs because each subsystem is solved separately with a cost that no longer depends on the size of the global problem. The individual subsystems are coupled to each other by a common Fermi level set at time t_0. It is currently unknown if there are any time-dependent constraints when considering the system to be at equilibrium with respect to the chemical potential at all times, when there is also a time-dependent contribution to consider. It will be shown that the method works under certain conditions and that the method does not diverge immediately for conditions when it fails, indicating that the method partially works.

Within DCTDDFT the global time-dependent density matrix is divided up into individual subsystem density-matrices weighted by a normalised partition function;

$$\sum_{\alpha} \mathbf{P}_{ij}^{\alpha} = 1, \tag{5.37}$$

where α is the subsystem index, and where i and j are orbital indices. The partition function, $\mathbf{P}_{ij}^{\alpha}$, is defined by a Mulliken-type [129] weight matrix (suitable for subsystems consisting of one core atom):

$$\mathbf{P}_{ij}^{\alpha} = \begin{cases} 1 & \text{if } i \in \alpha \text{ and } j \in \alpha \\ 1/2 & \text{if } i \in \alpha \text{ and } j \notin \alpha \\ 0 & \text{if } i \notin \alpha \text{ and } j \notin \alpha \end{cases} \tag{5.38}$$

Defining the time-dependent Kohn-Sham electron density;

$$\rho(\mathbf{r}, \mathbf{r}', t) = 2 \sum_{m}^{N/2} \psi_m(\mathbf{r}, t) \psi_m(\mathbf{r}', t) = \sum_{ij} \rho_{ij}(t) \phi_i(\mathbf{r}) \phi_j(\mathbf{r}') \tag{5.39}$$

where the electron density is defined in the space of the Kohn-Sham orbitals, $\{\psi_m(\mathbf{r}, t)\}$. The density matrix, $\rho_{ij}(t)$, is defined in the atomic orbital space, $\{\phi_i(\mathbf{r})\}$, and is given by the linear coefficients, $\{C_{im}(t)\}$, as follows:

$$\rho_{ij}(t) = 2 \sum_{m}^{N/2} C_{im}(t) C_{jm}(t) \tag{5.40}$$

The density matrix can be divided into subsystem contributions. The density matrix is then a sum of contributions from all subsystems, weighted by the partition matrix:

$$\rho_{ij}(t) \equiv \sum_{\alpha} \mathbf{P}^{\alpha}_{ij}\rho_{ij}(t) = \sum_{\alpha} \rho^{\alpha}_{ij}(t) \tag{5.41}$$

The local nature of the density matrix allows each subsystem density matrix contribution to be approximated by;

$$\rho^{\alpha}_{ij}(t) = 2\mathbf{P}^{\alpha}_{ij} \sum_{m} f_{\beta}(\epsilon_F - \epsilon^{\alpha}_m) C^{\alpha}_{im}(t) C^{\alpha}_{jm}(t) \tag{5.42}$$

where f_β is the Fermi function approximating an occupation number, β is the inverse electronic temperature, ϵ_F is the Fermi level common to all subsystems and ϵ_m is the orbital energy. The approximate occupation number, f_β, is calculated once at t_0 and kept constant throughout the propagation.

Propagation of the subsystem eigenfunction coefficients is accomplished by using the subsystem CN propagator, $\hat{U}^{\alpha}_{CN}$;

$$\hat{U}^{\alpha}_{CN}(t + \Delta t, t) = \frac{S^{\alpha} - \frac{i}{2}\Delta t \hat{H}^{\alpha}(t + \Delta t/2)}{S^{\alpha} + \frac{i}{2}\Delta t \hat{H}^{\alpha}(t + \Delta t/2)} \tag{5.43}$$

where $\hat{H}^{\alpha}$ is the subsystem Hamiltonian and S^{α} is the subsystem overlap matrix.

5.4.1 Implementation Details

The real-time DCTDDFT method is implemented to work in conjunction with the SIESTA package. The linear-scaling assembly of the Hamiltonian and overlap matrices are handled by SIESTA. Partitioning and communication of the subsystem data is readily handled by the ground state D&C code, see chapter 3. The DCTTDFT code is based on the original SIESTA TDDFT code of Tsolakidis *et al* [201]. The optical response code was provided by Dr. Daniel Sánchez-Portal.

5.4.2 Algorithm

The general overview of the DCTDDFT implementation within the SIESTA code is represented in a flowchart shown in Figure 5.1. The flowchart has been appropriately marked to indicate which parts of the code involve the original SIESTA routines (solid box), parallel communication (dotted box) and the present DCTDDFT module (dashed box). The algorithm begins by reading the spatial locations of all atoms and options to perform the DFT run. Once the atom specifics have been read into SIESTA it will distribute the atom information across the compute nodes according to a domain decomposition algorithm. In short, each compute node will be responsible for a subset of orbitals localised in a region of space and all the corresponding electronic information pertaining to those orbitals. Each node then generates the elements of the Hamiltonian and overlap matrices that it is uniquely responsible for. The Hamiltonian includes a perturbation from a static homogeneous electric field. The ground state density matrix is then calculated using the D&C method, as described in section 3.4. The DCTDDFT section of the code then begins from this point.

Before the first propagation step, the system will be divided into subsystems, α. This entails creating a data structure to store the orbital information for each subsystem with lists distinguishing the core and buffer atoms. If running in parallel, the matrix elements belonging to core or buffer orbitals that reside on other compute nodes need to be communicated to the nodes with ownership of subsystems requiring that data. Because of the spatial locality of the domain decomposition, the number of compute nodes that communicate with each other should remain constant or decrease as the system size increases, according to whether the number of processors employed scales with the system size or remains fixed, respectively.

Once all the non-local data has been transmitted, the initial eigenfunction coefficients of each subsystem are calculated. These subsystem eigenfunction coefficients are then stored in a data structure separate from the D&C data structures. It should be noted that only the *occupied* eigenfunctions are cached and propagated. The subsystem Hamiltonians are also stored, as they will be used in the next cycle to extrapolate a Hamiltonian at a future time.

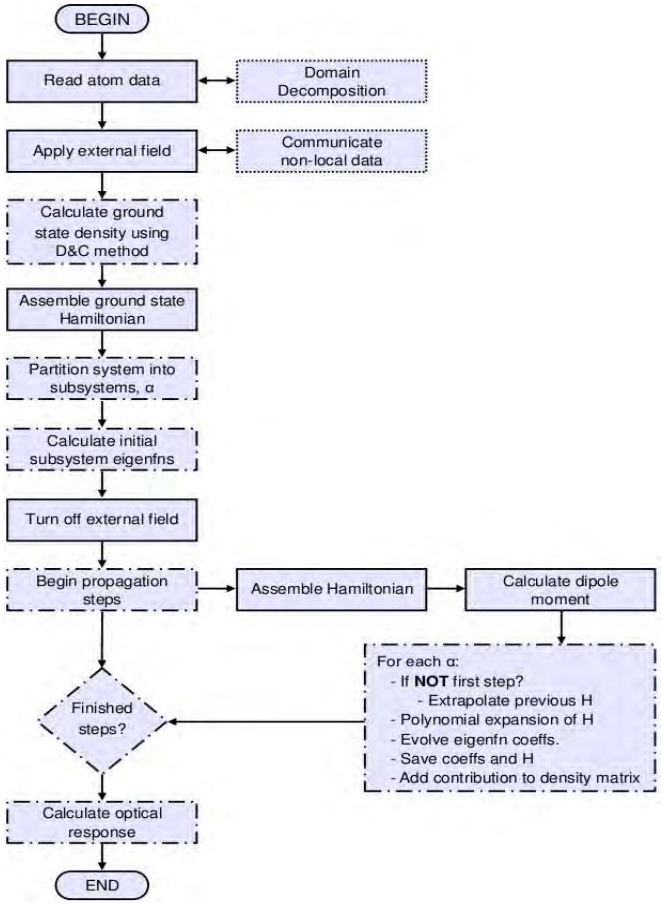


Figure 5.1: Schematic outlining the major implementation sections and process flow for the implementation of DCTDDFT within the SIESTA code. The original SIESTA routines are represented with solid boxes, the newly implemented DCTDDFT modules are shown in boxes with dashed lines and any parallel communication modules are shown in boxes with dotted lines.

The external field is switched off, which sets the electrons in motion as they oscillate about their equilibrium state. The propagation cycles begin at this stage. The number of propagation steps and the duration of each step, Δt, is specified by the user and will be specific to the type of system. In general, Δt will be small ($\leq 0.002 fs$) when using the DCTDDFT method so as to ensure stability of the propagator. The propagation algorithm evolves the eigenfunctions of each subsystem individually, reducing the memory footprint of the propagation algorithm.

For each propagation cycle a new global Hamiltonian is calculated, the in-

duced dipole moment for the current time is calculated and then each subsystem eigenfunction is propagated to a future time using the CN propagator. The newly evolved coefficients replace the previous coefficients and the subsystem Hamiltonian is cached for the next propagation step. These steps occur for each subsystem. A new density matrix is assembled from the propagated eigenfunctions in a similar fashion to assembling the ground state density matrix, with appropriate partition weights for the core and buffer orbitals. The subsystem Hamiltonian from the previous step is used in an extrapolation with the current subsystem Hamiltonian, to determine $\hat{H}(t + \Delta t/2)$ as used by the CN propagator. For a more accurate result, a more advanced predictor-corrector scheme can be used, although, *in theory* this is not necessary as using a smaller time step can produce accurate results.

Once the propagation is completed, the induced time-dependent dipole moment is used to calculate the optical response, allowing a direct comparison with experiment.

5.4.3 Memory Considerations

Each subsystem stores the dense 2-dimensional subsystem eigenfunction coefficients and the previous 2-dimensional subsystem Hamiltonian matrix. The scaling of the memory is then approximately $\sum_i^{N_\alpha} [(N_m^\alpha)^2 + (N_m^\alpha N_{occ}^\alpha)]$, where N_α is the number of subsystems, N_m^α is the total number of orbitals in the subsystem (subsystem Hamiltonian storage) and N_{occ}^α is the number of occupied orbitals (subsystem eigenfunction coefficients storage). Generally, for most common systems, N_{occ}^α is approximately half of N_m^α. For a linear molecule, such as the alkane molecule studied in section 5.5, the scaling does not cause difficulties. For larger systems the memory scaling will inhibit the use of this method, although modern computer architectures are equipped with large amounts of memory. With hardware increases in the future, the method can easily be applicable to very large systems.

There are ways to reduce the memory scaling. The first step is to reduce the number of subsystems, N_α, by using multiple core atoms within each subsystem, instead of Mulliken-type single core atom per subsystem partitioning.

Depending on the system in question, the calculation time will also be significantly reduced. By decreasing N_α, N_m^α will increase, although this increase is a smaller factor than the factor N_α for very large systems. The second step is to reduce the scaling of the $N_m^{\alpha^2}$ factor. As stated earlier, this factor is due to the storing of the subsystem Hamiltonian matrix used with the extrapolation scheme found within the propagation scheme. Taking advantage of the sparsity of the Hamiltonian, this matrix can be represented in sparse matrix form and hence reduce to linear scaling, in the limit of large subsystems. The final term in the scaling ($N_m^\alpha N_{occ}^\alpha$) is not altered as this represents the scaling due to the storage of the eigenfunction coefficients. Incorporating these memory saving steps would not drastically change the computing load and in most cases the reduction of subsystems by using multi-core subsystems will decrease the computational cost required to propagate each quantum state.

5.4.4 Parallelisation

The parallelisation of the real-time DCTDDFT implementation uses the same paradigm as the ground state D&C implementation. As with the D&C implementation, the domain decomposition load balancing scheme included within the SIESTA package, specifically designed for the KMG order-N method, is used to distribute the atoms amongst the compute nodes. The contributions to the Hamiltonian, overlap and density matrices from each atom are then stored on the corresponding compute nodes. With this scheme the only global communication occurs when constructing the density matrix; see section 3.5.3 for more information.

5.5 Results

The insulating 1D linear alkane molecule (C_nH_{2n+2}) is a favourable case for ground state linear-scaling methods as a closed-shell, wide gap material with low dimensionality, which should translate to the dynamic case. The first set of results uses a 26-atom molecule while the rest of the results use a 194-atom molecule.

Calculations were performed in parallel using 8 cpus for the DCTDDFT simulations and on a single cpu for the conventional TDDFT simulations[c]. The machines used were a 32 processor SGI Altix machine (1.6 GHz) with 64 GB of RAM and a 1472-processor SUN Constellation (2.93GHz Intel Nehalem cpus) machine. The calculations were carried out using a 150 Ry cut-off for the real-space integration grid used to represent the density, an energy shift of 0.02 Ry for the PAO orbital confinement and a density matrix convergence criteria of 1 x 10^{-4} for self-consistency. The PBE [54] form of the GGA was used for the XC functional in the adiabatic approximation. A SZ basis set[d] is used in all sets of calculations. Norm-conserving Troullier-Martins pseudopotentials [88, 91] in the Kleinman-Bylander factorised form [89, 90] were used. The choice of using a low quality basis set might not provide an accurate representation of the alkane molecule, though it does allow for relatively fast simulations. The eigenstates were propagated for 10 fs with a time step of 0.002 fs and an initial external electric field perturbation of 0.1 V/Å.

The first task is to verify that the DCTDDFT method works as intended, in particular concerning the application of the partition function (5.38) to a time-dependent density matrix. This is achieved by running a calculation for a 26-atom linear alkane molecule. The external electric field is orientated along the direction of the molecule (x-axis). The x-axis component of the dipole moments of the conventional TDDFT (black line) and the DCTDDFT method (blue line) are plotted in Figure 5.2. Each subsystem of the DCTDDFT method encapsulates the complete system, which in this limit, the method should be equal to the conventional method. The dipole moments of both methods are equal, which validates that the proposed method is conceptually correct in terms of using a partition function to assemble the global density matrix.

[c]Conventional TDDFT calculations were achieved by creating a single subsystem consisting of only core atoms which encapsulated the complete system.

[d]Although using such a small basis set will not produce accurate results for the system in question, the aim of the following work is to assess the performance of the DCTDDFT method and not produce chemically accurate results.

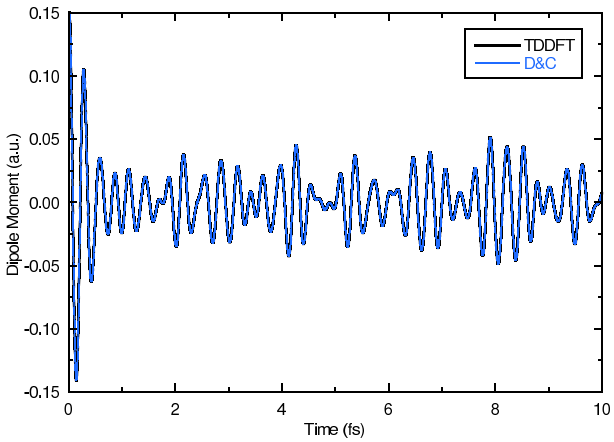


Figure 5.2: The induced dipole moment for a 26-atom linear alkane molecule, using standard TDDFT propagation (black line) and DCTDDFT propagation (blue line). Each DCTDDFT subsystem encapsulates the complete system. The dipole moments are equal to each other at all times. Thus, providing evidence that the DCTDDFT method is equivalent to the standard method in this limit.

5.5.1 Polarisation Direction

An investigation is performed on a larger 194-atom linear alkane molecule. The orientation of the external electric field is aligned normal to the alkane molecule along the y-axis. Here, the electrons are expected to stay close to their ground state equilibrium positions, as charge can not flow easily in a direction normal to the molecule. If the DCTDDFT method can not handle this scenario then the method might not handle the more common scenario when the electrons flow along the molecule. Figure 5.3 shows the dipole moment along the y-axis for a 25 Å subsystem (blue line), 25 Å subsystem plus a 5 Å outer buffer region (pink line) and 25 Å subsystem with a 25 Å outer buffer region (green line). The DCTDDFT results are compared with the standard TDDFT (black line) dipole moment. The use of large subsystems is to ensure that the method works as intended and that the accuracy is not questioned. The 25 Å subsystem corresponds to the converged localisation region sizes of ≈ 25 Å when using the LDM method [179–182] for this system. All DCTDDFT cases show good convergence with the conventional case. The 25 Å subsystem and 25 Å plus 5 Å subsystems have a slightly larger magnitude after 4 fs, although they are still in phase with

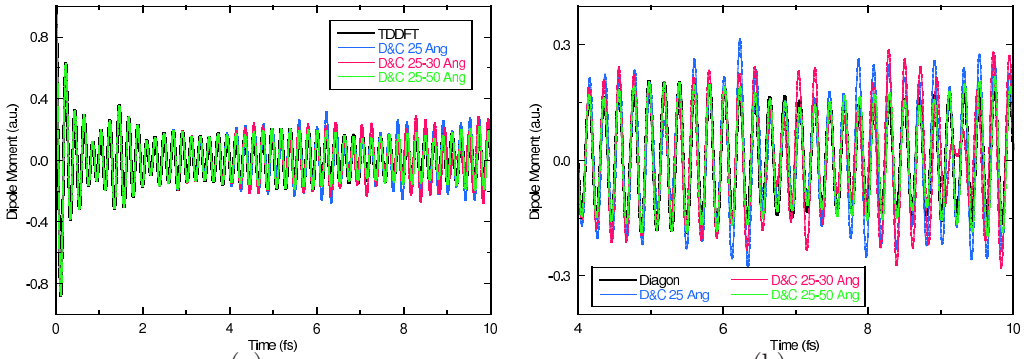


Figure 5.3: A comparison of dipole moments calculated from conventional TDDFT and DCTDDFT for a 194-atom linear alkane molecule with the external electric field perturbation applied perpendicular to the linear molecule. a) A complete view of the time-dependent dipole moment. b) A close up view of the time-dependent dipole moment when slight differences occur between the methods.

the conventional case. The 25 Å plus 5 Å subsystem case has a slight difference (resembling a beat) at ≈ 9.3 fs compared to the conventional case. The 25 Å plus 25 Å subsystem is equivalent to the conventional case. In this case, when the electrons are more confined, the DCTDDFT method works as intended, producing accurate dipole moments.

As the propagation proceeds to long time lengths, longer wavelengths of density fluctuations can be accurately extracted from the real-time method. It is unclear the effect of the partitioning will have on density fluctuations with wavelengths greater than the subsystem size. The small differences in the dipole moment found after 4 fs are most likely due to the approximation made in representing long wavelength fluctuations across subsystem borders with the partition function (3.11).

These are promising results for the DCTDDFT method. The following investigations are based on the electric field being aligned parallel to the molecule. In this case, the DCTDDFT method is tested when electrons flow along the molecule and across subsystem boundaries.

5.5.2 Subsystem Size Dependence

The next set of results looks at the dipole moments of the 194-atom linear alkane molecule for varying subsystem radii, with the external field oriented along the direction of the molecule in the x-axis. All following plots of the dipole moment will only be concerned with the x-component, unless otherwise specified. Figure 5.4 shows the results for subsystem radii ranging from 10 Å to 20 Å. The plots in the right column are close up views of the plots in the left column. The same format is followed for subsystem radii ranging from 25 Å to 35 Å in Figure 5.5 and for the 50 Å subsystem radius shown in Figure 5.6. The plots indicate that the dipole moment of the DCTDDFT method is approximately equal to the conventionally found dipole moment up until some time in the propagation when the dipole moment diverges. The specific time the divergence occurs is dependent on the size of the subsystem, with larger subsystems diverging at later times. For instance, a subsystem with radius 15 Å diverges at $\approx$ 1.3 fs while a subsystem with radius 30 Å diverges at $\approx$ 2.7 fs. It is expected that the localisation region in the time-dependent density would be larger than the ground state density [178], though the method still diverges even for a very large subsystem radius of 50 Å, at a time of $\approx$ 4.5 fs.

The divergence obviously limits the use of this method. Realistically, only the dipole moment up to the point of the divergence can be used. In most cases this is a short period, which in effect will reduce the resolution of the optical response calculations. Long propagation times are required to be able to extract the majority of the frequency response. The current results suggest that a very large subsystem radius is required to be able to propagate the system for a sufficient enough time, which is not desirable or more likely not possible.

There are a couple of points to be taken from these results. The first is the magnitude of the divergent peaks begin relatively small then generally increases in size during the propagation. This seems to be indicative of an unphysical density (and not obeying the conservation of energy), either caused by anomalous effects with the subsystem boundary or stability issues with the method. This knowledge can direct the effort of finding a possible solution. The second point is to do with the convergence of the dipole moment as the subsystem size is increased.

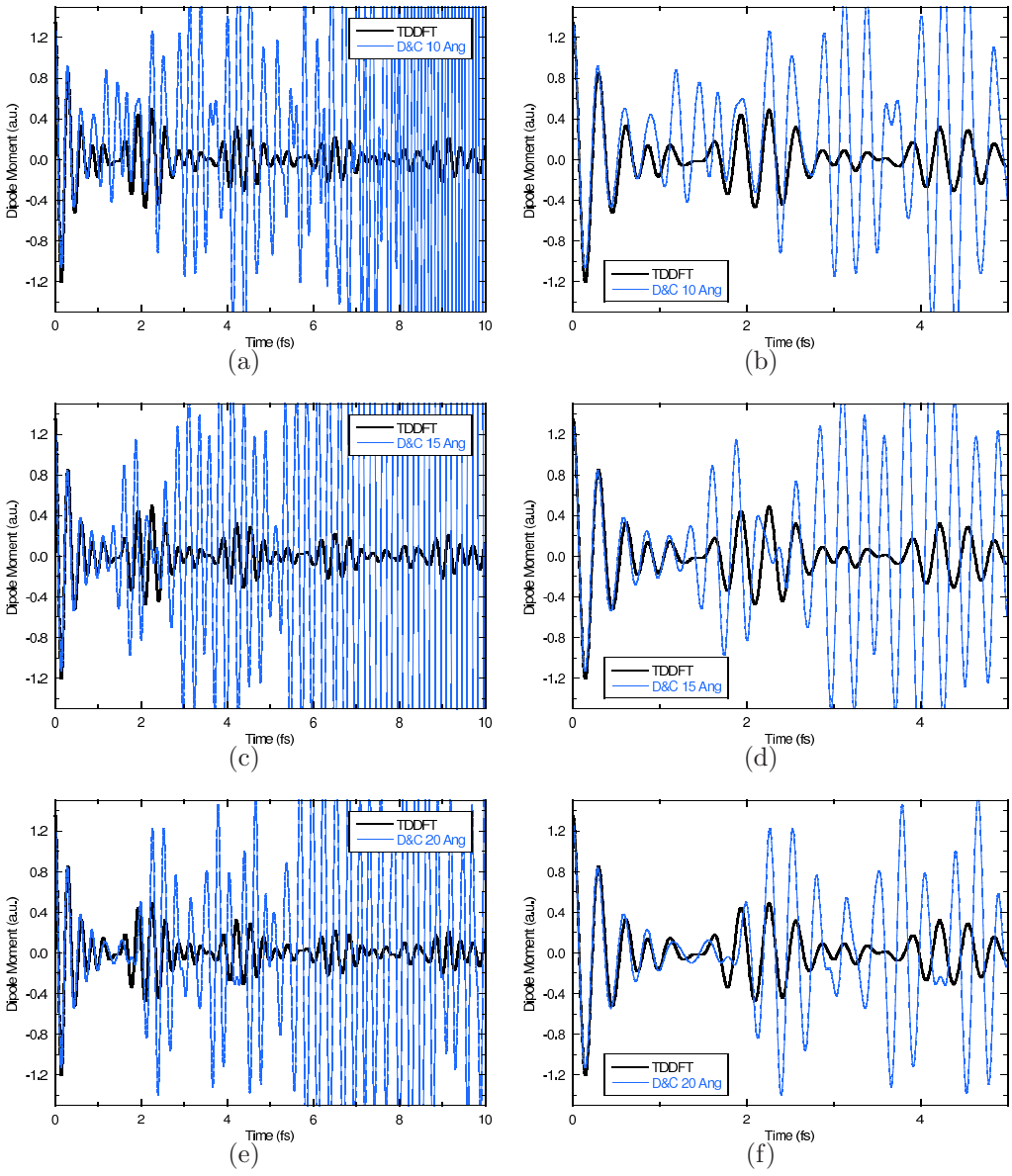


Figure 5.4: The induced dipole moments of a 194-atom linear alkane molecule using the DCTTDFT method with various subsystem sizes (blue lines) compared to the standard TDDFT method (black line). Figure a) uses a 10.0 Å subsystem, Figure c) uses a larger 15.0 Å subsystem and finally Figure e) uses a 20.0 Å subsystem. Figures b), d) and f) show a close up view of a), c) and e), respectively. The point in time of the divergence of the DCTDDFT method increases as the subsystem radius is increased.

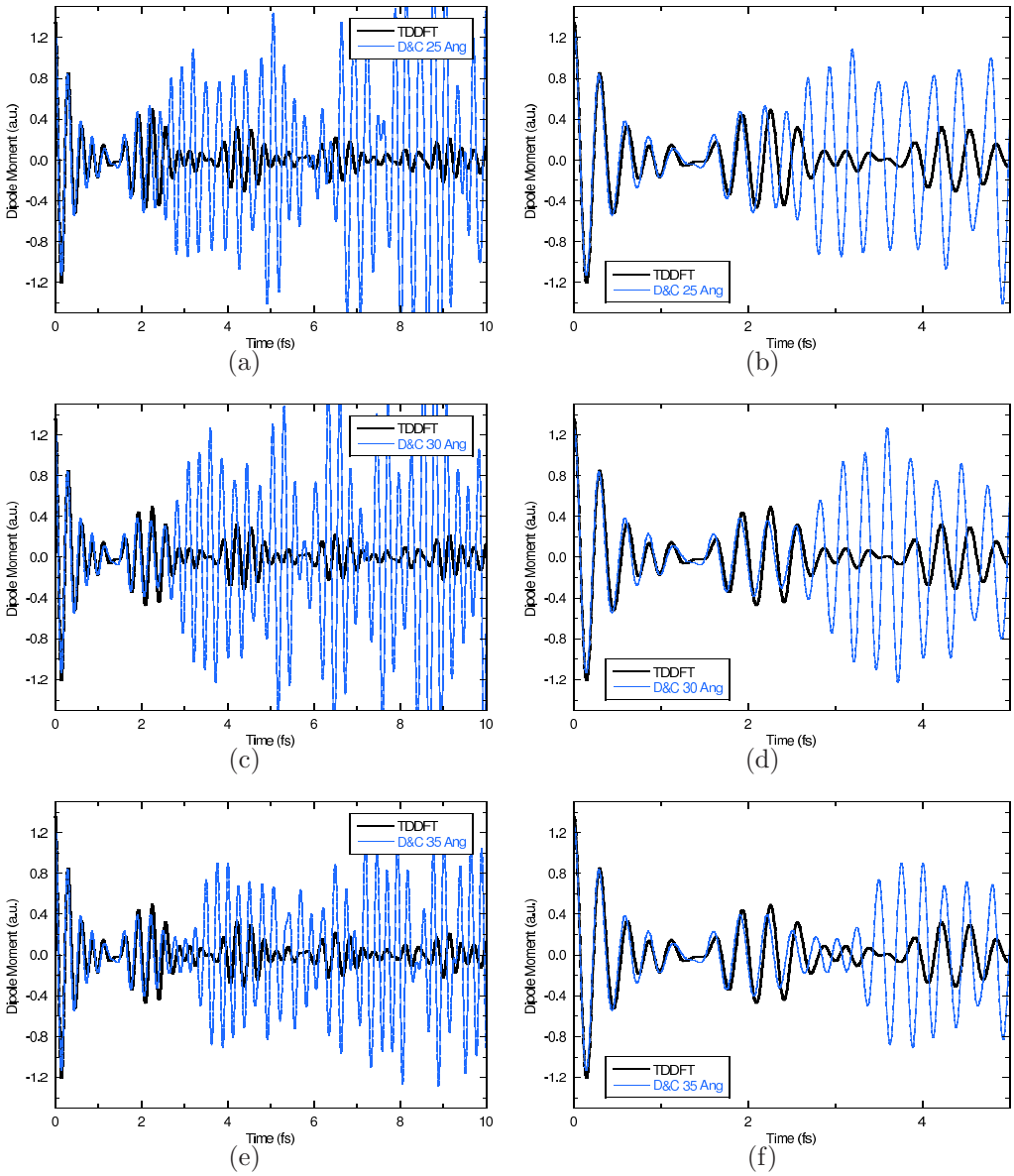


Figure 5.5: The induced dipole moments of a 194-atom linear alkane molecule using the DCTTDFT method with various subsystem sizes (blue lines) compared to the standard TDDFT method (black line). Figure a) uses a 25.0 Å subsystem. Larger subsystems are also shown with Figure c) using a 30.0 Å subsystem and Figure e) using a 35.0 Å subsystem. Figures b), d) and f) show a close up view of a), c) and e), respectively. The point in time of the divergence of the DCTDDFT method increases as the subsystem radius is increased. All subsystem radii above 25.0 Å can be considered equivalent up to ≈ 2.5 fs, indicating that convergence in the locality aspect of the method is reached with a subsystem radius of 25.0 Å and that the divergence is caused by other factors.

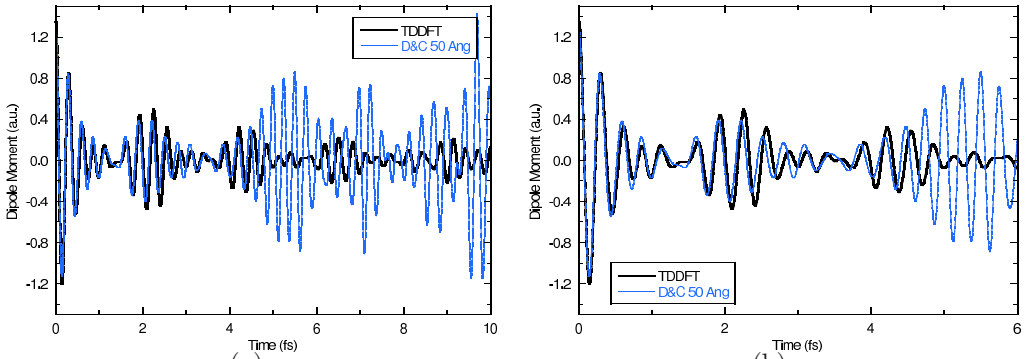


Figure 5.6: The induced dipole moments of a 194-atom linear alkane molecule using the DCTTDFT method with a 50 Å subsystem radius (blue line) compared to the standard TDDFT method (black line). a) the full time evolution. b) close up view, up to 6 fs.

Disregarding the diverged sections of the dipole moment, the DCTDDFT method for this particular system converges, with respect to the dipole moment, at a subsystem radius of $\approx$ 25 Å. That is, for subsystem radii greater than 25 Å, the dipole moments are considered equal, up to the point the 25 Å subsystem calculation diverges (i.e. no improvements to the accuracy of the method are made with increasing the subsystem radius above 25.0 Å). For instance, the 50 Å results are equal to the 25 Å results up to a time of $\approx$ 2.5 fs. Hence, if the method worked for all times, then a subsystem radius of 25 Å would be sufficient for converged (in the dipole moment) results. For converged results, the size of the subsystems are consistent with findings of the localisation regions in the LDM method [179–182], for similar linear alkane molecules. This validates the fact that the method converges at a subsystem radius of 25.0 Å and that any divergences at a later time are caused by other factors. Larger subsystems are required than in the ground state case, due to the longer coherence lengths in the excited case. Large subsystems increase the memory usage and the prefactor of the method. Reductions in the prefactor and memory usage can be made with using subsystems with non-overlapping multi-atom core regions. This type of partitioning reduces the overlap in the buffer regions between neighbouring subsystems.

Before continuing to try to find the cause of the divergence, it is important to

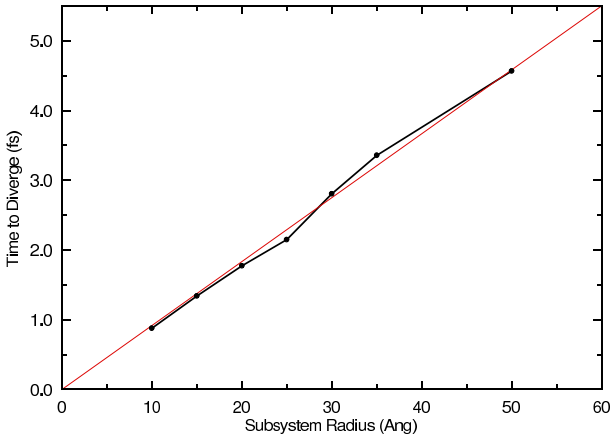


Figure 5.7: The approximate time required for DCTDDFT method to diverge for increasing subsystem sizes. The near-linear relationship can be interpreted as the time required for the subsystem density interacting with the subsystem border to contribute to the global density and cause the divergence in the dipole moment.

understand the relationship between the time for the dipole moment to diverge and the subsystem size. This relationship is found to be nearly linear, as shown in Figure 5.7, and is a strong indication that the divergence in the dipole moment is caused by the subsystem boundary. It can be considered that as the subsystem eigenfunction coefficients evolve in time, there can be errors caused by subsystem boundary effects (as in the ground-state case). These errors will take a finite time, dependent directly on the size of the subsystem, to reach the core of the subsystem where the largest contributions to the global density is made. Once these errors reach the core, the density will become unphysical and cause the dipole moment to diverge.

In the following, an investigation of the stability of the CN propagator and the DCTDDFT method takes place to rule out any possible links to this aspect of the method. Subsequently, the effects of the subsystem boundary on the propagation is investigated as the *probable* source of error.

5.5.3 Stability

The first aspect of the method to check regarding the source of error in the dipole moment is the stability of the DCTDDFT method and the CN propagator. In

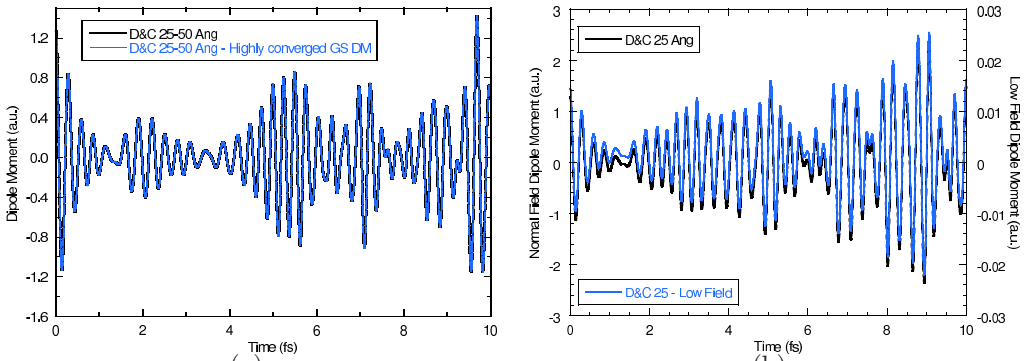


Figure 5.8: Stability of the DCTDDFT method and the CN propagator. a) Highly converged ground state density matrix propagation. b) Propagation with a low strength external field.

the following section different facets of the propagation are investigated.

The first investigation checks the effect of contributions from virtual orbitals on the ground state density during a propagation. A less converged ground state density will have contributions from partially occupied states above the Fermi level at a finite temperature. These contributions might play a significant part in the propagation, leading to a different final state. Shown in Figure 5.8(a), the propagation begins from a highly converged ground state density (blue line), in this case with a density matrix convergence criteria of 1 x 10^{-6} compared to the standard 1 x 10^{-4} (black line). The highly converged density based dipole moment (blue line) is equivalent to the less converged density based dipole moment (black line). Hence, at a convergence tolerance of 1 x 10^{-4}, the density is considered to be converged and does not effect the outcome of the propagation.

The second investigation looks into the effect of the strength of the external electric field. The stability of the DCTDDFT might be strongly dependent on the electric field strength. Shown in Figure 5.8(b), the strength of the field is reduced from 0.1 V/Å (black line) to 0.001 V/Å (blue line) for a subsystem radius of 25 Å. The low-field dipole moments are scaled to match the larger field results. There are no changes in the shape and form of the low-field moment. Hence, the propagation is stable with an electric field strength of 0.1 V/Å.

The next investigation checks the stability of the DCTDDFT when no external field is applied. Shown in Figure 5.9(a), the dipole moment of a conventional

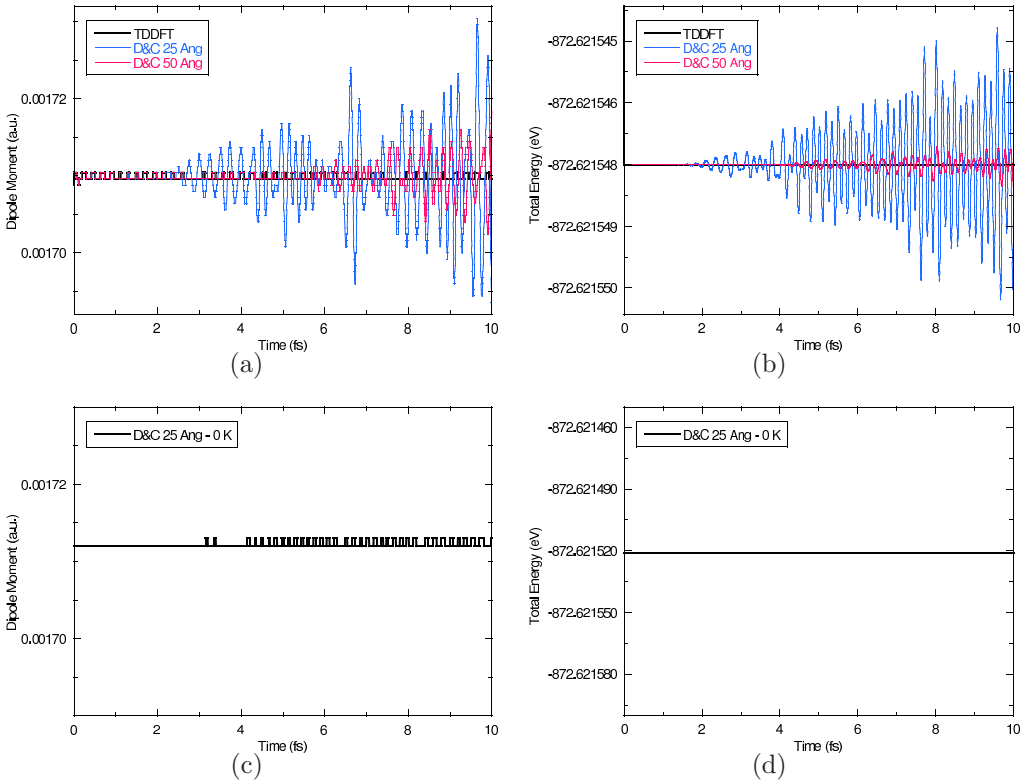


Figure 5.9: Further stability checks of the DCTDDFT method and the CN propagator. The induced dipole moment when no external field is applied is shown in a) with a finite temperature of 100 K, and in c) at zero Kelvin. The total energy when no external field is applied is shown in b) with a finite temperature of 100 K, and in d) at zero Kelvin.

TDDFT calculation (black line) is compared with the dipole moment of the DCT-DDFT method using a subsystem radius of 25 Å (blue line) and 50 Å (pink line), when no external field is applied. The conventional calculation oscillates, with a square waveform, about zero due to the finite electronic temperature. Because the maximum amplitude is small at 0.001713 a.u, the dipole moment amplitude can be effectively regarded as zero. The 25 Å subsystem case follows closely the conventional calculation up until ≈ 2.5 fs, where larger sinusoidal oscillations begin. These oscillations enlarge during the evolution. The oscillations occurs at the point in time the dipole moment diverges for the 25 Å subsystem case when an external field has been applied. The fact the error increases as time increases correlates to the behaviour of the divergences seen in Figures 5.4, 5.5 and 5.6.

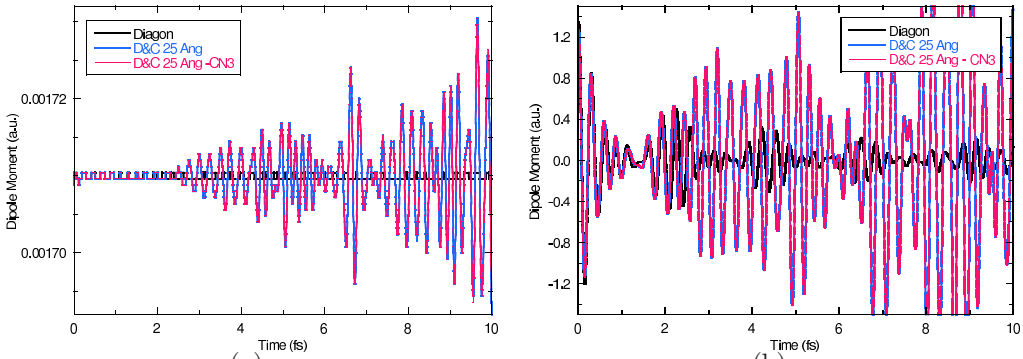


Figure 5.10: Stability of the DCTDDFT method using the third order CN propagator. a) The higher-order propagator is unstable when no external electric field is present. b) The dipole moment diverges with the higher-order propagator.

However, the 50 Å subsystem results are slightly different; there are changes in the square waveform oscillations at ≈ 4.5 fs, corresponding to the time when the dipole moment diverges when an external field has been applied, though the sinusoidal oscillations begin at ≈ 5.5 fs. The total energy during the propagation is also shown in Figure 5.9(b), indicating that the total energy is not conserved.

This is a clear indication that the subsystem eigenfunctions are not orthonormalised and there exists an unphysical instantaneous density. The times at which the total energy begins to oscillate match up precisely to the times the dipole moments diverge for both subsystem sizes.

Running the same calculations at zero Kelvin temperature does not produce the oscillations as did the propagations with a finite electronic temperature, see Figures 5.9(c) and 5.9(d). This clearly indicates that the oscillations were not generated by instabilities in the method but are an indication that the conservation of energy has been violated. This is easily explained when considering subsystem boundary effects and the finite electronic temperature. The temperature introduces momentum into the electrons which when they interact with subsystem boundaries, cause the same divergences as seen when an external field is applied.

Further tests are carried out to completely rule out any intrinsic instability as a probable cause of the dipole moment divergence. An increase in stability can be

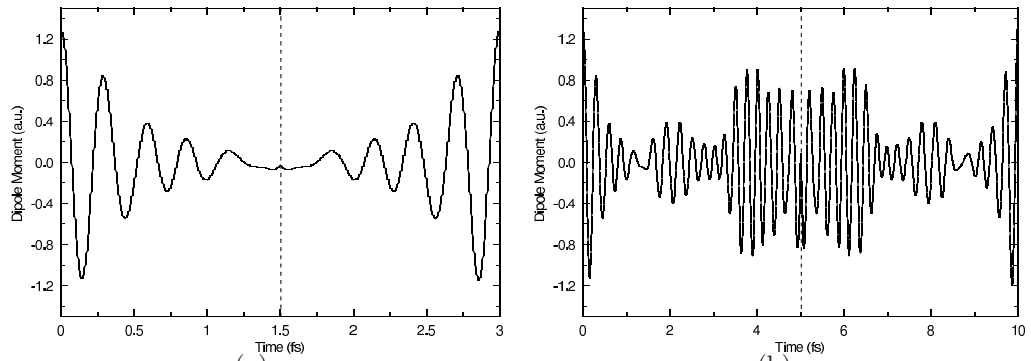


Figure 5.11: The stability of the CN propagator within the DCTDDFT method is tested by propagating forward in time for set a period then backwards in time. The CN propagator is found to be stable even when there is divergence in the dipole moment. a) 1.5 fs propagation time with no divergence. b) 5 fs propagation time including divergence.

achieved with a higher-order expansion of the CN propagator. Shown in Figure 5.10(a), the third-order CN propagator is used with no applied external electric field for a 25 Å subsystem radius calculation. Figure 5.10(b) shows the evolution of the dipole moment for the third-order CN propagator. This higher-order propagator also exhibits divergences during the propagation. Hence, proving that either the first-order CN propagator is stable or that both propagators are unstable. A comparison with another propagator will have to be examined in the future to prove that the DCTDDFT method is stable. A high-order Magnus expansion based propagator [200] could be used for this purpose.

Other issues could be due to the propagation itself, where a predictor-corrector method might be suitable in this situation. A calculation with a very short timestep at 0.0002 fs was performed that produced the same dipole moment as the longer timestep at 0.002 fs, suggesting that a predictor-corrector method might not have a positive effect.

The final check of the stability investigates the time-reversal symmetry property of the CN propagator. Shown in Figure 5.11(a), is the plot of the dipole moment when propagating for 1.5 fs forward in time and then back in time to 0 fs using a 35 Å subsystem. At 1.5 fs there is no divergence and the DCTDDFT method obeys the time-reversal symmetry property. When propagating up to 5

fs, shown in Figure 5.11(b), divergences occur, however, the time-reversal symmetry property is still not broken. Another indicator that the DCTDDFT method and the CN propagator are stable.

All investigations into the stability of the DCTDDFT method and the CN propagator have shown that the method is stable and that no fundamental problems exist with the propagation. In the following section, the subsystem boundary effects are investigated.

5.5.4 Subsystem Boundary Effects

After ruling out any instability issues with the method, the effect of the subsystem boundary on the dipole moment is investigated. It was pointed out by Dr. Daniel Sánchez-Portal [204] that the density could be reflecting off the subsystem boundaries back into the core of the subsystem. There is a high probability that this is occurring, though difficult to prove. Regardless, for the time being it is assumed that reflections off the subsystem boundaries occur.

The reflections are not the only reason for concern. Another possible issue with the subsystem boundary is the indirect influence of the density near the boundary on the density central to the subsystem. For example, the density at the boundary of the subsystem can become polarised, due to a lack of formal mixing of the eigenfunctions with a confining potential, which will influence the density central to the subsystem. This effect may be alleviated by using the D&C method proposed by Zhao *et al* [120], where they altered the standard D&C method by using so-called positive and negative fragments instead of spatial partition functions, which when combined in a specific way can cancel out artificial boundary effects and probably any spurious polarisation moments. In this thesis other approaches have been considered.

If it is assumed the divergence is caused by anomalous boundary effects, the first attempt to alleviate these effects is to use a subsystem outer buffer region. As with the ground state case, the outer buffer region is used to propagate the subsystem eigenfunctions, although, when constructing the instantaneous density matrix, those contributions from the outer buffer region are not included. The purpose of this will be to reduce any boundary effects. Figure 5.12 shows the

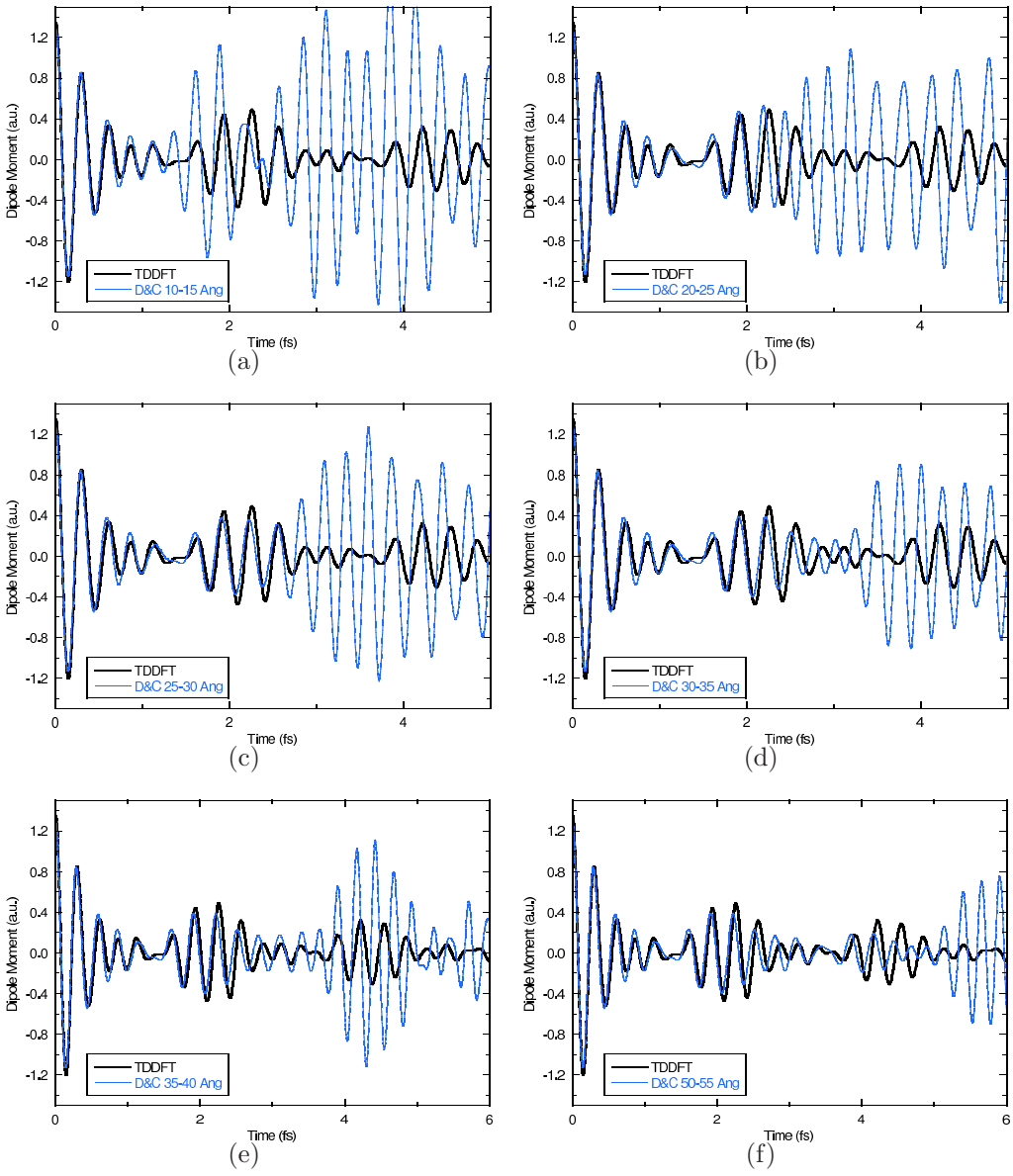


Figure 5.12: The induced dipole moments of a 194-atom linear alkane molecule using the DCTTDFT method with various subsystem sizes including outer buffer regions (blue lines) compared to the standard TDDFT method (black line). The dipole moment of a particular complete subsystem size (including outer buffer region) is equivalent to the dipole moment of a standard subsystem at the same complete size. Any proposed boundary effects on the dipole moment persist.

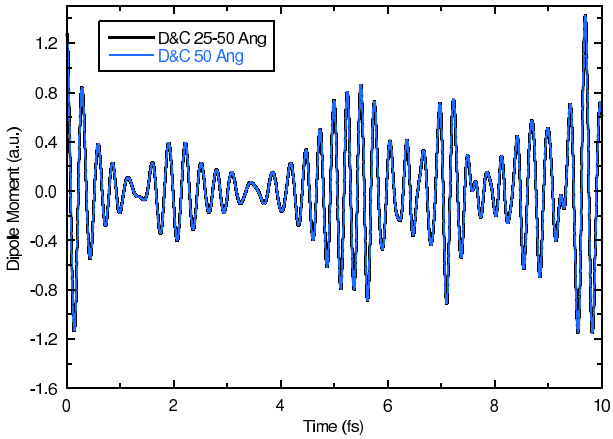


Figure 5.13: Congruence in the dipole moment is demonstrated with a comparison between a 25 Å subsystem with an outer buffer region of 25 Å and a 50 Å subsystem. The dipole moments are shown to be equal.

dipole moments of all previously considered subsystem radii with the addition of a 5 Å outer buffer region. What is found is that the size of the complete subsystem (including the outer buffer region) is equivalent to using a standard subsystem of the same size. For example, a subsystem with a 25 Å radius plus a 5 Å outer buffer region produces the same dipole moment as a 30 Å standard subsystem. It seems that the outer buffer region only prolongs the divergence (i.e. the boundary effects persist).

The congruence of the dipole moment for a subsystem with an outer buffer region and a standard subsystem at the equivalent size is shown in Figure 5.13. The dipole moment is plotted for a subsystem radius of 25 Å with an outer buffer region of 25 Å (black line), in effect producing a 50 Å subsystem. This is plotted against the dipole moment from a standard 50 Å subsystem (blue line). The dipole moments are equal at all times. Unfortunately even with the use of an outer buffer region of 25 Å the proposed boundary effects still exist after a time of ≈ 4.5 fs, that is when the dipole moment of the standard 50 Å subsystem diverges. For the above cases, the subsystem outer buffer region has no effect on alleviating any errors caused by the proposed anomalous effects of the subsystem boundary. The application of a switching function to taper the boundary of the subsystem Hamiltonian, as in section 4.1, is now employed as another possible remedy to the dipole moment divergence. By tapering the interactions, it is possible the

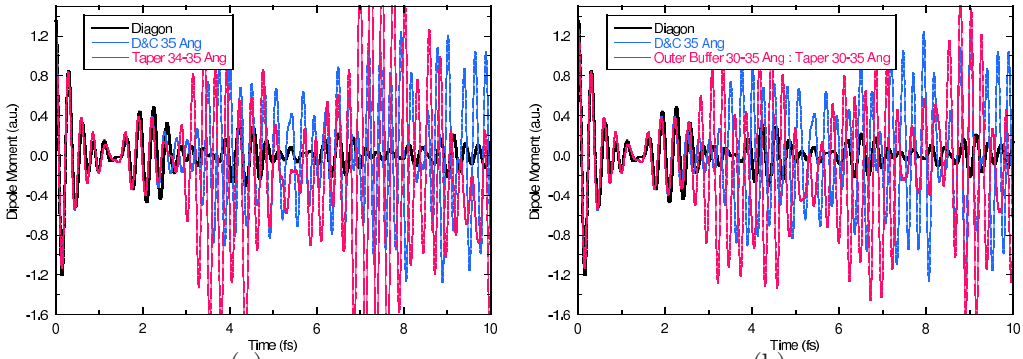


Figure 5.14: An attempt to alleviate the proposed boundary effects on the dipole moment with the use of a switching function to taper the subsystem Hamiltonian interactions at the boundary. The dipole moment divergence persists. a) 2 Å taper region starting from 1 Å before the subsystem boundary. b) 2.5 Å taper region starting from 0.5 Å before the subsystem boundary.

boundary effects will be reduced. However, this is also not the case, with the dipole moments shown in Figure 5.14 for a subsystem radius of 35 Å. Figure 5.14(a) has a 1 Å taper region that goes to zero at the subsystem boundary. The divergence still exists, with an additional problem with the evolution operator not being unitary, seen as shifts in the dipole moment from the oscillations about zero. Because the total energy of the system is fluctuating, the eigenfunctions will not be normalised, resulting in an unphysical density. The amplitude of the oscillations with the tapered run are also larger than the standard D&C calculation. Shown in Figure 5.14(b), a larger 5 Å taper region is used inside a 5 Å outer buffer region. The larger region also has failed to alleviate the divergence. The amplitude of the oscillations are also larger than the standard case.

A similar technique as the tapering mechanism is attempted by applying a masking function [205–207] to each subsystem. The masking function dampens the eigenstate amplitude for only the basis functions in the proximity of the boundary. The masking function is used by real-space TDDFT methods to stop any wavefunction reflections off the boundary of the grid and back into the simulation space. It is anticipated that the masking function will have a similar effect on the subsystem boundary as on the boundary of a real-space grid. The masking function is applied to each subsystem eigenfunction, after each propagation step

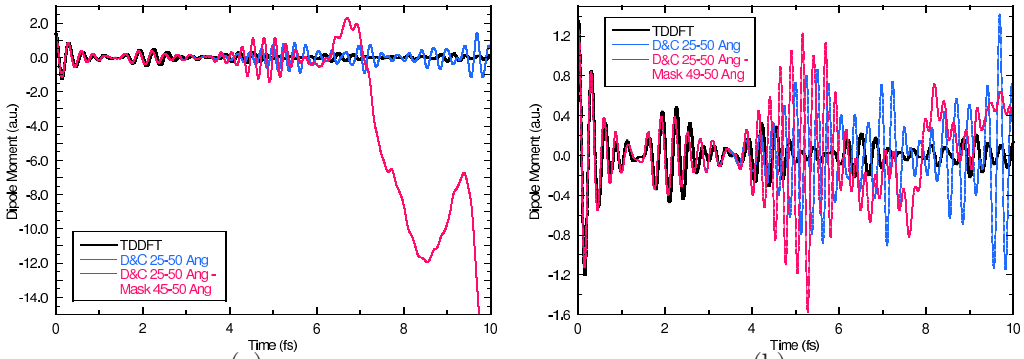


Figure 5.15: An attempt to alleviate the proposed boundary effects on the dipole moment with the use of a masking function on the subsystem eigenfunctions in close proximity to the subsystem boundary. a) 5 Å mask region starting from 5 Å before the subsystem boundary (outer buffer region inclusive). b) 1.0 Å taper region starting from 1.0 Å before the subsystem boundary (outer buffer inclusive).

and before the construction of the density matrix, namely;

$$\phi^{\alpha}_{i,mask} = \phi^{\alpha}_{i} M(r) \tag{5.44}$$

where ϕ^{α}_{i} is the eigenfunction, i, of subsystem α. The masking function used by Burnus *et al* [208] for their time-dependent electron localisation function work is used here, defined as;

$$M(r) = \begin{cases} 0, & r > r_{max} \\ f(r), & r_{ab} < |r| < r_{max} \\ 1, & r < r_{ab} \end{cases} \tag{5.45}$$

where r_{max} is the extent of the masking region and $r_{max} - r_{ab}$ is the width of the masking region. $f(r)$ is a smooth function defined as:

$$f(r) = \left[\cos\left(\frac{\pi(r - r_{ab})}{2(r_{max} - r_{ab})} \right) \right]^{\frac{1}{4}} \tag{5.46}$$

The masking function is applied to a 25 Å subsystem with a 25 Å outer buffer

region using different sized masking regions, as shown in Figure 5.15. Similar to the tapered Hamiltonian case, the modification of the eigenfunction coefficients after they have been propagated, changes the norm of the eigenfunctions and hence the conservation of energy is violated. All the different masking regions exhibit this behaviour, with the dipole moment either increasing or decreasing to large and unrealistic values. The first masking function uses a masking region size of 5 Å located at a distance of 45 Å from the centre of the subsystem, shown in Figure 5.15(a). This masking region produces large deviations of the dipole moment, indicative of an unrealistic density. Figure 5.15(b) uses a masking region of 1 Å. Even with only a small amount of subsystem eigenfunctions actually affected by the masking function, the eigenfunctions will eventually not be normalised.

As an alternative to a masking function, a complex absorbing potential [209–212] can be added to each subsystem Hamiltonian to absorb the eigenfunctions at the subsystem boundary. The potential can be turned on very slowly which can possibly remedy the above mention eigenfunction normalisation issue. The complex absorbing potential can also be made to cater for a small range of oscillatory wavelengths where as the masking function is applied to all wavelengths. Coupled with using a monochromatic laser pulse as the excitation source the absorbing potential could help reduce the eigenfunction normalisation issue. This potential has not been implemented and is a possible avenue to explore in the future.

The final attempt to solve the dipole moment divergence problem is shown in Figure 5.16. In this case, orbitals within a certain distance from the subsystem boundary are kept frozen during the propagation for a 35 Å subsystem. That is, those particular orbitals were not allowed to evolve and remained with constant coefficient values throughout the propagation. Only orbitals close to the core of the subsystem are allowed to evolve. Figure 5.16(a) has a frozen region of 5 Å and Figure 5.16(a) has a frozen region of 10 Å. In both cases, the frozen orbital regions produced divergences at an earlier time than the standard DCTDDFT method. The frozen regions acted as new boundaries, in effect reducing the size of the subsystem. The amplitude of the divergences are also much larger than

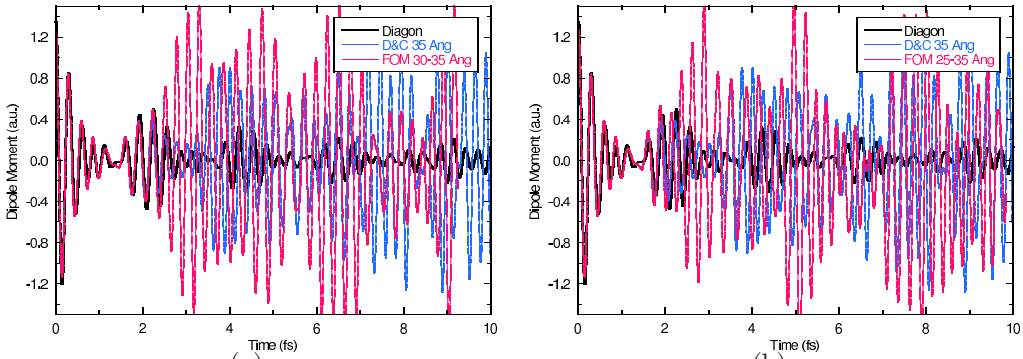


Figure 5.16: The dipole moments generated by not propagating orbitals within a certain distance from the 35 Å subsystem boundary. Divergences in the dipole moment still occur. a). 5 Å frozen orbital region. b) 10 Å frozen orbital region.

those found using the standard DCTDDFT method.

All attempts at alleviating the boundary effects that were proposed to be the cause of the dipole moment divergence have failed. This suggests that either the method parameters need to be optimised or that the anomalous subsystem boundary effects will always occur and that other means must be used to avoid these effects. Other approaches to the DCTDDFT method need to be investigated e.g. different partition functions. Constraints on the time-dependence also needs investigation.

5.5.5 Computational Scaling

The computational scaling of the real-time DCTDDFT method for the 194-atom linear alkane molecule (using the same settings as for all previous calculations) is shown in Figure 5.17. The required time for a single propagation of the eigenfunctions is shown for the conventional TDDFT case (black line + circles) and for the DCTDDFT cases with a 10 Å subsystem (blue line + squares) and a 25 Å subsystem (pink line + diamonds). The standard TDDFT method exhibits O (N^3) scaling while the DCTDDFT method is shown to be linear-scaling. The prefactor for the 25 Å subsystem is larger than the 10 Å subsystem. The cross-over point for the 25 Å subsystem occurs at about 950 atoms. For this 1-dimensional case

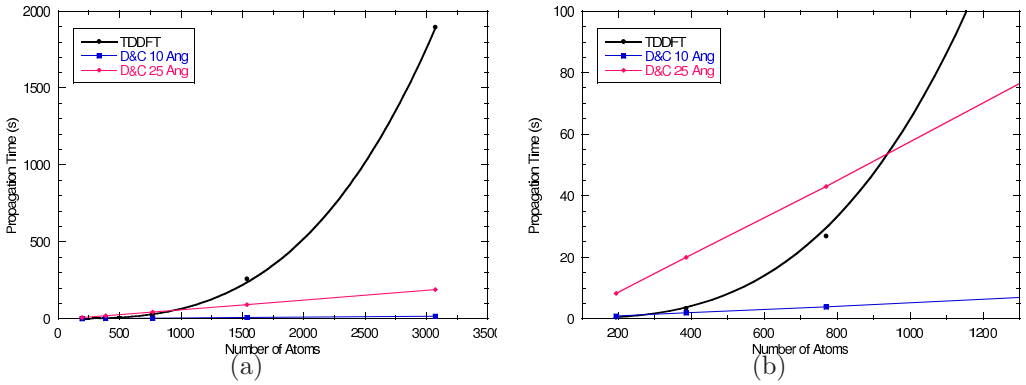


Figure 5.17: The time required for a single propagation of the eigenfunctions of a 194-atom linear alkane molecule using standard TDDFT and DCTDDFT. The standard calculation scales as $\mathrm{O}\left(N^3\right)$ while the DCTDDFT method is shown to scale linearly. The cross-over point for the 25 Å subsystem occurs at about 950 atoms using SZ basis set. a) shows the complete plot. b) close up view to highlight the cross-over point.

the prefactor is not so great as to be the cause of large computation times. In this case, there is a clear advantage of using DCTDDFT rather than standard TDDFT when comparing calculation times only. For 3-dimensional dense systems the prefactor might be a factor when considering the time required for the calculation, causing the cross-over point to be pushed out to much larger atom numbers.

5.5.6 Optical Response

The optical response of the method is now examined. It is accepted that the dipole moment diverges, so what is done here is to calculate the imaginary polarisability for each subsystem size by only considering the dipole moment up to the point of the divergence, shown in Figures 5.18 and 5.19. Figure 5.18(a) shows the limitations in the resolution of the frequency response as the propagation time is decreased from 10 fs to 5 fs for the conventional TDDFT case. Although all the peaks are still present, they are less defined in the 5 fs case. For the DCTDDFT cases, because the longest valid propagation time occurred for the 50 Å subsystem case at ≈ 4.5 fs, the optical response will be compared to the 5 fs optical response of the conventional TDDFT case.

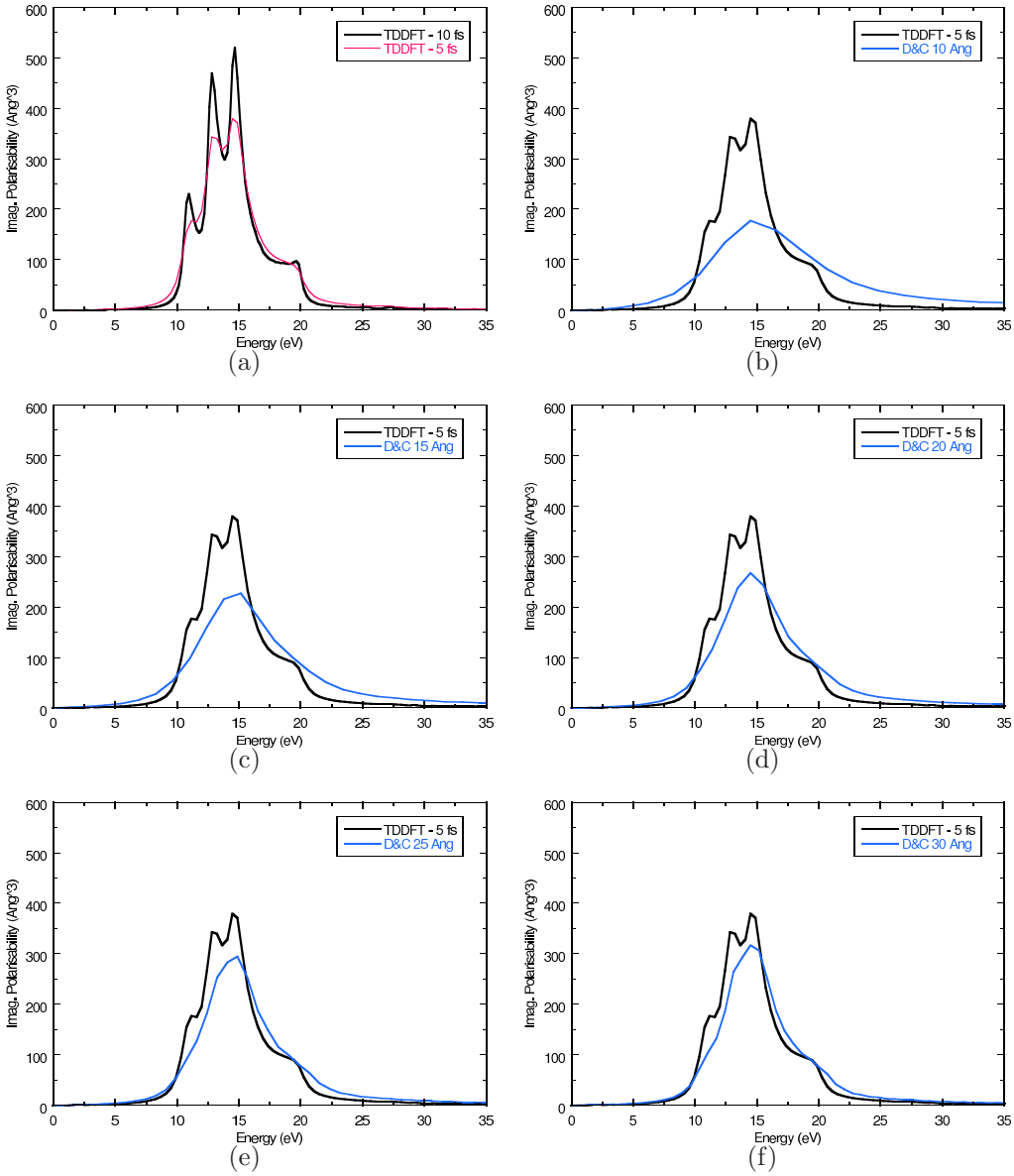


Figure 5.18: The optical response for subsystem sizes from 10 Å up to 30 Å. Only the dipole moment up to the point of divergence is used in the calculation.

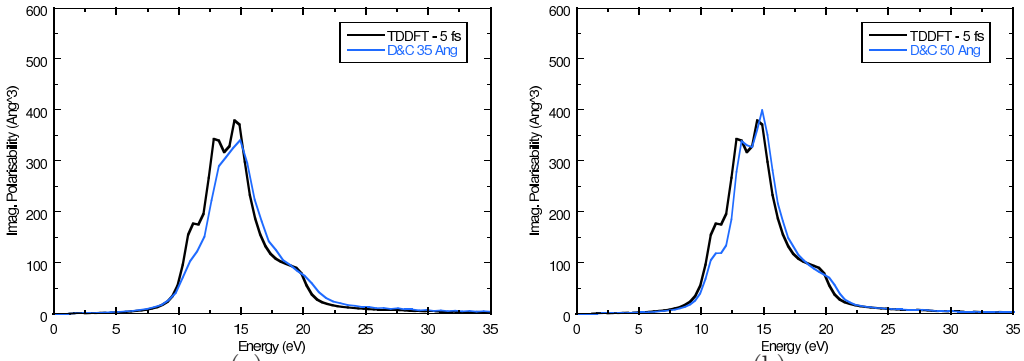


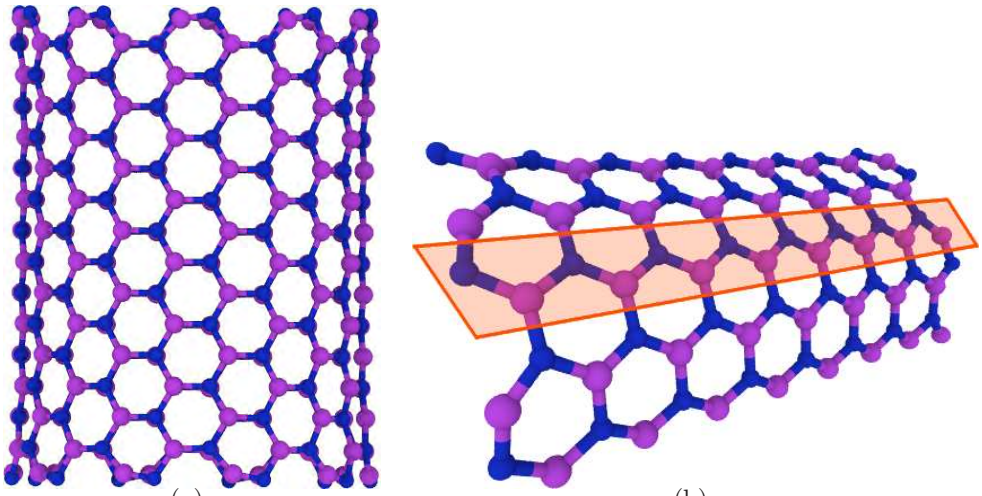
Figure 5.19: The optical response for subsystem sizes of 35 Å and 50 Å.

For small subsystem radii, the resolution of the optical response is low. There is a single peak at $\approx 15\,\text{eV}$ with a long extended tail towards higher energy. The long tail eventually forms into a shoulder peak at $\approx 22\,\text{eV}$ for subsystem radii greater than 25 Å. The 50 Å subsystem has the longest propagation time and hence the highest resolution, producing all four peaks of the conventional TDDFT case. The peaks in this case are at slightly shifted frequencies and the intensities are not exact, though this is still a comparable result for the approximations being made.

It is worth emphasising that the length of the propagation determines the resolution of the optical response. As was found earlier, convergence in the dipole moment is achieved with a 25 Å subsystem, which means that if the 25 Å subsystem calculation were to propagate successfully up to 5 fs then it would produce comparable results as the 50 Å subsystem, with a large reduction in the prefactor.

5.6 2-dimensional Partitioning

The divergence in the dipole moment due to subsystem boundary effects is attempted to be remedied by using a 2-dimensional partitioning scheme. A 2-dimensional partitioning scheme is a scheme with subsystems that are infinite in size (or periodic) in one dimension, while the other dimensions are finite in size, as in the normal scheme. In essence, the 2-dimensional partition can be thought of as cylindrical in shape. This type of partitioning will not be strictly linear-



(a) (b)

Figure 5.20: Ball and stick representation of the complete BNNT structure and a single multiple core subsystem. a) Unit cell of 19.5Å (10,10) BNNT structure used in the DCTDDFT 2-dimensional partition propagations. b) a subsystem with multiple core atoms along the length-wise axis of the nanotube. The core atoms are within the shaded area. The subsystem has no boundaries along this axis, as it encapsulates the complete length of the nanotube.

scaling, although it will in most cases reduce the scaling significantly. Cases for which linear-scaling occurs will be systems that grow in directions orthogonal to the direction of the infinite dimension. This type of partitioning is thought to alleviate any anomalous subsystem boundary effects with electron charge density flowing parallel with the dimension, as there is no boundary. Any electron flow crossing the subsystem boundaries of the other dimensions will still be subject to subsystem boundary effects. Typically, the direction of the infinite dimension will be aligned parallel to the external electric field.

Testing of the 2-dimensional partitioning scheme is carried out on a $\approx 19.5\mathring{A}$ (principal axis) boron-nitride nanotube (BNNT) with (10,10) chirality, see Figure 5.21(a). This insulating system has a structure that is well suited to the cylindrical shape of the subsystems used in the 2-dimensional partitioning scheme. For the BNNT, the infinite dimension of the subsystem is directed along the length-wise axis of the molecule. Instead of having a single core atom, in this case, there are multiple atoms in the core region, see Figure 5.20(b). Having multiple core

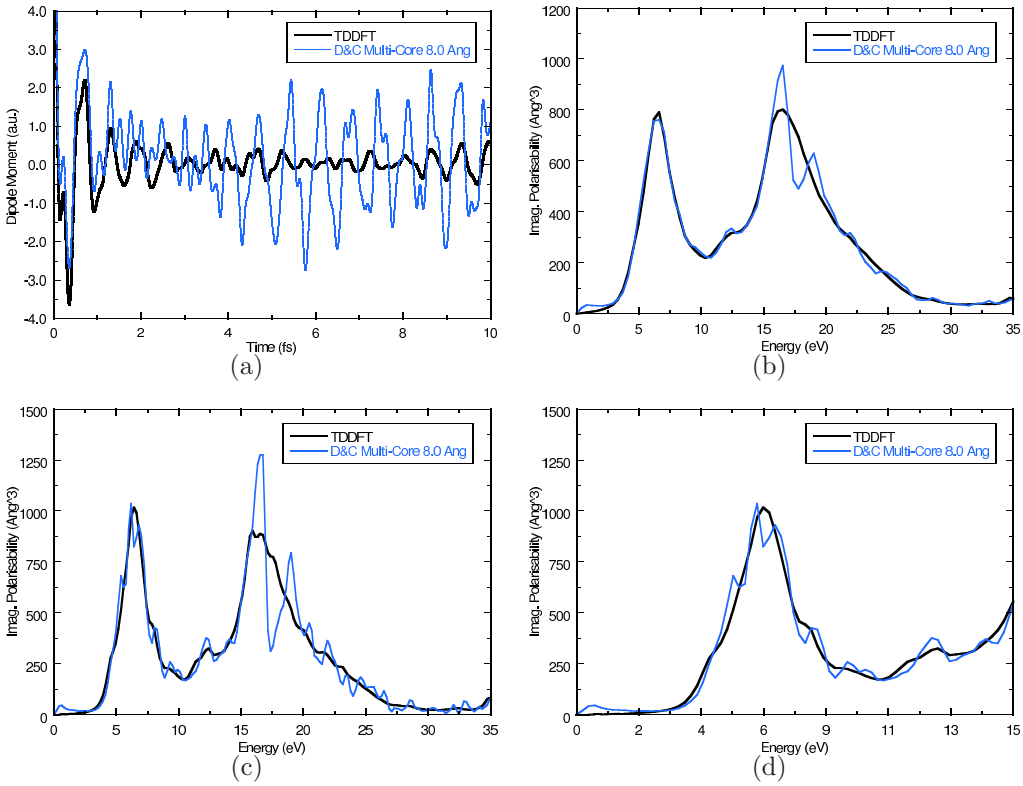


Figure 5.21: The dipole moment and optical response of the ≈ 19.5Å (principal axis) (10,10) BNNT structure using standard TDDFT and the D&C method employing a 2-dimensional partitioning scheme with multi-core atom subsystems. a) induced dipole moment up to 10 fs. b) imaginary polarisability using 5 fs of induced dipole moment data. c) imaginary polarisability using 10 fs of induced dipole moment data. d) close up view of c).

atoms reduces overlap amongst neighbouring subsystems, which in turn, speeds up the calculation.

The calculations were carried out using a 100 Ry cut-off for the real-space integration grid used to represent the density, an energy shift of 0.02 Ry for the PAO orbital confinement and a density matrix convergence criteria of 1 x 10^{-4} for self-consistency. The PBE [54] form of the GGA was used for the XC functional in the adiabatic approximation. A SZ basis set is used in all sets of calculations. Norm-conserving Troullier-Martins pseudopotentials [88, 91] in the Kleinman-Bylander factorised form [89, 90] were used. The eigenstates were propagated for 10 fs with a time step of 0.002 fs and an initial external electric field perturbation of 0.1 V/Å aligned along the length-wise axis of the nanotube.

Figure 5.21(a) shows the induced dipole moment for the BNNT system for both standard TDDFT (black line) and the 2-D partition DCTDDFT method (blue line) with a buffer radius of 8.0 Å. Unfortunately, the 2-dimensional partitioning does not stop the dipole moment from diverging. The divergence is not as pronounced as found in the previous sections with the standard partitioning for the polymer system. The major peaks of the DCTDDFT method remain in phase at the beginning and end of the propagation with that found with standard TDDFT propagation. There is also noise in the DCTDDFT case, which will add peaks to the optical response.

The optical response is shown in Figure 5.21(b), for a propagation of 5 fs, and Figure 5.21(c), for a propagation of 10 fs. A close up view of 5.21(c) is shown in Figure 5.21(d). For the 5 fs propagation, the DCTDDFT imaginary polarisability is equal to the standard TDDFT imaginary polarisability up to $\approx$ 16 eV, after which the major peak at $\approx$ 17 eV is found with DCTDDFT. The small peak at $\approx$ 12.5 eV is even found with the DCTDDFT method. At the higher energies, the DCTDDFT introduces large noise spikes which do not exist for the standard TDDFT propagation. These are due to the extra features found in the dipole moment. The DCTDDFT imaginary polarisability is relatively smooth, due to the short propagation time. With a 10 fs propagation, the DCTDDFT imaginary polarisability still has all the major peaks at the correct energies, although there are many smaller peaks which do not exist in the standard propagation. The majority of the extra peaks match up with the shoulder peaks found with the standard propagation.

The 2-dimensional partitioning still produced divergence in the dipole moment, although, the divergence was not as great as found with standard D&C partitioning. The DCTDDFT dipole moment remained relatively in phase with the standard TDDFT dipole moment. This fact, along with the success of the DCTDDFT with the alkane molecule with external fields aligned normal to the molecule (which in this case, is equivalent to the 2-dimensional partitioning scheme), suggests that further improvements to the 2-dimensional partitioning scheme are worthwhile.

5.7 Concluding Remarks

A real-time TDDFT method using the D&C paradigm has been proposed in this chapter. The method was tested on a linear alkane molecule and produced accurate dipole moments for the situation when the electric field was aligned normal to the principal axis of the molecule. However, divergence in the dipole moment occurred for the case where the electric field was aligned parallel to the principal axis of the alkane molecule. The point in time of the divergence is related to the subsystem size, where larger subsystems diverge at longer times. The effects of the subsystem boundary on the divergence were investigated. Attempts to eliminate subsystem boundary effects using outer buffer regions, Hamiltonian tapering, masking functions and a 2-dimensional partitioning scheme were not successful. The stability of the method was demonstrated by not applying an external field and showing that the dipole moment and total energy do not diverge. The stability of the method was further reinforced by showing that the DCT-DDFT CN propagator maintained the time-reversal symmetry property of stable propagators. The optical response was calculated and produced fairly reasonable results when compared to standard the TDDFT method, despite the issues with the dipole moment. The actual cause of the dipole moment divergence is found to be most probably related to subsystem boundary effects.

Further investigations are required into improving the method. One improvement is the determination and handling of any time-dependent constraints within the method, as it is unclear if the partition function and the subsystem propagation should explicitly handle any time constraints. Application of a complex absorbing potential [209–212] to each subsystem needs to be investigated. Finally, instead of exciting all frequencies at once with the external electric field, using a monochromatic laser field to excite a single frequency might be more effective when applying any of the previous measures (and proposed measures) to counter the subsystem boundary effects. For example, with a monochromatic laser field perturbation, the wavefunction masking function and complex absorbing potential can be optimised to operate more effectively at the single excitation frequency.

6

Divide and Conquer Applications

The performance and capabilities of the current D&C method implementation are investigated in this chapter. Two different deoxyribonucleic acid (DNA) are first considered. The effects and performance of the tapering mechanism on λ-DNA [1] is investigated. Then a comparison is made of the energy convergence of a 1WQZ DNA structure [35] with the DMM method. Furthermore, the electronic structure of 1WQZ is examined. Finally, the electronic structure of a zeolitic imidazolate framework (ZIF) crystal, ZIF-100 [40], is found.

6.1 Deoxyribonucleic acid

The applicability of the D&C method for two different strands of DNA is investigated in this section. The first is a periodic dry λ-DNA system, previously studied by de Pablo *et al* [1]. The convergence of the total energy with respect to increasing subsystem sizes is investigated when the tapering mechanism is applied. Acceleration of the SCF convergence rate is also examined. The second DNA strand is the 1WQZ (Protein Data Bank Id) structure. For this system, convergence in the Harris functional energy [213, 214] found by the D&C method is compared to results published by Otsuka *et al* [35], who have used the DMM method [36] within the CONQUEST code [37–39]. The computational resources required to run the convergence are investigated, along with a comparison of the electronic structure of hydrated and dehydrated 1WQZ DNA. In particular, a comparison of the partial atomic charges are is made. This information will be

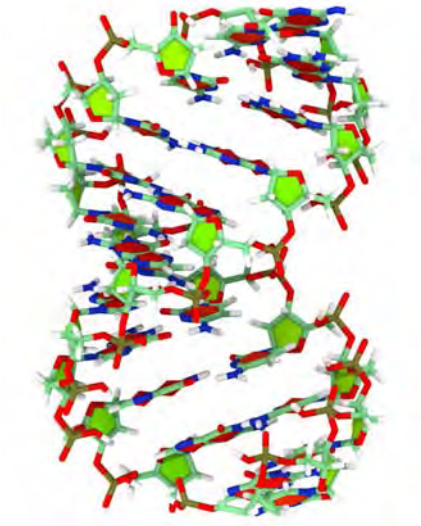


Figure 6.1: Schematic stick representation of the 715-atom dry λ-DNA unit cell. The cell repeats along the vertical axis. Atoms are coloured according to the following scheme; Hydrogen - white, oxygen - red, carbon - green, nitrogen - blue and phosphorus - brown.

particular useful in the setup of the electrostatic interactions within force fields used in MD simulations.

6.1.1 λ-DNA

The dehydrated λ-DNA system was previously examined in a study by de Pablo *et al* [1]. They used this structure to demonstrate computationally an absence of conductivity in dry λ-DNA (experimental verification was also given in the same work). This study was the first to apply a first principles linear-scaling technique to the study of DNA. The linear-scaling method de Pablo *et al* used was the KMG functional minimisation method [32] implemented in SIESTA [2]. Although this DNA strand only has 715 atoms, which is now well within the domain of standard diagonalisation techniques with the use of appropriate computing resources, it is a DNA system known to converge. Hence it should represent a robust test case to examine in this thesis. The unit cell of the DNA strand repeats along the axis of the strand, as shown in Figure 6.1.

It has been shown that discontinuities can occur in the energy surface when subsystem atom memberships are altered (section 4.1). A method to attempt to

alleviate the sudden changes in energy by applying a switching function to taper interactions in the Hamiltonian was proposed. The performance of the tapering mechanism is investigated on the larger λ-DNA system by testing the convergence of the total energy with respect to the subsystem size. The SCF convergence rate when the tapering mechanism is applied is also investigated.

Calculations were performed in parallel using 24 processors of a 11,900 processor SUN Constellation computer (2.93GHz Intel Nehalem cpus). All calculations were performed using the memory conservation scheme, as described in section 3.5.2, which on average will increase the computational time by 50 %. The calculations were carried out using a 150 Rydberg cut-off for the real-space integration grid used to represent the density, a DZP basis set on all atoms with an energy shift of 100 meV for the PAO orbital confinement, and a density matrix convergence criteria of $1x10^{-4}$ for self-consistency. The PBE [54] form of the GGA was used for the XC functional. Norm-conserving Troullier-Martins pseudopotentials [88,91] in the Kleinman-Bylander factorised form [89,90] were used. The study done by de Pablo *et al* [1] used a DZ basis set on all atoms except phosphorus and for atoms involved in hydrogen bridges, where a DZP basis set is used. For the DZ basis set atoms the orbital radii were manually set as follows; H states at 4.2 Bohr; C states at 4.1 Bohr; N states at 3.6 Bohr; and O states at 3.2 Bohr. Within the DZP basis set used by de Pablo *et al* the non-polarisation functions are larger and were set as follows; H states at 5.5 Bohr; N states at 4.6 Bohr; and O states at 4.2 Bohr.

6.1.1.1 Results

Six sets of calculations are performed for each subsystem size on the optimised λ-DNA geometry from de Pablo [1], shown in Figure 6.2. The first set is a standard D&C calculation, while the rest of the calculations use the tapering mechanism with varying taper regions. The taper regions range from 1.0 Å up to 2.0 Å in 0.25 Å increments. The energy curve for the standard D&C calculation between a subsystem radius of 8.0 Å and 11.0 Å is smooth. The discontinuity in the energy occurs at a subsystem radius 7.0 Å. The standard D&C energy converges to the energy found via standard diagonalisation (dashed line), at −121821.7687 eV,

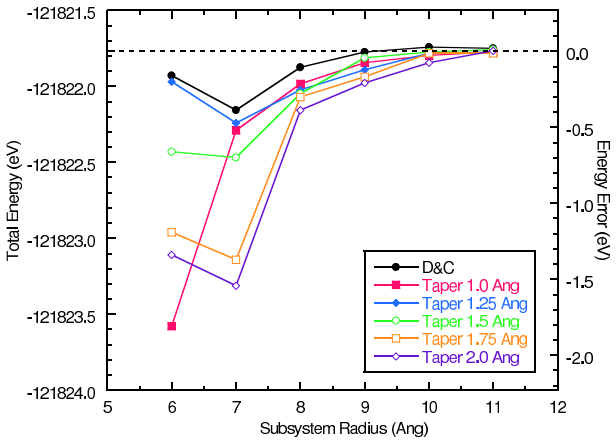


Figure 6.2: A comparison of standard D&C and D&C with the application of the tapering mechanism for the total energy convergence with respect to increasing subsystem size for the λ-DNA system. The dashed line at $-121821.7687\,\text{eV}$ is the total energy found using standard diagonalisation techniques.

quite rapidly. Convergence can be considered to have occurred with a subsystem radius of 9.0 Å with a error in the energy of $-6.4\,\text{meV}$. A 7.0 Å subsystem has an error of $-388.5\,\text{meV}$ which amounts to only $-0.54\,\text{meV/Atom}$.

For all taper regions, the taper mechanism tends to reduce the total energy relative to the standard D&C calculation. For subsystems greater than and including 7.0 Å this implies the energy will converge at a slower rate. The trend indicates that larger taper regions produce lower energies and hence will converge at a slower rate. The 1.0 Å taper region is the only taper region to produce a systematically increasing curve. All other taper regions only slightly lowered the energy. However, it can be argued that the energy for the 1.0 Å taper region at a subsystem radius of 6.0 Å is too low and is in fact still a discontinuity. This does *not* signify that the taper mechanism has not functioned as intended. It more than likely has to do with the type of simulation performed here. The D&C SCF procedure does not necessarily converge to the correct energy with increasing subsystem size in a systematic fashion and will typically converge in an oscillatory fashion.

Table 6.1 shows a comparison of the absolute errors per atom between the KMG functional [32] work of de Pablo *et al* [1] with the standard D&C work of this thesis. de Pablo *et al* claim to use a LWF localisation region between

Table 6.1: Comparison of the absolute error per atom (meV) between the de Pablo *et al* work [1] using the KMG functional minimisation method and the standard D&C method for the λ-DNA system.

KMG	D&C						
$4-5$ Å	5 Å	6 Å	7 Å	8 Å	9 Å	10 Å	11 Å
5.0	1.87	0.23	0.54	0.15	0.01	0.04	0.03

$4.0-5.0$ Å. The exact localisation region size is unclear. An extra D&C method calculation at a subsystem radii of 5.0 Å was carried out to ensure a fairer comparison. A 4.0 Å D&C calculation was also carried out, although failed to converge. All D&C subsystem sizes were found to produce smaller errors than the KMG method. The closest comparison that can be made is with the 5.0 Å subsystem with an error 1.87 meV/Atom being much smaller than the KMG error of 5.0 meV/Atom

The SCF iteration counts for the above calculations are shown in Figure 6.3. The Broyden mixing scheme was employed with a weight of 0.05 for all sets of calculations. A low mixing weight is required due to convergence issues when using larger mixing weights with small subsystem sizes. To determine if the tapering mechanism helps accelerates SCF convergence all things must be equal i.e. the mixing weight could not be modified during the self-consistency process to help with convergence.

The assistance of the tapering mechanism on the self-consistency rate is evident for the λ-DNA case. The largest difference in the SCF iterations occurs with a subsystem radius of 6.0 Å, where the standard D&C calculation took 210 iterations to converge, while only 48, 51, 35, 32 and 28 iterations were required for the taper regions from 1.0 Å up to 2.0 Å. For this subsystem size, as the taper regions grew in size, the number of SCF iterations reduced. For a 7.0 Å subsystem, only taper regions 1.25 Å to 1.75 Å had lower SCF iteration counts than the 40 iterations required for the standard D&C case. The taper mechanism produced lower counts for the 8.0 Å subsystem for all taper regions. Subsystems at 9.0 Å and larger converged at similar rates regardless if the tapering mechanism was applied. With such large subsystems, the subsystem will capture most of the

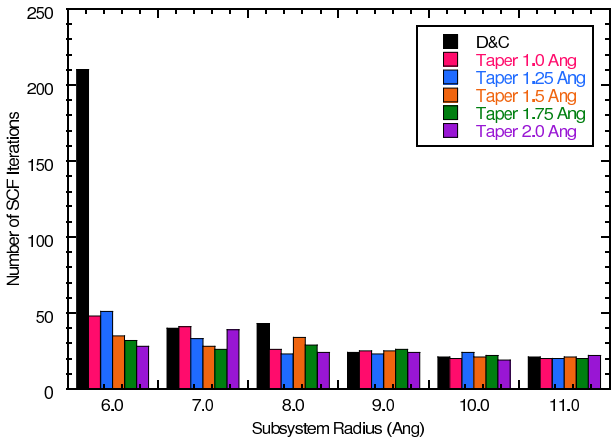


Figure 6.3: The SCF iteration counts for the total energy convergence calculations on the λ-DNA system. The tapering mechanism is found to reduce the number of iterations for small subsystem sizes.

information of interacting orbitals and should converge at a similar and constant rate to standard diagonalisation, which in this case requires 20 iterations.

The de Pablo *et al* [1] relaxation calculation ran with an average of 7 SCF iterations per relaxation step using the KMG functional minimisation method, which seems much lower than the SCF counts reported in thesis using the D&C method. This does not necessarily highlight any deficiencies with the D&C method, as the optimised geometry is the one found using the KMG method and a low mixing weight is used in this thesis. In addition, the average number of SCF iterations can not really be compared to the SCF counts reported in this thesis as the SCF procedure during a geometry relaxation starts using a density matrix from a previous relaxation step (i.e. closer to convergence), while the SCF procedure from a single-point calculation starts from isolated atomic densities (i.e. far from convergence).

The taper mechanism has been found to slightly lower the total energy in the convergence tests of the λ-DNA system, with only the 1.0 Å taper region producing a systematically converged curve over the whole range. The biggest benefit of the tapering mechanism is the acceleration of the SCF process for small subsystem sizes. The tapering mechanism can be of great benefit in accelerating the convergence process, which for large systems is greatly beneficial.

6.1.2 1WQZ DNA Structure

Recently, the CONQUEST code [37–39] has introduced the option to use the SIESTA PAOs [2, 5, 84, 215] for its basis set functions, offering a unique opportunity to compare two different linear-scaling methods which use the same basis set functions and are also implemented in different codes. This makes possible a comparison between the DMM implementation [36] of CONQUEST with the current D&C implementation in SIESTA. Using the 1WQZ DNA structure from the work of Otsuka *et al* [35][a], the convergence of the CONQUEST Harris energy functional [39] using a SZ basis set is compared with the convergence of the SIESTA Harris energy functional [2, 213, 214], as the subsystem size is increased. In addition, the average calculation time of the SCF iterations for each of the convergence runs is reported. The electronic structure of hydrated and dehydrated 1WQZ DNA is investigated using a high quality calculation with a DZP basis set, where partial atomic charges are calculated and compared. The work presented in this section extends the previous study by Otsuka *et al* [35].

The atomic structure of the 1WQZ DNA system is shown in Figure 6.4, consisting of a total of 3,439 atoms. Specifically, it contains 634 DNA atoms where the B-DNA decamer 5'-d(CCATTAATGG)$_2$-3' is used. The DNA molecule is hydrated with 932 water molecules and 9 Mg counter-ions are included for charge neutrality. The original 2.9 Å-resolution x-ray diffraction data set included a few deuterium atoms. These have been replaced with normal hydrogen atoms, as there is only interest in the electronic structure. The unit cell dimensions are 39.74 Å by 31.03 Å by 27.09 Å. This is a reasonably large system, which requires the use of a linear-scaling method.

Currently the non-self-consistent Harris energy functional implemented within SIESTA [2] only functions within the LDA using the Perdew-Zunger parameterisation [52]. Because GGA potentials are used here, the Harris energy is taken from the zeroth SCF iteration before any density matrix mixing can occur i.e. self-consistent Harris energy functional. This is possible because the first ap-

[a]The structure and calculation details were graciously made available by Dr. Tsuyoshi Miyazaki and Dr. Takao Otsuka, who have optimised the structure with the Amber9 package [216].

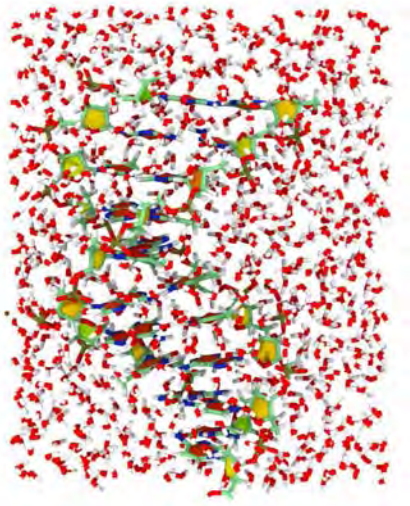


Figure 6.4: Schematic stick representation of the 3,439-atom wet 1WQZ (Protein Data Bank Id) DNA unit cell. The unit cell dimensions are 39.74 Å by 31.03 Å by 27.09 Å. Hydrogen - white, oxygen - red, carbon - green, nitrogen - blue and phosphorus - brown.

proximation to the density just consists of the atomic densities, precisely what is required for the Harris energy functional.

Calculations were performed in parallel using either 32 processors (for the hydrated 1WQZ structure) or 24 processors (for the dry 1WQZ structure) of a 11900-processor SUN Constellation (2.93GHz Intel Nehalem cpus), using 3 GB or less of RAM per CPU for all calculations. Calculations were performed using the memory conservation scheme, as described in section 3.5.2, which on average will increase the computational time by 50 %. The calculations were carried out using a 400 Rydberg cut-off for the real-space integration grid used to represent the density, a density matrix convergence criteria of $1\text{x}10^{-4}$ for self-consistency, an energy shift of 100 meV for the PAO orbital confinement for all basis sets. The Harris functional energy calculations use both SZ and DZP basis sets for the different simulations, while the high-quality electronic structure calculations use a DZP basis set on all atoms. The PBE [54] form of the GGA was used for the XC functional. Norm-conserving Troullier-Martins pseudopotentials [88, 91] in the Kleinman-Bylander factorised form [89, 90] were used. The SZ basis set and all other settings are consistent with the settings used by Otsuka *et al* [35].

6.1.2.1 Results

The convergence of the Harris functional energies with respect to the subsystem size for both the hydrated and dehydrated 1WQZ DNA systems are shown in Figure 6.5. The calculations are run using a SZ basis set and a DZP basis set. The error in the energy for all plots is taken to be relative to the energy found using standard diagonalisation (as indicated by the dashed lines). Standard diagonalisation calculations could not be run on the hydrated system using a DZP due to the larger number of basis functions used for the system. The energy for the hydrated SZ 1WQZ system (Figure 6.5(a)) converges systematically and resembles the energy curve reported by Otsuka *et al* [35]. The energy converges quickly, with an error of 0.015396 eV for the 8.0 Å subsystem down to an error of 0.000052 eV for the 12.0 Å subsystem. Even a 6.0 Å subsystem can produce accurate results with an error of 0.390052 eV, which amounts to only 0.000113 eV/Atom. The energy for the hydrated DZP system (Figure 6.5(b)) is only calculated up to a subsystem radius of 8.0 Å. For larger subsystems the memory requirements per compute node are found to be greater than the available 3 GB per compute node. Using high quality settings with large basis sets requires large amounts of memory for each compute node. A new memory model needs to be implemented to handle these situations. In short, the subsystem data needs to be stored on multiple compute nodes as this will allow for larger subsystems (see section 3.5.2).

The energy convergence for the dehydrated SZ system (Figure 6.5(c)) converges quickly and systematically for subsystem sizes 7.0 Å and larger. It is unclear if the low energy found at a subsystem radius of 6.0 Å is a discontinuity or if it is due to oscillatory behaviour. For well converged energies, the errors range from 0.015835eV for 8.0 Å subsystem calculations down to 0.000273 eV for 12.0 Å subsystem calculations. These errors are comparable to the errors found with the hydrated SZ system. The energy for the dehydrated DZP system (Figure 6.5(d)) converges in a systematic fashion, although the energy converges from lower energies. The convergence is found to be slower than the SZ results reported here, especially for subsystems 7.0 Å and below. For subsystems above 7.0 Å the convergence rate is faster, although still slower than the SZ dehydrated

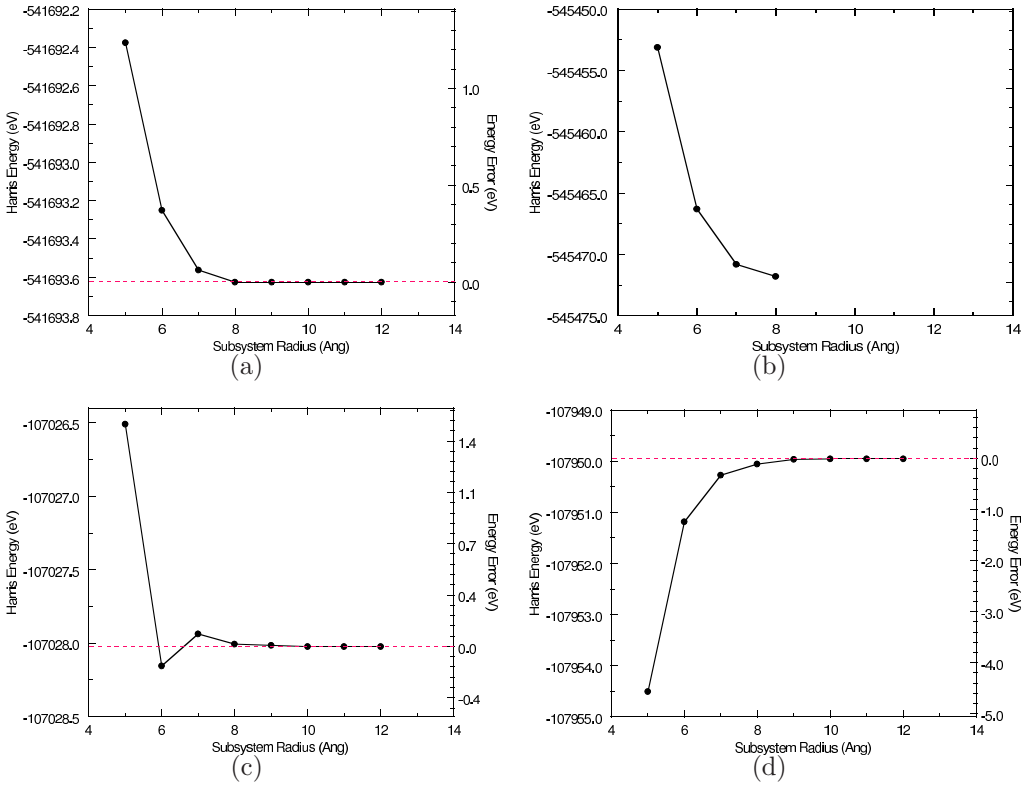


Figure 6.5: Convergence in the Harris functional energy of the 1WQZ DNA structure as the subsystem size is increased. The energy is considered to be converged at a subsystem radius of 8 Å. The dashed lines indicates the energy found using standard diagonalisation. A small subsystem radius can be used for qualitative data due to the small energy error. a) Hydrated system (3439 atoms) with SZ basis set b) and DZP basis set. The Dehydrated system (634 atoms) with a SZ basis set d) and a DZP basis set.

system. The error for the 8.0 Å subsystem is found to be −0.106747 eV which is roughly an order of magnitude greater than the SZ dehydrated system result. The error for a 12.0 Å subsystem is −0.000398 eV, which is well within typical energy convergence criteria used in the SCF process.

Comparisons with the energy curves found by Otsuka *et al* [35] can be made firstly by examining the energy range in which the curves reside. Only SZ results can be used as Otsuka *et al* have only reported results with this basis set. For the Otsuka *et al* hydrated curve, the energy range between a cut-off distance[b]

[b]The spatial cut-off distance in DMM refers to the threshold distance that sets elements in the density matrix, as proposed by Li, Nunes and Vanderbilt [105], to zero for distances

from 7 Å to 12 Å for the hydrated system is $-541,533.6$ eV to $-541,534.1$ eV. For equivalent subsystem sizes, the DMM energy range is larger than the energy range found with the D&C method at $-541,693.55$ eV to $-541,693.65$ eV. For the dehydrated system, the Otsuka *et al* energy range is between $-107,258.9$ eV to $-107,259.2$ eV and for the work presented here it is between $-107,950.0$ eV to $-107,950.5$ eV. The similar energy values are due to the use of the same SIESTA PAO basis set functions by both codes. Since the PAOs and pseudopotentials are equivalent, the main difference in the energy ranges will be due to the difference in numerical integration techniques.

The Otsuka *et al* [35] hydrated energy curve converges from a maximum difference of ≈ 0.45 eV at a cut-off of ≈ 7.4 Å to the fully converged energy at a cut-off distance of 13.23 Å. The rate of convergence of the D&C method is faster in this case; for a subsystem radius of 7 Å the energies fully converge from an energy difference of -0.06 eV. Due to the faster convergence rate of the D&C method, the size of the DMM cut-off distance is larger than the D&C subsystem radius for converged values. In this regard, Otsuka *et al* report a cut-off distance of 10 Å is large enough for quantitative results using the DMM method, while it is found here that a subsystem radius between 7-8 Å is required for the D&C method. The D&C method is found to converge to the real energy at a faster rate than the DMM method.

The average time to complete an SCF cycle for the energy convergence runs is shown in Figure 6.6(a). The time takes into consideration the assembly of the Hamiltonian and the diagonalisation of the Hamiltonian using the D&C method. The average time required to communicate data amongst the compute nodes is shown in Figure 6.6(b).

With moderate computing resources, the D&C implementation is found to be very efficient for both the hydrated and dehydrated 1WQZ DNA systems. A converged subsystem radius of 8 Å for the hydrated SZ system takes only an average of 1.3 mins to complete an SCF cycle using 32 cpus. The hydrated DZP system requires on average 36.77 mins with an 8 Å subsystem. For the dehydrated

greater than the cut-off distance. This introduces sparsity and allows the minimisation method to scale linearly. The cut-off distance is considered as a localisation region and is analogous to the localisation regions of the D&C method i.e. the subsystems.

Table 6.2: The distribution of atoms amongst the compute nodes for the hydrated and dehydrated 1WQZ DNA structures. The hydrated system uses 32 cpus, while the dehydrated system uses 24 cpus. The dehydrated system is poorly load balanced which in turn increases the time required to communicate data amongst the compute nodes.

Hydrated system - 3439 atoms		Dehydrated system- 634 atoms	
Compute Node	Atom Count	Compute Node	Atom Count
0	89	0	30
1	94	1	29
2	99	2	17
3	94	3	17
4	101	4	39
5	94	5	33
6	89	6	13
7	107	7	29
8	105	8	40
9	97	9	26
10	97	10	21
11	125	11	34
12	122	12	26
13	101	13	43
14	110	14	3
15	126	15	29
16	99	16	35
17	99	17	5
18	99	18	29
19	110	19	18
20	110	20	24
21	90	21	33
22	94	22	5
23	110	23	56
24	114		
25	112		
26	115		
27	136		
28	130		
29	119		
30	118		
31	134		

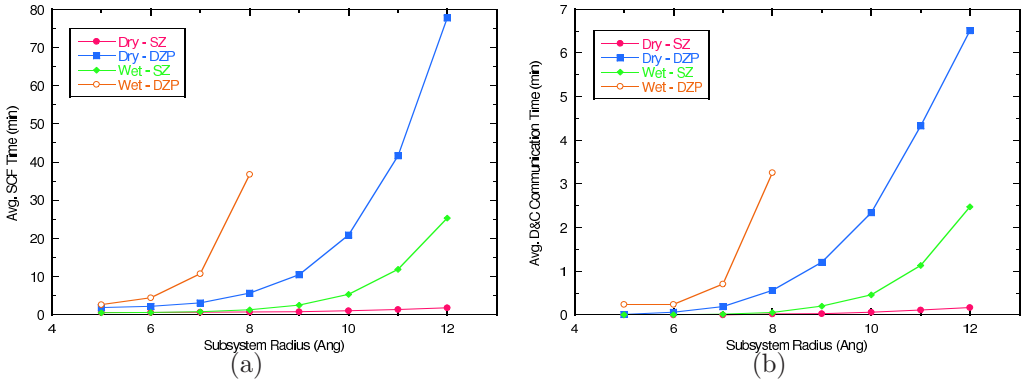


Figure 6.6: The a) average SCF time and b) average time spent communicating data amongst compute nodes for both dry and hydrated 1WQZ DNA systems with increasing subsystem size.

SZ system, a converged subsystem radius of 9 Å completes an SCF cycle in an average time of 0.85 mins using 24 cpus, while the DZP system requires an average of 10.6 Å. For each subsystem size, the calculation time required for dehydrated DZP system is larger than the hydrated SZ system, even though the hydrated SZ system has an extra 2805 atoms and an extra 1045 orbitals. This is due to the poor load-balancing of the dehydrated system data across the 24 cpus. Table 6.2 is a listing of the distribution of the atoms across the cpus. The hydrated system has relatively equal number of atoms on each cpu, while the dehydrated system is clearly poorly load-balanced, with some cpus having as low as 3 atoms on a single cpu, conversely, the largest number of atoms on a cpu is 56. Typically, this means that a small number of cpus are doing the majority of the work, while the rest are idling. The average data communication times during each SCF cycle also reflect the poor load-balancing, as shown in Figure 6.6(b). The smaller dehydrated DZP system spends more time on average communicating data between the compute nodes, than the other systems except for the hydrated DZP system. The CONQUEST times are not available at the time of writing the thesis. It would be interesting to compare the different codes as the prefactor of the DMM method should be less than the D&C method. This makes for a better comparison between the localisation regions of both linear-scaling methods.

The electronic structure of the hydrated and dehydrated 1WQZ DNA systems is now examined using a larger DZP basis set. Based of the SZ and DZP energy

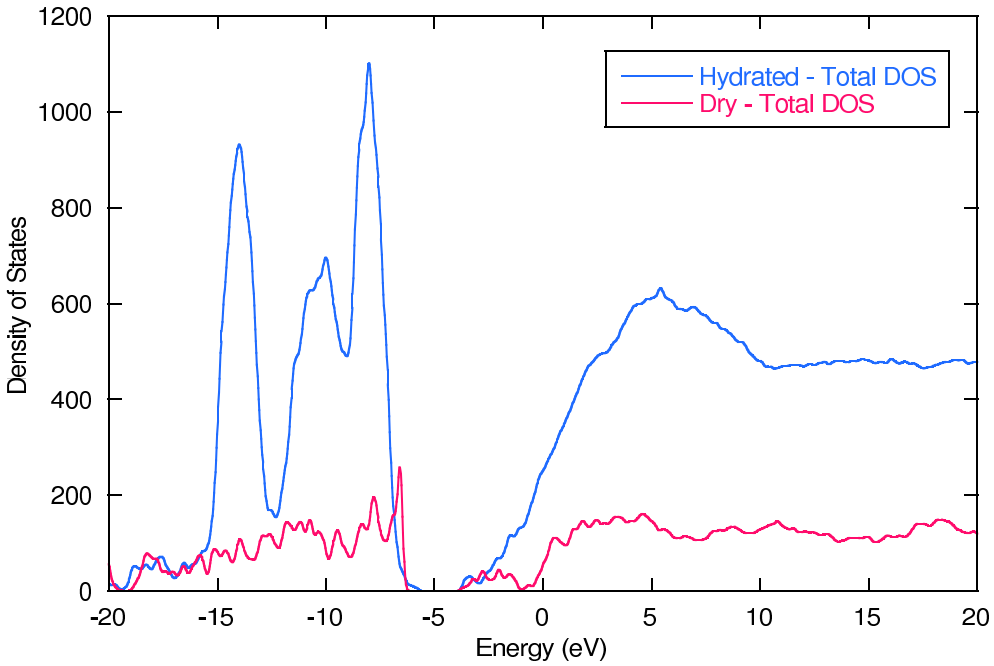


Figure 6.7: The total DOS of the dehydrated and hydrated 1WQZ DNA data structure. The band gap of the dehydrated system is ≈ 2.13 eV and the band gap of the hydrated system is ≈ 1.5 eV.

convergence calculations, a 8 Å subsystem is used for both sets of calculations. It is also the largest subsystem size that can fit into the available memory with a DZP basis set. The total DOS of the hydrated and dehydrated systems is shown in Figure 6.7. The detailed landscape of the dry system is broadened with the inclusion of water molecules, producing three major peaks in the occupied space. The valence band states are shifted towards lower energies in the hydrated system. The band gap of the dehydrated system is ≈ 2.13 eV, while the band gap of the hydrated system is ≈ 1.5 eV. The peak near the band-edge of the hydrated system now has a small tail which reduces the band gap of this system. This is a reduction of 0.63 eV, which is smaller, although comparable to difference found by Kratochvilova *et al* [217] at 0.79 eV for a smaller DNA strand consisting of 5 base pairs, 5-d(TCGGA)-3, and solvated with only 110 water molecules. To determine which states are responsible for the reduction in the band gap the PDOS is analysed. Shown in Figure 6.8, the PDOS indicates that the carbon 2s and 2p states have the largest contribution to the valence band edge of the hydrated system, followed by the nitrogen 2s and 2p states and the hydrogen

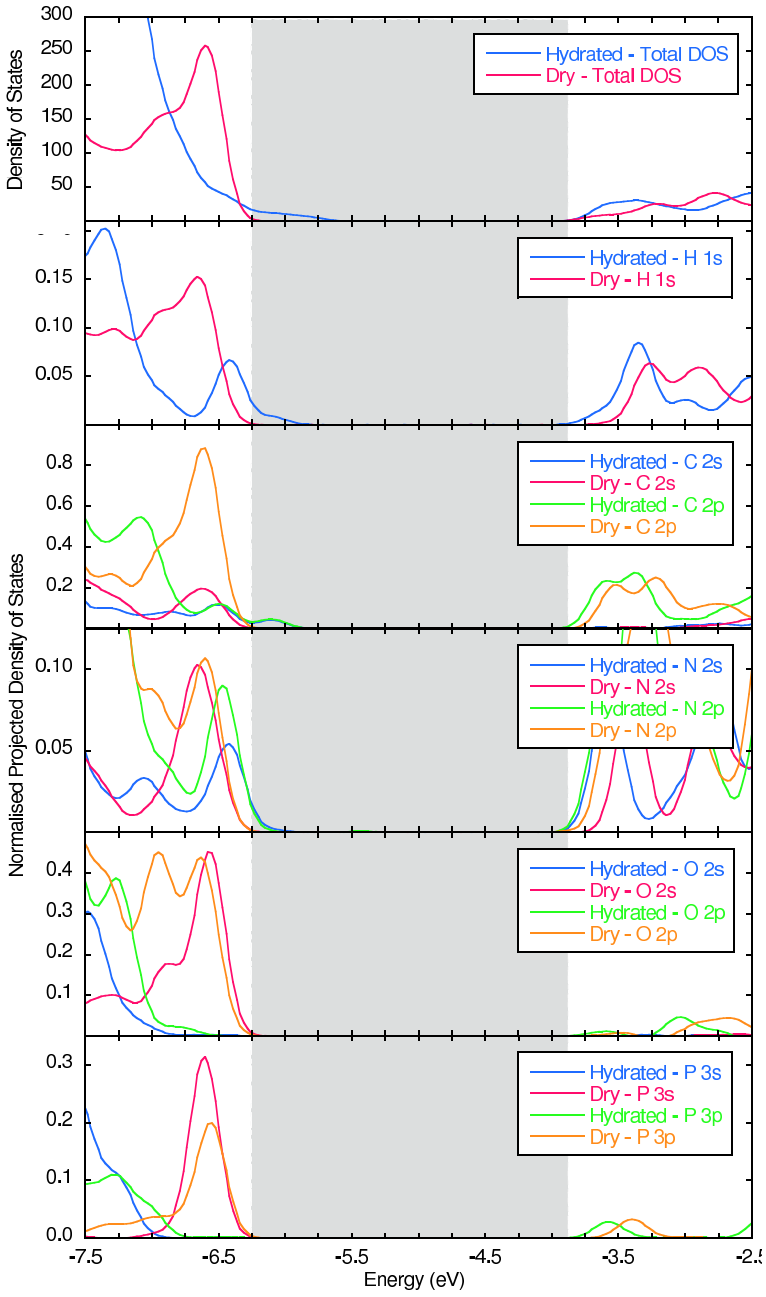


Figure 6.8: The normalised PDOS located about the band gap for both the hydrated and dry 1WQZ DNA systems. The shaded area indicates the band gap for the dry DNA system. The hydrated DNA system reduces the size of the band gap predominantly due to the carbon states, with smaller contributions from the hydrogen and nitrogen states.

1s states. These states are shifted towards higher energies in the valence band, while they are shifted towards lower energies in the conduction band. Conversely, the hydrated oxygen and phosphorus states are shifted towards lower energies in both bands. The p states have the largest contributions in the conduction band in both systems.

The partial atomic charges of the dehydrated and hydrated 1WQZ DNA systems are now compared, using the Mulliken population analysis method [177] to calculate the charges. The atomic charges can be used in the electrostatic energy of MD force fields [218]. Mulliken population analysis has been successfully used in the study of DNA [219,220]. Others have examined Mulliken charges in similar applications, such as Bende *et al* [221] who have used Mulliken population analysis to calculate the charge transfer between the DNA phosphate group and lysine (and arginine) side chains of histone proteins in water. Using the D&C method allows for large realistic DNA systems to be modelled, which can produce charges more appropriate to the system under investigation in the MD simulation.

The nucleotides, phosphate groups and the 2-deoxyribose molecules found within the DNA molecule are shown in Figure 6.9. The atoms are labelled for reference when listing the partial atomic charges. The labelling is arranged in a clockwise fashion for all molecules. The 5' end of the DNA strand has a terminal phosphate group and the 3' end a terminal hydroxyl group. Conventionally, the direction along the backbone starting from the 5' end to the 3' end is called *downstream*, conversely, the *upstream* direction is from the 3' end to the 5' end.

The partial atomic charges for the hydrated 1WQZ DNA molecule can be found in the following tables. The partial charges for each individual atom are listed for the nucleotides, while for the phosphate groups and 2-deoxyribose molecules the average partial charges of each atom are listed. The values in the parenthesis are the partial charges of the hydrated molecule relative to the partial charges of the dehydrated molecule. For example, the cytosine N1 atom located at base 1 (Table 6.3) has a partial atomic charge of 0.438 a.u., which is −0.017 a.u. more negatively charged than the equivalent atom in the dehydrated system. The total partial charges for each molecule are also listed.

Each class of molecule within the dehydrated system is found to participate

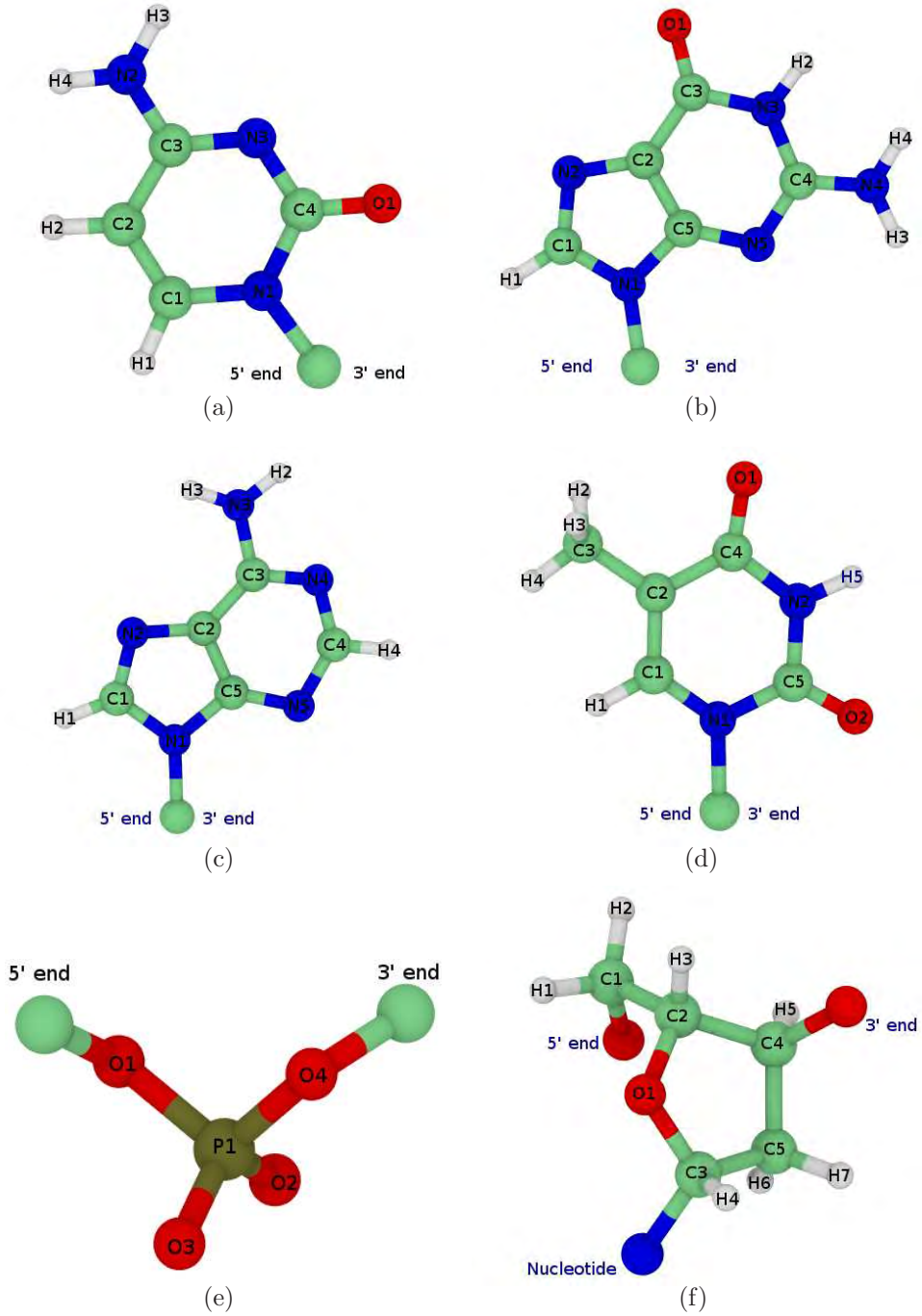


Figure 6.9: Labelling of atoms within the nucleotides and the backbone of the 1WQZ DNA structure. The numbering scheme works in a clockwise fashion. The molecules are orientated so that left of the image is the 5' end and right of the image is the 3' end. a) cytosine b) guanine c) adenine d) thymine e) phosphate group f) 2-deoxyribose.

in charge transfer when the system becomes hydrated, as can be seen from the total charge differences. The cytosine nucleotides (Table 6.3), the thymine nucleotides (Table 6.3) and the 2-deoxyribose molecules (Table 6.8) have the least amount of charge transfer. The hydrated cytosine nucleotides are found to be more negatively charged while the thymine nucleotides and the 2-deoxyribose molecules are slightly more positively charged than their equivalent dehydrated molecules. The hydrated cytosines are generally more positively charged than their dehydrated counterparts, conversely the majority of the thymines and the 2-deoxyribose molecules are slightly more negatively charged. The largest charge transfers occur with the phosphate groups, the guanine nucleotides and the adenine nucleotides. These molecules are all typically found to be more negatively charged than their equivalent dehydrated molecules.

To clarify that the difference in charge between the dehydrated and hydrated DNA systems is due to charge transfer from the water molecules and not just an rearrangement of charge, the total charge of the DNA molecules is calculated. The dehydrated DNA molecule is found to be only slightly positively charged at 0.167 a.u., while the hydrated DNA molecule has a total charge of -5.357 a.u. A charge transfer of 5.357 a.u. has occurred from the solvent to the DNA molecule. The phosphate groups are found to be the predominant molecule in which the charge is transferred from the solvent. For each phosphate group there is a charge transfer of 0.166 a.u. on average for downstream phosphate groups and 0.189 a.u. on average for upstream phosphate groups. Because there are a total of 18 phosphate groups in the backbone of 1WQZ DNA this amounts to 3.201 a.u. of the total charge transfer that has occurred.

6.1.3 Concluding Remarks

Comparison with the energy curve of Otsuka *et al* [35], shows that for the convergence in energy, a smaller localisation region is required in the D&C method than the DMM method, and that convergence occurs at a faster rate with the D&C method. The calculation times for converged subsystem radii have been found to be very efficient on modest computing resources. These calculations all ran below using 3 GB per CPU RAM. High quality calculations were used to

Table 6.3: The Mulliken partial atomic charges (a.u.) of the cytosine nucleotides in the hydrated 1WQZ DNA system. The values in parentheses are the charge transfer of the hydrated DNA relative to dehydrated DNA. The bases are numbered starting from the 5' end *downstream* to the 3' end using the numbering scheme shown in Figure 6.9(a).

Atom	Base 1	Base 2	Base 9	Base 10
N1	0.438 (-0.017)	0.480 (-0.017)	0.462 (-0.012)	0.492 (-0.015)
C1	-0.090 (0.003)	-0.128 (0.000)	-0.141 (-0.019)	-0.113 (0.005)
H1	0.059 (0.014)	0.069 (0.012)	0.059 (-0.002)	0.062 (0.012)
C2	-0.025 (0.005)	-0.041 (-0.011)	-0.051 (-0.006)	-0.043 (-0.020)
H2	0.016 (0.006)	0.007 (0.008)	0.012 (0.001)	0.027 (0.022)
C3	-0.248 (0.015)	-0.294 (0.005)	-0.261 (-0.012)	-0.191 (0.018)
N2	0.289 (0.012)	0.321 (0.021)	0.300 (0.012)	0.278 (-0.005)
H3	0.029 (0.003)	0.016 (-0.003)	0.012 (-0.013)	0.014 (-0.014)
H4	0.014 (-0.017)	-0.006 (-0.009)	0.003 (-0.011)	0.039 (0.019)
N3	0.282 (-0.026)	0.297 (0.003)	0.273 (0.001)	0.208 (-0.004)
C4	-0.558 (0.007)	-0.464 (0.020)	-0.511 (-0.017)	-0.519 (0.011)
O1	-0.032 (0.000)	-0.055 (0.005)	-0.028 (0.024)	-0.040 (0.022)
Total	**0.174 (0.005)**	**0.202 (0.034)**	**0.129 (-0.054)**	**0.214 (0.051)**

Table 6.4: The Mulliken partial atomic charges (a.u.) of the guanine nucleotides in the hydrated 1WQZ DNA system. The values in parentheses are the charge transfer of the hydrated DNA relative to dehydrated DNA. The bases are numbered starting from the 5' end *downstream* to the 3' end using the numbering scheme shown in Figure 6.9(b).

Atom	Base 1	Base 2	Base 9	Base 10
N1	0.484 (-0.004)	0.512 (-0.001)	0.530 (-0.002)	0.492 (-0.006)
C1	-0.332 (-0.014)	-0.401 (-0.028)	-0.355 (-0.030)	-0.348 (-0.044)
H1	0.064 (-0.007)	0.073 (0.010)	0.076 (0.011)	0.062 (-0.007)
N2	0.148 (-0.037)	0.199 (0.019)	0.081 (-0.039)	0.199 (0.045)
C2	-0.140 (-0.025)	-0.076 (-0.049)	-0.058 (-0.028)	-0.119 (-0.040)
C3	-0.283 (0.004)	-0.231 (0.005)	-0.275 (-0.017)	-0.344 (-0.009)
O1	-0.053 (-0.035)	-0.132 (-0.046)	-0.027 (0.004)	-0.012 (-0.027)
N3	0.408 (0.006)	0.393 (0.010)	0.404 (-0.003)	0.434 (-0.001)
H2	-0.008 (-0.002)	-0.015 (0.004)	-0.015 (0.001)	-0.028 (0.002)
C4	-0.390 (-0.014)	-0.399 (-0.006)	-0.467 (-0.021)	-0.446 (-0.011)
N4	0.270 (-0.007)	0.271 (-0.026)	0.295 (-0.018)	0.293 (-0.031)
H3	0.031 (-0.004)	0.025 (-0.007)	0.015 (-0.028)	0.021 (-0.011)
H4	0.012 (-0.004)	0.006 (0.000)	0.004 (-0.021)	0.011 (-0.015)
N5	0.169 (-0.026)	0.210 (0.029)	0.171 (-0.035)	0.164 (-0.027)
C5	-0.251 (-0.006)	-0.280 (-0.007)	-0.306 (-0.022)	-0.249 (-0.001)
Total	**0.129 (-0.175)**	**0.155 (-0.093)**	**0.073 (-0.248)**	**0.130 (-0.183)**

Table 6.5: The Mulliken partial atomic charges (a.u.) of the adenine nucleotides in the hydrated 1WQZ DNA system. The values in parentheses are the charge transfer of the hydrated DNA relative to dehydrated DNA. The bases are numbered starting from the 5' end *downstream* to the 3' end using the numbering scheme shown in Figure 6.9(c).

Atom	Base 3	Base 4	Base 5
N1	0.436 (-0.006)	0.497 (0.000)	0.499 (-0.006)
C1	-0.323 (-0.037)	-0.315 (-0.023)	-0.357 (-0.022)
H1	0.078 (-0.002)	0.061 (-0.012)	0.088 (-0.003)
N2	0.171 (0.053)	0.139 (0.023)	0.172 (0.032)
C2	-0.117 (-0.008)	-0.129 (-0.021)	-0.127 (-0.022)
C3	-0.244 (-0.013)	-0.263 (-0.014)	-0.235 (-0.005)
N3	0.197 (-0.017)	0.258 (-0.035)	0.237 (-0.024)
H2	0.005 (-0.015)	0.019 (-0.011)	0.013 (-0.013)
H3	0.005 (-0.007)	0.019 (0.002)	0.004 (-0.002)
N4	0.310 (-0.004)	0.204 (-0.014)	0.253 (-0.001)
C4	-0.299 (-0.012)	-0.342 (-0.046)	-0.317 (-0.037)
H4	0.052 (0.000)	0.082 (0.019)	0.067 (0.008)
N5	0.178 (0.014)	0.173 (-0.006)	0.170 (-0.008)
C5	-0.266 (-0.005)	-0.265 (-0.014)	-0.247 (-0.024)
Total	**0.183 (-0.059)**	**0.138 (-0.152)**	**0.220 (-0.127)**
	Base 6	Base 7	Base 8
N1	0.477 (0.006)	0.524 (0.003)	0.499 (-0.004)
C1	-0.256 (-0.017)	-0.321 (-0.006)	-0.345 (-0.010)
H1	0.075 (0.004)	0.075 (0.009)	0.073 (0.002)
N2	0.093 (-0.023)	0.124 (0.034)	0.155 (0.023)
C2	-0.103 (-0.004)	-0.097 (-0.005)	-0.109 (-0.001)
C3	-0.294 (-0.014)	-0.227 (-0.014)	-0.317 (-0.025)
N3	0.279 (-0.022)	0.281 (-0.004)	0.258 (-0.026)
H2	0.006 (-0.020)	0.018 (-0.006)	0.006 (-0.006)
H3	0.007 (-0.024)	0.017 (-0.012)	0.001 (-0.005)
N4	0.247 (0.000)	0.227 (-0.010)	0.236 (0.005)
C4	-0.330 (-0.021)	-0.347 (-0.018)	-0.340 (-0.016)
H4	0.066 (0.008)	0.058 (-0.001)	0.057 (0.009)
N5	0.207 (-0.003)	0.170 (0.006)	0.183 (0.015)
C5	-0.287 (-0.028)	-0.264 (-0.006)	-0.268 (-0.005)
Total	**0.187 (-0.158)**	**0.238 (-0.030)**	**0.089 (-0.044)**

Table 6.6: The Mulliken partial atomic charges (a.u.) of the thymine nucleotides in the hydrated 1WQZ DNA system. The values in parentheses are the charge transfer of the hydrated DNA relative to dehydrated DNA. The bases are numbered starting from the 5' end *downstream* to the 3' end using the numbering scheme shown in Figure 6.9(d).

Atom	Base 3	Base 4	Base 5
N1	0.414 (-0.013)	0.491 (-0.009)	0.492 (-0.014)
C1	-0.198 (-0.014)	-0.188 (-0.008)	-0.201 (-0.006)
H1	0.064 (-0.002)	0.073 (0.001)	0.053 (-0.008)
C2	0.140 (-0.010)	0.149 (-0.012)	0.135 (0.003)
C3	-0.124 (-0.026)	-0.113 (-0.015)	-0.120 (-0.013)
H2	0.038 (0.018)	0.017 (0.003)	0.033 (-0.003)
H3	0.029 (0.000)	0.014 (-0.013)	0.030 (0.000)
H4	0.015 (-0.025)	0.030 (-0.001)	0.016 (0.004)
C4	-0.238 (0.005)	-0.284 (-0.004)	-0.299 (0.003)
O1	-0.062 (0.008)	-0.053 (0.019)	-0.051 (-0.003)
N2	0.407 (0.013)	0.444 (0.004)	0.421 (0.005)
H5	-0.011 (0.004)	-0.008 (-0.002)	-0.031 (0.004)
C5	-0.430 (0.011)	-0.416 (-0.013)	-0.399 (-0.015)
O2	-0.010 (0.016)	-0.090 (0.011)	-0.043 (0.015)
Total	**0.034 (-0.015)**	**0.066 (-0.039)**	**0.036 (-0.028)**
	Base 6	Base 7	Base 8
N1	0.472 (-0.001)	0.418 (0.001)	0.471 (-0.005)
C1	-0.190 (0.005)	-0.147 (0.007)	-0.195 (-0.007)
H1	0.074 (0.016)	0.070 (0.002)	0.065 (-0.004)
C2	0.138 (0.000)	0.102 (-0.009)	0.142 (-0.008)
C3	-0.122 (-0.021)	-0.134 (-0.034)	-0.118 (-0.018)
H2	0.057 (0.034)	0.050 (0.008)	0.017 (-0.015)
H3	0.024 (-0.008)	0.046 (0.004)	0.030 (0.001)
H4	0.020 (-0.007)	0.027 (0.005)	0.032 (0.007)
C4	-0.347 (0.017)	-0.331 (0.011)	-0.288 (0.000)
O1	-0.022 (0.007)	-0.040 (0.019)	-0.030 (0.006)
N2	0.487 (0.003)	0.501 (0.003)	0.472 (0.011)
H5	-0.007 (0.001)	-0.010 (0.002)	0.002 (0.000)
C5	-0.454 (-0.008)	-0.458 (-0.010)	-0.456 (0.004)
O2	-0.007 (0.013)	-0.008 (0.022)	-0.015 (0.006)
Total	**0.123 (0.051)**	**0.086 (0.031)**	**0.129 (-0.022)**

Table 6.7: The *average* Mulliken partial atomic charges (a.u.) of the phosphate groups found in the backbone. The values in parentheses are the average charge transfer of the hydrated DNA relative to dehydrated DNA. The average charges are grouped by the two backbones of the DNA. The atoms are numbered using the numbering scheme shown in Figure 6.9(e)

Atom	Downstream	Upstream
O1	0.005 (-0.019)	0.033 (-0.013)
P1	-0.291 (-0.033)	-0.326 (0.036)
O2	-0.186 (-0.096)	-0.162 (-0.089)
O3	-0.156 (-0.065)	-0.194 (-0.110)
O4	0.034 (-0.020)	0.044 (-0.014)
Total	**-0.593 (-0.166)**	**-0.605 (-0.189)**

Table 6.8: The *average* Mulliken partial atomic charges (a.u.) of the 2-deoxyribose molecules found in the backbone. The values in parentheses are the average charge transfer of the hydrated DNA relative to dehydrated DNA. The average charges are grouped by the two backbones of the DNA. The atoms are numbered using the numbering scheme shown in Figure 6.9(f)

Atom	Downstream	Upstream
C1	-0.066 (-0.017)	-0.059 (-0.022)
H1	0.050 (0.006)	0.050 (0.003)
H2	0.066 (-0.003)	0.066 (-0.006)
C2	0.018 (-0.005)	0.003 (-0.006)
H3	0.051 (0.008)	0.049 (0.007)
O1	-0.012 (-0.007)	0.003 (-0.005)
C3	-0.146 (-0.009)	-0.146 (-0.013)
H4	0.074 (0.012)	0.072 (0.008)
C4	0.025 (-0.020)	0.018 (-0.018)
H5	0.052 (-0.004)	0.058 (0.002)
C5	-0.035 (-0.007)	-0.032 (-0.008)
H6	0.030 (0.002)	0.035 (0.005)
H7	0.034 (0.003)	0.031 (0.002)
Total	**0.140 (-0.040)**	**0.147 (-0.050)**

compare the electronic structures of the hydrated and dehydrated systems. The hydrated system reduces the band-gap of the dry system. Charge transfer from the water solvent molecules to the phosphate groups in the DNA backbone alters the partial atomic charges of the nucleotides, in particular for the guanine and adenine nucleotides. For the studied DNA systems, the D&C method proves to be a very efficient method.

6.2 Zeolitic Imidazolate Framework - ZIF100

In this final section, the ability of the current D&C implementation is tested on a very large system that is well suited to the method. The system under study belongs to a recent class of materials named zeolitic imidazolate frameworks (ZIF) [222]. ZIFs are porous crystalline materials that resemble natural aluminosilicate zeolites with their cage-like structure. The tetrahedrally coordinated silicon found in the zeolite is replaced by a transition metal, in this case zinc, and the oxygen bridges are replaced by imidazolate links, in this case 5-chlorobenzimidazolate (cbIM), see Figure 6.10(a). The links and the metals determine the ZIFs resulting structure. By substituting linkers, different ZIF topologies can be found and predicted [223, 224], similar to metal-organic framework (MOF) materials [225–227].

The particular ZIF crystal examined in this section is named ZIF-100, first synthesised by Wang *et al* [40]. The ZIF-100 cubic unit cell is very large containing 13,584 atoms with a lattice constant of 71.9797 Å. The porous structure and local nature of the chemistry of this system is ideally suited for the D&C method.

ZIFs were created for the purpose of selectively capturing specific molecules [228] from a mixture of different gases. The ZIF-100 crystal can selectively capture carbon dioxide from several different gas mixtures and is capable of storing large quantities of the gas at standard room temperature and pressure. Wang *et al* [40] report that one litre of ZIF-100 material can store up to 28.1 litres of carbon dioxide at 273 K and 15.9 litres at 298 K. Wang *et al* showed this by measuring the adsorption isotherms of various gases, including carbon dioxide,

methane, carbon monoxide and nitrogen. They found the ZIF-100 structure to have a high storage capacity and a partiality to carbon dioxide. The selective absorption characteristics of ZIFs and their sponge-like behaviour can be used in the absorption of green house gases from industrial gas emission streams. leading to a reduction of the uptake of green house gases into the atmosphere. Though it does not provide a full solution to the problem, it can help reduce the impact of current fossil fuel burning technologies on the environment.

There have been numerous studies involving MOFs [229–231] and in particular their hydrogen adsorption properties using DFT [232–234] and MD [235–237] simulations. Computational studies on various ZIFs have also been performed [238–240]. The most interesting are the investigations of the ZIF-68, ZIF-69 and ZIF-70 structures [241–243], where the partial charges on the atoms were calculated using DFT and then used to set up electrostatic interactions within force fields, such as the Universal Force Field (UFF) [244]. These MD simulations are used to investigate adsorption properties of various gases, with the atomic partial charges derived from either the full unit cell of the ZIF or a small fragment moiety. There have been no first principles studies involving ZIF-100 due to its large size and the computational demands of standard diagonalisation techniques. A linear-scaling method, such as D&C, is required for this task and is used in the following work.

Due to its capabilities, the ZIF-100 structure is potentially a very important material. Studies on the electronic properties are necessary to better understand these capabilities and to highlight possible improvements. In this thesis, the electronic structure of the full unit cell of the ZIF-100 is found using the current D&C implementation. The full unit cell is compared with a fully optimised cluster fragment moiety of the complete unit cell and a single cbIM molecule. Comparing the cluster to the full unit cell will help determine if fast cluster calculations can be used to represent the complete system. The findings from the electronic structure calculations of the full unit cell of ZIF-100 can be used to parameterise potentials within MD simulations, as this has not been previously been achieved.

Firstly, the convergence of the total energy with respect to the subsystem size

is investigated using a low quality calculation. Knowledge of the required size of the subsystem for a low quality calculation can provide an estimate when considering high-quality calculations. The reason for not using high-quality calculations in the convergence test is due to the large amounts of memory required for large subsystem sizes. A reasonably sized subsystem needs to be chosen that will provide accurate results and be able to reside in the available memory. The partial atomic charges, using Mulliken population analysis [177], of the full ZIF-100 unit cell are found and compared to the partial charges of the small fragment moiety and the cbIM molecule. The charge densities and the (P)DOS are also found for the ZIF-100 structure and are compared to those of the small fragment moiety.

6.2.1 Geometry Disorder

Crystallographic data for the ZIF-100 unit cell, as determined via X-ray diffraction, can be found in the Cambridge Crystallographic Data Centre (CCDC) with the deposition code, 668215. The atomic positions as they stand within the CCDC database exhibit various amounts of disorder. Before any simulations can be performed the disorder within the structure must be removed. The following is a brief account of the clean up procedures performed on the ZIF-100 structure.

Before any disorder can be removed, the geometric structure of the cbIM molecule needs to be found to determine optimal bond lengths and angles between certain atoms. A DFT geometry optimisation of a single cbIM molecule terminated with hydrogens was performed to find the optimal bond lengths and bond angles between the carbon, chlorine, nitrogen and hydrogen atoms (Figure 6.10(a)). Employing the SIESTA [2–6] method, the calculations were performed using the PBE [54] parameterisation of the GGA XC and norm-conserving Troullier-Martins pseudopotentials [88, 91] in the Kleinman-Bylander factorised form [89,90]. The employed pseudopotential reference valence configurations were $3s^2 3p^6 3d^{10} 4f^0$, $2s^2 2p^2 3d^0 4f^0$, $2s^2 2p^3 3d^0 4f^0$, $3s^2 3p^5 3d^0 4f^0$ and $1s^1 2p^0 3d^0 4f^0$ for Zn, C, N, Cl and H, respectively. The cut-off radii for each angular momentum channel, s, p, d and f, were specified as follows; for Zn, $0.80a_0$, $1.15a_0$, $0.80a_0$ and $2.14a_0$; for C, $1.19a_0$ (all channels); for N, $1.14a_0$ (all channels); for Cl, $1.66a_0$, $1.66a_0$, $1.88a_0$, $1.88a_0$; and for H, $1.25a_0$ (all channels). A DZP basis set on all

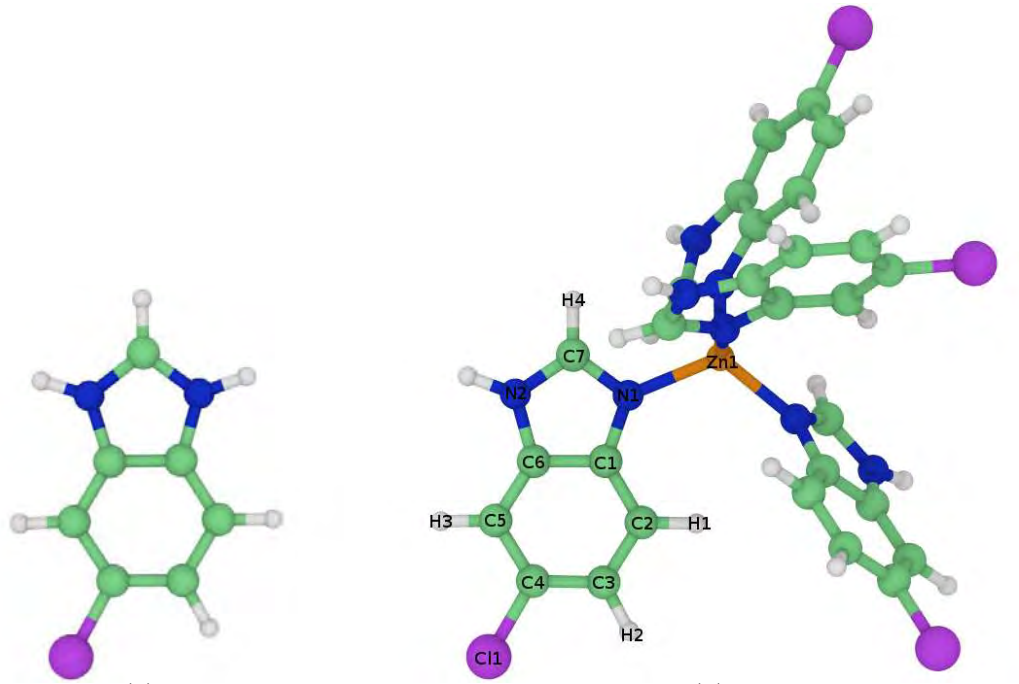


Figure 6.10: Fragment components of the ZIF-100 structure. Hydrogen atoms are white, carbon atoms are green, nitrogen atoms are blue, chlorine atoms are purple and zinc atoms are orange. a) Hydrogenated cbIM molecule used in the clean up procedure of the disordered ZIF-100 unit cell. b) Tetra-Zn-cbIM - A small fragment moiety of ZIF-100. Consists of four hydrogenated cbIM molecules tetrahedrally bound to a central zinc atom. The atom labelling is used for the partial atomic charges.

atoms with an orbital confinement energy of 0.005 Ry was used in the present calculation. The cut-off energy for the real-space integration grid was set to 350 Ry and where self-consistency was achieved a tolerance of $1x10^{-5}$ in the density matrix convergence criteria was used. The forces on the atoms were calculated and were allowed to relax using the CG minimisation technique until their residual forces had converged to less than $0.02\,\text{eV}\ \text{Å}^{-1}$.

The average carbon-hydrogen bond length was found to be 1.100 Å and the carbon-carbon-hydrogen bond angle on average was 120.0°. The carbon-chlorine bond length was found to be 1.732 Å. Now that the structural properties of the free standing cbIM molecule are known, disorder within the ZIF-100 X-ray-diffraction structure data can now be fixed according to these properties.

The solvent water molecules in this case do not contribute to the stability of the ZIF-100 structure and are removed. Through visual examination of the asymmetric unit, two whole cbIM molecules were found to exhibit disorder, where for each cbIM molecule another overlapping cbIM molecule was found. One of the cbIM molecules had to be removed; the choice was determined by which of the overlapping molecules was the least skewed i.e. the isolated cbIM molecule is planar. The positions of the uppermost carbon-hydrogen atoms on one of the cbIM molecules that was not removed were incorrect, found by comparing with the other cbIM molecules in the asymmetric unit cell. The position of these nitrogen-carbon-hydrogen atoms were corrected to the bond angle of 120° and the carbon-hydrogen bond length set to 1.100 Å. Again, in the asymmetric unit cell, a single chlorine atom was found unbonded to its closest carbon atom. The chlorine atom was moved and placed in the plane of the cbIM molecule at a distance of 1.732 Å (as found in the cbIM geometry optimisation) from the closest carbon atom. On the majority of the cbIM molecules there exists two bonded chlorine atoms instead of one. The chlorine atoms have a 50% occupancy which allows an easy conversion to hydrogen, where a choice is made to convert one of the chlorine atoms to hydrogen and shorten the bond length to 1.100 Å.

All the disorder discernible from the asymmetric unit cell has been found and corrected. The full unit cell must be constructed to remove the remaining disorder. Constructing the full unit cell from the asymmetric unit cell produces more chlorine atoms in place of a hydrogen atom. A python script was written to substitute every *other* chlorine atom, with a 50% occupancy, with a hydrogen and the bond length is appropriately shortened to 1.100 Å. There remains two sources of disorder still left in the full unit cell. One is chlorine-chlorine bonded atoms and the other is when no chlorine atoms are found on the cbIM molecules. Once again a python script was used to find when these conditions occur and substitute the particular atom with either a chlorine or a hydrogen.

Once all the disorder has been removed, the positions of the hydrogens needed to be optimised. This was done using the molecular mechanics program, GULP [7]. The positions of all non-hydrogen atoms are fixed while the hydrogen atoms are constrained by potentials that limit the bond length, the bond angle of the

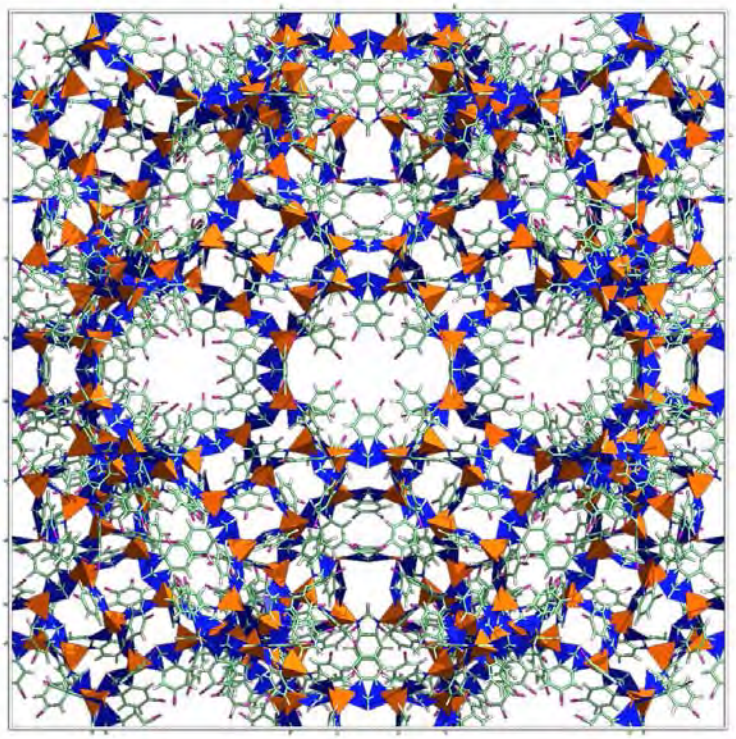


Figure 6.11: The 13,584 atom unit cell of cubic ZIF-100 shown in the $(00\bar{1})$ plane. The portals into the hollow centre of the structure are easily discernible. Hydrogen atoms are white, carbon atoms are green, nitrogen atoms are blue (represented as tetrahedrons), zinc atoms are orange (represented as tetrahedrons) and chlorine atoms are purple.

carbon-carbon-hydrogen atoms and potentials to keep the hydrogen atom in the plane of the cbIM molecule. The bond length is governed by a harmonic potential with a force constant set to $100.0\,\mathrm{eV\AA}^{-2}$, the bond angle is governed by the three body harmonic potential with a force constant of $100.0\,\mathrm{eV\ rad}^{-2}$ and the torsional constraint is governed by the out of plane potential with a force constant of $100.0\,\mathrm{eV\AA}^{-2}$.

The final clean structure can now be used in any forthcoming calculations. Figure 6.11 shows a visualisation of the unit cell of cubic ZIF-100 shown in the $(00\bar{1})$ plane as used in the following electronic structure calculations.

6.2.2 Results

There have been two primary investigations performed on ZIF-100. The first investigation looks into the convergence of the total energy with respect to the subsystem radius (including an outer buffer region of 1.0 Å) using a SZ basis set on all atoms with an orbital confinement of 0.01 Ry and a 450 Ry integration grid. The calculation time of the first SCF cycle is also shown for the convergence test. The ZIF-100 is the largest structure reported in this thesis[c], and with the current D&C memory model (see section 3.5.2) it was difficult to complete the convergence tests using a larger basis set.

The second investigation is a comparison of the electronic structure of the complete ZIF-100 unit cell with the electronic structure of a small cluster fragment moiety and the cbIM molecule. Aspects of the electronic structure that are compared are the (P)DOS, the partial atomic Mulliken charges and the charge density. These results can be used to set up force fields in a molecular mechanics simulation. A DZP basis set is used in this case, with an orbital confinement of 0.01 Ry on a 450 Ry integration grid. The radius of the standard subsystem was set to 7.337 Å which is 80 % of the maximum Hamiltonian interaction length.

For both of the above investigations the same pseudopotential configuration as the previous cbIM calculation is employed. The calculations were performed in parallel using 64 processors of a 1472-processor SUN Constellation (2.93GHz Intel Nehalem cpus) using up to 3 GB of RAM per CPU for all calculations.

6.2.2.1 Energy Convergence

Convergence of the total energy (including the energy error per atom) with respect to subsystem size is shown in Figure 6.12(a). Each subsystem has an extra 1.0 Å outer buffer region included. The error per atom is calculated relative to the total energy obtained with a subsystem radius of 10.0 Å. The energy differences are found to be small, with accurate results obtained at an error of −0.00005 eV/atom, using a subsystem radius of 7.0 Å. Even for a 5.0 Å subsystem radius, the error is small at 0.00036 eV/atom. The plot suggests a 9.0 Å subsys-

[c]The current D&C implementation has achieved a single SCF calculation of a 512,000-atom bulk silicon unit cell, with a SZ basis set, using 256 cpus.

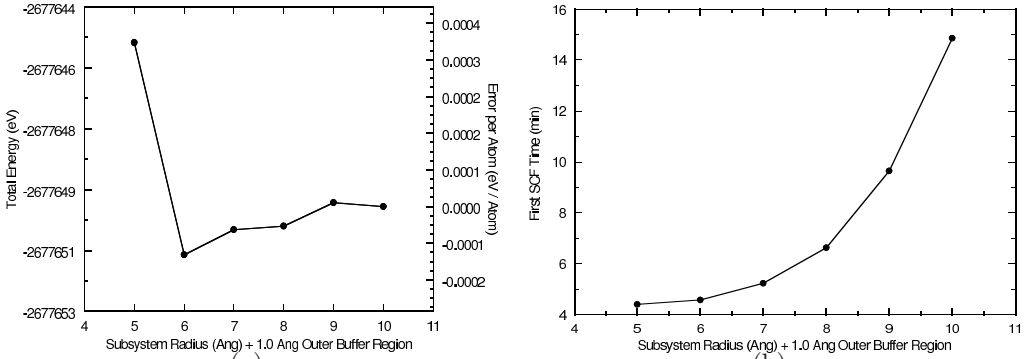


Figure 6.12: Subsystem size dependent computation details of the ZIF-100 structure, using a SZ basis set on 64 cpus. Each subsystem has a 1.0 Å outer buffer region. a) Convergence in the total energy of the ZIF-100 structure as the subsystem size is increased. The total energy converges quickly, with full convergence considered with a 9.0 Å subsystem radius. b) The calculation time of the first SCF iteration for each subsystem size. Efficient calculations are possible with all subsystem sizes, with the SZ basis set.

tem with an error of 0.000008 eV/atom is required for effective full convergence in the total energy.

The calculation time of the first SCF iteration is shown in Figure 6.12(b) for each subsystem size. The times range from 4.4 mins for the 5.0 Å subsystem up to 14.9 mins for the 10.0 Å subsystem. Calculations run with a 7.0 Å subsystem will produce accurate results, which has a first SCF iteration calculation time of only 5.2 mins with the SZ basis set. The first SCF iteration completes in the longest time compared with subsequent SCF iterations, hence, the average SCF time will be lower. An improvement on the times can be achieved through better load-balancing of the atoms on the compute nodes. It is expected with near-perfect load-balancing (which is quite achievable in this case), a reduction of 50% in calculation time is possible. Regardless, the D&C method is found to be very efficient for this system, as is expected due to the porous nature of the material[d].

[d]The D&C method works efficiently with porous materials because neighbouring atoms are sparsely located for any particular subsystem. The subsystems will tend to have a smaller number of basis functions for any given subsystem size, resulting in a smaller subsystem memory footprint and faster subsystem diagonalisation times. The duplication of effort caused by overlapping subsystems is also reduced. Conversely, dense systems with the same volume as

6.2.2.2 Electronic Structure

The electronic structure of the complete ZIF-100 unit cell is compared to the electronic structure of a small ZIF-100 fragment moiety and the cbIM molecule. These results can be used in the setting up of force fields in molecular mechanics simulations. Determination of the (P)DOS, the partial atomic charges (via a Mulliken population analysis [177]), and the charge density is achieved using the same parameters as the SZ basis set calculations, except with a larger DZP basis set. It was found that a subsystem radius of 7.337 Å is close to the largest subsystem that would fit the program into the available memory of 3 GB per cpu when using 64 cpus. This subsystem radius amounts to 80 % of the maximum interaction length within the Hamiltonian, at 9.1713 Å. The employed subsystem size is in the range of converged results according to the SZ calculations, with an error per atom of $\approx$ 0.00005eV. Also, due to the porous nature of the material, it is expected that any chemistry will be primarily governed by local effects which would not require such large subsystems. Finally, the average time per SCF computation was found to be approximately 30 mins.

Using larger subsystem radii would require more cpus to be able to access more memory on the symmetric multiprocessor machine. Only a certain amount of extra cpus can be used before higher scaling occurs due to poor load balancing. More than likely, there will be a point where the data will not be able to reside in the available memory on each compute node, regardless of the combined total memory. In this case, a new memory model will have to be used that distributes the subsystem data amongst many compute nodes; refer to section 3.5.2[e].

The large size of the ZIF-100 structure lends itself to long computational times, even when using order-N methods. It would be worthwhile to investigate if using a small moiety of the structure is representative of the complete structure, within any particular simulation of localised properties. To this end, the electronic structure of the complete ZIF-100 structure is compared to a small fragment moiety, as shown in Figure 6.10(b). The moiety consists of four hy-

the porous system and the same subsystem size will have a larger amount of duplication, a larger subsystem memory footprint and longer subsystem diagonalisation times.

[e]This option has not been implemented at the time of running the calculations, and is a feature that may be investigated in the future.

drogenated cbIM molecules that are tetrahedrally bound to a central zinc atom. The moiety will be referenced as Tetra-Zn-cbIM from this point on. The DOS for the ZIF-100 and along with the PDOS for the Tetra-Zn-cbIM structures are shown in Figure 6.13. The ZIF-100 structure is found to be semi-conducting with a band-gap of $\approx 1.4\,\text{eV}$. DFT typically underestimates band-gaps and is expected that the actual band-gap will be larger. Due to finite-size effects and confinement, the tetra-Zn-cbIM fragment moiety has a larger band-gap of 3.3 eV. Most features found in the ZIF-100 DOS are found in the Tetra-Zn-cbIM DOS. The carbon states contribute to the majority of the DOS. The major peaks of the both sets (ZIF-100 and Tetra-Zn-cbIM) are aligned for states in the conduction band, while for states in the valence band, the Tetra-Zn-cbIM DOS is slightly shifted towards lower energies. The shift is the cause of the larger band-gap for the tetra-cbIM structure.

The PDOS of Tetra-Zn-cbIM is plotted below the DOS in Figure 6.13. The PDOS of the ZIF-100 structure is not available, as the current D&C PDOS algorithm cannot run within the available memory. The PDOS shows that the carbon 2p and nitrogen 2p states near the band edges are hybridised. The chlorine 3p state has a peak near the valence band edge as well. As the total DOS of the ZIF-100 structure is similar to Tetra-Zn-cbIM, it is expected that the PDOS will also be similar.

Most force fields within molecular mechanics use partial atomic charges to define the electrostatic contributions to the potential [218]. The partial charges can be populated from the results of first principles calculations. The partial atomic charges for the ZIF-100 structure have not been found before, due to its large size, and are provided here for the first time. Mulliken population analysis [177] is used to calculate the partial atomic charges in this thesis. Liu *et al* [241] successfully used Mulliken charges for their UFF forcefields of ZIF-68 and ZIF-69, producing accurate carbon dioxide adsorption isotherms. The comparison of the partial charges will be the best indicator of the representability of the Tetra-Zn-cbIM moiety. Table 6.9 lists the partial atomic charges for the cbIM molecule, and a single cbIM molecule plus a zinc atom from the Tetra-Zn-cbIM moiety and the ZIF-100 structure. Averaged charges for each atom are

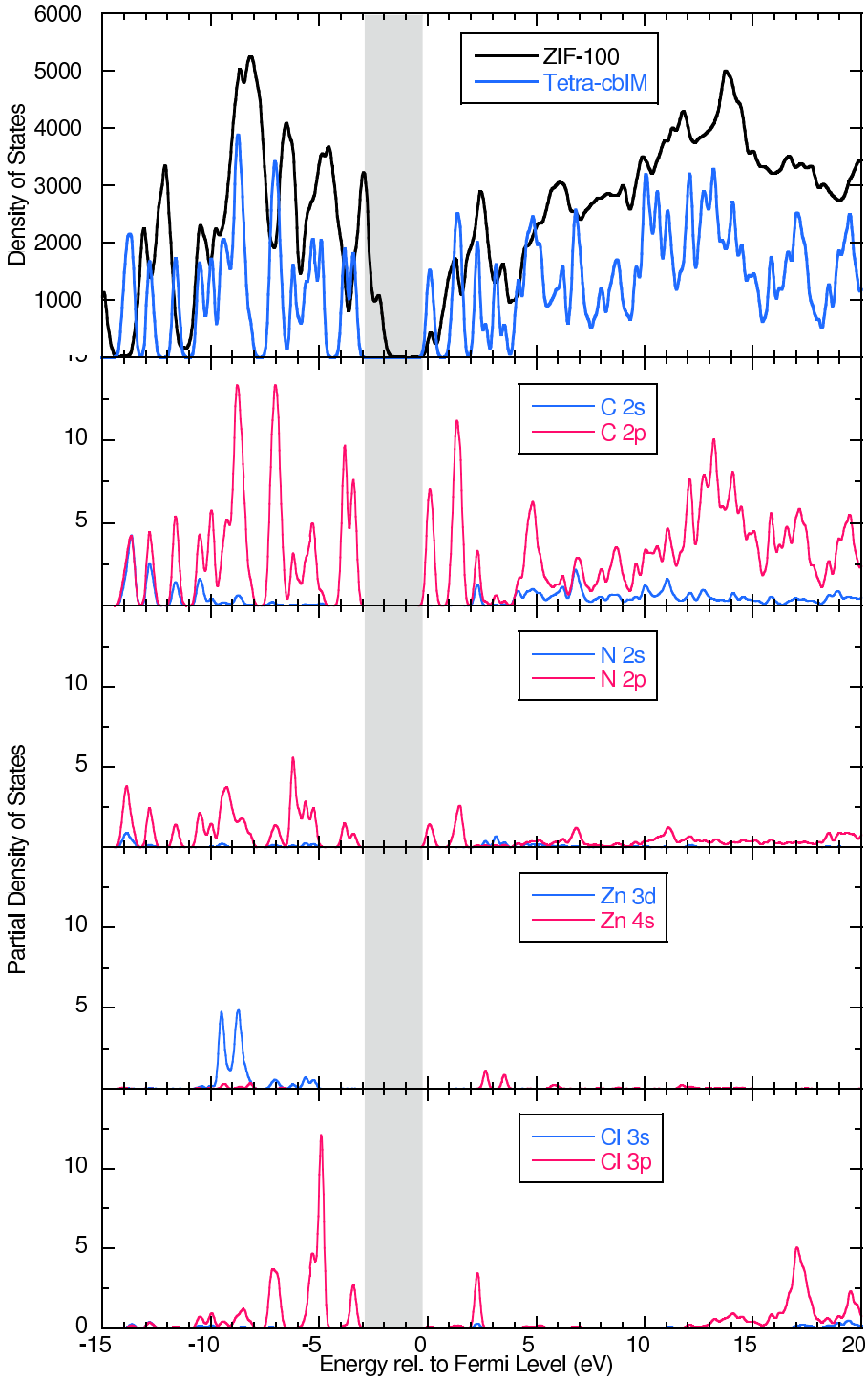
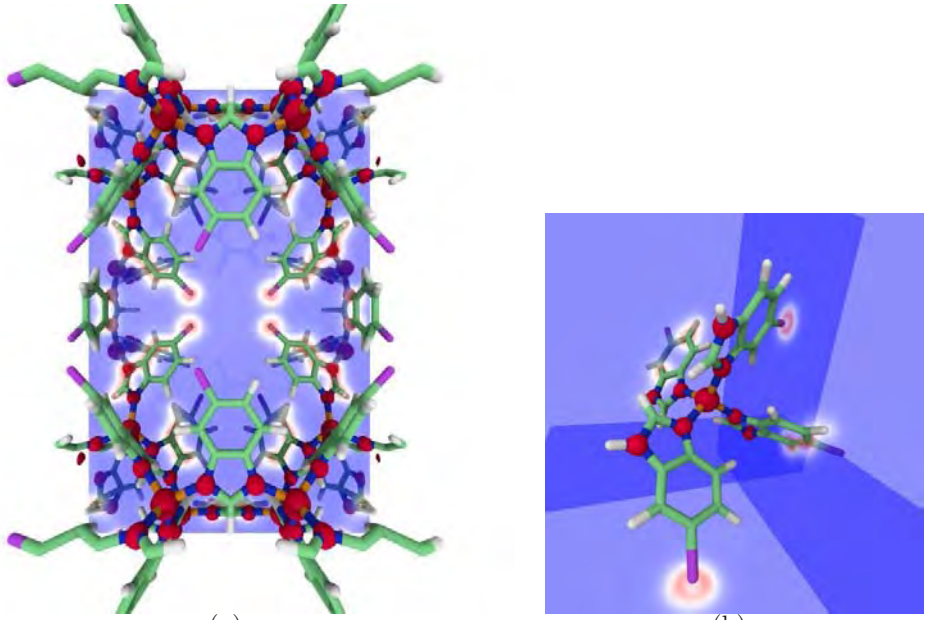


Figure 6.13: The DOS for the ZIF-100 structure and Tetra-Zn-cbIM moiety are shown in the top panel. The size of the ZIF-100 band-gap is $\approx 1.4\,\text{eV}$, while the tetra-Zn-cbIM band-gap is 3.3 eV. The lower panels show the PDOS for Tetra-Zn-cbIM, where the carbon 2p and nitrogen 2p states near the band edges are hybridised.

Table 6.9: The Mulliken partial atomic charges (in units of a.u.) of the cbIM molecule, the Tetra-Zn-cbIM moiety and the ZIF-100 structure. The atoms are numbered according to the numbering scheme shown in Figure 6.10(b). The average charges for a particular atom are shown in parentheses.

Atom	cbIM	Tetra-Zn-cbIM	ZIF-100
N1	0.351	0.114 (0.128)	0.154 (0.165)
N2	0.354	0.350 (0.358)	0.148 (0.170)
C1	-0.082	-0.106 (-0.094)	-0.114
C2	-0.066	-0.045 (-0.050)	-0.021
C3	-0.052	-0.045 (-0.043)	-0.041
C4	0.030	0.031 (0.033)	0.036 (0.032)
C5	-0.102	-0.093 (-0.081)	-0.058
C6	-0.077	-0.079 (-0.080)	-0.076
C7	-0.307	-0.289 (-0.286)	-0.286
H1	-0.029	-0.020 (-0.015)	-0.011
H2	-0.020	-0.012 (-0.011)	-0.030
H3	-0.032	-0.026 (-0.023)	-0.023
H4	0.001	0.017 (0.020)	0.030
Cl1	-0.020	-0.010 (-0.010)	0.030 (0.032)
Zn1	—	0.505	0.476 (0.481)

also listed. Refer to Figure 6.10(b) for a the numbering scheme used to label the individual atoms. The are only small differences in the charges between the three structures, with a few exceptions. For example, because the N1 atom is bonded to the Zn1 atom in the Tetra-Zn-cbIM and ZIF-100 structures, while it is bonded to a hydrogen in the cbIM molecule, there will greater charge transfer to the zinc atom and hence a difference in the charges. In cbIM, the nitrogen atom attracts more charge from the lighter hydrogen and hence has more charge than the N1 atom in Tetra-Zn-cbIM and ZIF-100. The Tetra-Zn-cbIM N1 atom has a charge of 0.114 a.u. which is less than the ZIF-100 N1 atom at 0.154 a.u.. The greater charge on the ZIF-100 N1 atom is due to a greater charge transfer from the Zn1 atom, which is reflected in the lower charge of 0.476 a.u. than in the Tetra-Zn-cbIM (0.505 a.u.). This difference has to be due to the cluster nature of the moiety, as there is no additional charge available from the boundary of the cluster and that long-range potentials help stabilise the greater charge as



(a) (b)

Figure 6.14: Charge density plots of a) an opening into the centre of the ZIF100 structure and b) the Tetra-Zn-cbIM moiety. The red isosurface centred at the nitrogen and zinc atoms has a value of 0.4 electrons/bohr^3. The highest density concentration can be found very close to origin of the zinc atoms at 2.1 electrons/bohr^3. The contour plot shows that the rest of atoms have low density, where the scale range is from 0.0 electrons/bohr^3 (blue) to 0.15 electrons/bohr^3 (white) to 0.3 electrons/bohr^3 (light red). a) ZIF-100 b) Tetra-Zn-cbIM

in periodic ZIF-100. The Cl1 atom charge is negative on the Tetra-Zn-cbIM moiety while it is positive on the ZIF-100 structure, albeit, the difference is only 0.004 a.u.. It must be noted that the approximation inherent within the D&C method will also contribute to the small differences in the charges. Even though there are some differences between the ZIF-100 and Tetra-Zn-cbIM partial atomic charges, the Tetra-Zn-cbIM charges are reasonably close to claim that the moiety is representative of the full periodic system in terms of the atomic charges.

Finally, the electron charge density of the ZIF-100 structure is compared with the charge density of the Tetra-Zn-cbIM moiety. Plots of the charge densities are shown in Figure 6.14. Both structures exhibit similar densities, providing more evidence that the Tetra-Zn-cbIM moiety is representative of the ZIF-100 structure. For both structures, the zinc atoms have the largest amount of electron

density at ≥ 2.1 electrons/bohr^3. The nitrogen atoms have a density concentration up to 0.4 electrons/bohr^3 and with the chlorine and hydrogen atoms up to 0.3 electrons/bohr^3.

6.2.3 Concluding Remarks

The successful determination of the electronic structure of the clean ZIF-100 structure using the D&C implementation has been accomplished. The convergence behaviour of the system relative the size of the subsystems has been found, indicating that a subsystem as small as 7.0 Å can be used for quantitative results. The electronic structure of the ZIF-100 using a large basis and high-quality settings was found. A 7.337 Å subsystem radius was used, as this was the largest radius that could fit into the available memory. The structure was found to be semi-conducting with a band-gap of ≈ 1.4 eV. The electronic structure of small fragment moiety was found and compared to the ZIF-100 structure. It was found that the moiety possessed similar properties and can be considered representative of the ZIF-100 structure when considering local properties. Further studies are required to determine the bonding nature of electrophilic carbon dioxide and ZIF-100. Adsorption isotherms also need to be computed and compared with experimental results [40].

7

Conclusions

The primary aim of this thesis was to investigate first principles linear-scaling methods [19] in the ground and excited states. This was successfully achieved for the ground state by combining the density matrix D&C method [20–22] with the linear-scaling assembly of the Hamiltonian and overlap matrices within the SIESTA DFT code [2–6]. The implementation successfully scaled linearly with respect to increasing system size. The applicability to a variety of systems with decreasing band gaps has been demonstrated. The current parallel implementation allows for electronic structure calculations of large systems consisting of thousands to tens of thousands of atoms with relatively modest computational resources. The electronic structure of systems consisting of hundreds of thousands is expected to be possible with the current implementation and with adequate compute resources. When executed in parallel, the D&C approach exhibits near perfect speedup providing the data distribution is well balanced. The performance of the parallel communication framework needs further examination with larger numbers of compute nodes, in a similar fashion to Bowler and Miyazaki [245] with the CONQUEST code [38]. The findings of the D&C method have been found to be comparable to the KMG functional minimisation method [32] in SIESTA. The standard D&C implementation is found to be robust and efficient.

Extensions to the standard density matrix D&C method that improve molecular dynamics and geometry optimisations were proposed and implemented. The first extension was shown to lessen the effects of discontinuities in the potential energy surface that occur when atoms enter or leave a subsystem. Here, a

switching function [144–151] is applied to each subsystem Hamiltonian and overlap matrix to taper the interactions between the core atoms and buffer atoms near the boundary of the subsystem. The tapering mechanism was applied to the calculation of the interaction energy between two hydrogenated 20.0 Å silicon clusters. The interaction energy curve calculated from standard D&C was discontinuous throughout the range of cluster separation distances. The application of the taper mechanism successfully produced continuous smooth energy curves for taper regions up to 1.25 Å. Larger taper regions reintroduced discontinuities. The tapering mechanism was compared with the outer buffer scheme of Dixon and Merz Jr. [25] and was found to produce smoother results. An added and equally important benefit of the tapering mechanism is the acceleration of the SCF process for small to medium sized subsystems, when the taper mechanism produces smooth energy curves. The cause of the reintroduction of the discontinuities using larger taper regions needs further investigation, although work carried out in this thesis suggests that the issue is caused by the SCF process finding different ground state densities for different taper regions.

A new method is proposed to produce continuous energy landscapes within the D&C method. Here, a potential is added to each subsystem that is the sum of the background charge found surrounding the subsystem. The addition of the potential to each subsystem is thought to indirectly expand the subsystem boundary, hence reducing the boundary effects on the orbitals central to the subsystem. The potential will be able to be calculated using the particle mesh Ewald method [157] (or one of its many improvements [158]) or more appropriately the fast multipole method [101,159]. It is proposed that this method will smooth the energy landscape and also improve the precision of the D&C method at the expense of larger computational cost than the tapering mechanism. The proposed method will be investigated in the near future.

The second extension generalised the standard D&C FDM method [27,33] by allowing each subsystem to have its own density matrix update rate. In effect, the designation of more than just the active and frozen regions is possible. This accelerates the SCF process and decreases the number of relaxation steps (in a local geometry optimisation) by allowing more effective electron density flow be-

tween active and non-active regions. In fact, there can be many designated active regions with the proposed scheme. This was demonstrated with a local geometry optimisation of 10 atoms at one end of a 194-atom linear alkane molecule. The charge transfer between the active region and the frozen region was found to be smoother with the scheme proposed in this thesis. Applying the tapering mechanism produced an even smoother flow of charge between the regions. The proposed scheme reduced the number of relaxation steps when compared to the standard FDM scheme. The issue with large memory requirements is still present, limiting this scheme to moderately sized systems.

A proposed scheme to help reduce the memory cost involves storing the subsystem eigenvectors and eigenvalues onto disk and then accessing them when required. Using this scheme the number of subsystems loaded into memory can be set at runtime. The efficiency of reading the data from disk is hardware dependent resulting in the competitiveness of this method may be put into question. Of course using partitioning schemes that allow many subsystem core atoms will reduce the duplication of effort of storing the same eigenstates, which in turn reduces the memory cost.

The D&C paradigm was then applied to real-time TDDFT [34, 72] to create a linear-scaling method for excited states (DCTDDFT). The method produced accurate dipole moments for a linear alkane molecule when the external electric perturbation was aligned normal to the principal axis of the molecule. However, divergence in the dipole moment occurred for the case when the electric field was aligned parallel to the principal axis. The effect of the subsystem boundary on the propagating density was found to be cause of the divergence. Attempts to eliminate subsystem boundary effects using outer buffer regions, Hamiltonian tapering, selective local orbital propagation and wavefunction masking functions were found not be effective. The optical response was calculated, despite the issues with the dipole moment, and produced fairly reasonable results when compared to standard TDDFT; that is the major peaks of the spectrum were found. The success of the method for electric field polarisations normal to the alkane molecule suggests that, in the future, the method may be made to work for partition schemes that produce subsystems which are infinitely sized in the direction

parallel to the external field, but partitioned orthogonal to this.

Further investigations are required into improving the DCTDDFT method, as follows:

- The determination and handling of any time-dependent constraints within the method, as it is unclear if the partition function and the subsystem propagation should explicitly handle any time constraints.
- Application of a complex absorbing potential [209–212] to each subsystem as an alternative to the tapering and masking functions.
- Instead of exciting all frequencies at once with the external electric field, a monochromatic laser field can be used to excite a single frequency. This might make any of the previous measures (and proposed measures) to counter the subsystem boundary effects more effective. For example, with a monochromatic laser field perturbation, the wavefunction masking function and complex absorbing potential can be optimised to operate more effectively at the single excitation frequency.

Although not an improvement to the DCTDDFT, other methodologies to apply the D&C paradigm need to be investigated, such as the coupled-perturbed Kohn Sham scheme [246, 247], similar to the work of Touma *et al* [195] who applied D&C to the coupled-perturbed Hartree-Fock method [248, 249].

The final work performed in this thesis highlights both the capabilities and limitations of the present D&C implementation by studying three large systems. The first of these systems is a 715-atom dehydrated λ-DNA structure, which was previously studied by de Pablo *et al* [1]. The tapering mechanism is applied to this system and the convergence of the total energy with respect to subsystem size was examined. The D&C method was found to converge rapidly with increasing subsystem size and to be very efficient. The tapering mechanism was found to accelerate the SCF convergence process for subsystems less than 9.0 Å in radius.

A comparison with a second DNA system is made with the published results of Otsuka *et al* [35], who used the density matrix minimisation (DMM) [36] linear-scaling technique implemented in CONQUEST [37–39]. The D&C method was found to converge to Kohn-Sham energies at a faster rate than the DMM

method. A comparison of the electronic structure for hydrated and dehydrated DNA structures is also made. It was found that the hydrated DNA structure reduced the band gap of the dry system by 0.63 eV. The atomic partial charges for all atoms is found for both hydrated and dehydrated systems. A charge transfer of 3.2 e^- from the solvent to the DNA molecule is found to occur.

For the final system, the electronic structure of the large 13,584-atom ZIF-100 [40] structure is found using only comparatively moderate computing resources. A 7.337 Å subsystem radius was found to be the largest radius that could fit into the available memory per compute node. The electronic structure of a small fragment was found and compared to the ZIF-100 structure. It was found that the fragment possessed similar properties and can be considered representative of the ZIF-100 structure when considering local properties. Further studies are required to determine the bonding nature of electrophilic carbon dioxide and ZIF-100. Adsorption isotherms also need to be computed and compared with experimental results [40].

The D&C method was found to be very efficient and precise for the above mentioned systems. The one problem encountered for all systems was the limitations of the available memory on each compute node. High quality calculations with large subsystems require large amounts of memory on each compute node. To overcome this issue, a new memory model is proposed in which the Hamiltonian, overlap and eigenstate data for each subsystem is distributed over many compute nodes. By spreading the data across the many compute nodes increases the amount of memory available to each subsystem. Another important addition required for the current D&C implementation is the ability to have multiple core atoms within each subsystem. By allowing this, duplication of eigenstates within adjacent subsystems is reduced leading to large reductions in memory cost as well as computational cost. Currently, the D&C implementation can handle multiple core atom subsystems when running in serial; this needs to extended to the parallel case where the true benefits of the method are shown.

A

Publication Report

A.1 Peer reviewed journal article

- "First principles calculations using density matrix divide-and-conquer within the SIESTA methodology", B. O. Cankurtaran, J. D. Gale and M. J. Ford, *Journal of Physics - Condensed Matter* **20**, 294208 (2008).

A.2 Oral presentation

- "First principles electronic structure calculations of large nanostructures", B. O. Cankurtaran, J. D. Gale and M. J. Ford, Australian Institute of Physics Congress (AIP), December 2008, Adelaide, Australia.

A.3 Poster presentations

- "First principles calculations of very large nanostructures", B. O. Cankurtaran, J. D. Gale and M. J. Ford, The World Association of Theoretical and Computational Chemists (WATOC), September 2008, Sydney, Australia.

- "Divide-and-Conquer Density Functional Theory within SIESTA", B. O. Cankurtaran, J. D. Gale and M. J. Ford, International Conference on Nanoscience and Nanotechnology (ICONN), February 2008, Melbourne, Australia.

- "The Density Matrix Divide-and-Conquer Approach within the SIESTA Framework", B. O. Cankurtaran, J. D. Gale and M. J. Ford, Molecular Modelling 2007, November 2007, Melbourne, Australia.

Bibliography

[1] P. de Pablo *et al.*, Physical Review Letters **85**, 4992 (2000).

[2] J. Soler *et al.*, Journal of Physics - Condensed Matter **14**, 2745 (2002).

[3] J. Junquera, Ó. Paz, D. Sánchez-Portal, and E. Artacho, Physical Review B **64** (2001).

[4] E. Artacho, D. Sánchez-Portal, P. Ordejón, A. García, and J. Soler, Physica Status Solidi B-Basic Research **215** (1999).

[5] D. Sánchez-Portal, P. Ordejón, E. Artacho, and J. Soler, International Journal of Quantum Chemistry **65**, 453 (1997), 37th Annual Sanibel Symposium, St Augustine, FL, Mar 01-07, 1997.

[6] P. Ordejón, E. Artacho, and J. Soler, Physical Review B **53**, 10441 (1996).

[7] J. Gale and A. Rohl, Molecular Simulation **29**, 291 (2003).

[8] S. Fleming and A. Rohl, Zeitschrift Für Kristallographie **220**, 580 (2005).

[9] W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graphics **14**, 33 (1996).

[10] P. of Vision Pty. Ltd. (2004), [Computer software] - Retrieved from http://www.povray.org/download/ (2004).

[11] J. Stone, An efficient library for parallel ray tracing and animation, Master's thesis, Computer Science Department, University of Missouri-Rolla, 1998.

[12] S. R. Bahn and K. W. Jacobsen, Comput. Sci. Eng. **4**, 56 (2002).

[13] A. Szabo and N. Ostlund, *Modern Quantum Chemistry* (Dover Publishing, Mineola, New York, 1996).

[14] W. Kohn and L. Sham, Physical Review **137**, 1697 (1965).

[15] W. Kohn and L. Sham, Physical Review **140**, 1133 (1965).

[16] E. Landau, *Handbuch der Lehre von der Verteilung der Primzahlen*2 vols (Leipzig, 1909).

[17] H. Sambe and R. Felton, Journal of Chemical Physics **62**, 1122 (1975).

[18] W. Kohn, Physical Review Letters **76**, 3168 (1996).

[19] S. Goedecker, Reviews of Modern Physics **71**, 1085 (1999).

[20] W. Yang and T. Lee, Journal of Chemical Physics **103**, 5674 (1995).

[21] W. Yang, Physical Review Letters **66**, 1438 (1991).

[22] W. Yang, Physical Review A **44**, 7823 (1991).

[23] F. Shimojo, R. Kalia, A. Nakano, and P. Vashishta, Computer Physics Communications **167**, 151 (2005).

[24] S. Dixon and K. Merz Jr., Journal of Chemical Physics **104**, 6643 (1996).

[25] S. Dixon and K. Merz Jr., Journal of Chemical Physics **107**, 879 (1997).

[26] T. Lee, J. Lewis, and W. Yang, Computational Materials Science **12**, 259 (1998).

[27] T. Lee and W. Yang, International Journal of Quantum Chemistry **69**, 397 (1998).

[28] T. Akama, M. Kobayashi, and H. Nakai, Journal of Computational Chemistry **28**, 2003 (2007).

[29] M. Kobayashi, T. Akama, and H. Nakai, Journal of Chemical Physics **125** (2006).

[30] C. F. Bunge and R. Carbo-Dorca, Journal of Chemical Physics **125** (2006).

[31] W. Li and S. Li, Journal of Chemical Physics **121**, 6649 (2004).

[32] J. Kim, F. Mauri, and G. Galli, Physical Review B **52**, 1640 (1995).

[33] M. Ermolaeva, A. van der Vaart, and K. Merz, Journal of Physical Chemistry A **103**, 1868 (1999).

[34] K. Yabana and G. F. Bertsch, Physical Review B **54**, 4484 (1996).

[35] T. Otsuka, T. Miyazaki, T. Ohno, D. Bowler, and M. Gillan, Journal of Physics - Condensed Matter **20**, (2008).

[36] E. Hernandez, M. Gillan, and C. Goringe, Physical Review B **53**, 7147 (1996).

[37] D. Bowler, R. Choudhury, M. Gillan, and T. Miyazaki, Physica Status Solidi B - Basic Solid State Physics **243**, 989 (2006).

[38] D. Bowler, T. Miyazaki, and M. Gillan, Journal of Physics - Condensed Matter **14**, 2781 (2002).

[39] T. Miyazaki, D. Bowler, R. Choudhury, and M. Gillan, Journal of Chemical Physics **121**, 6186 (2004).

[40] B. Wang, A. Cote, H. Furukawa, M. O'Keeffe, and O. Yaghi, Nature **453**, 207 (2008).

[41] R. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, The International Series of Monographs on Chemistry (Oxford University Press, 1994).

[42] W. Koch and M. Holthausen, *A Chemist's Guide to Density Functional Theory*, Second ed. (Wiley-VCH, 2001).

[43] M. Marques *et al.*, editors, *Time-Dependent Density Functional Theory*, The Lecture Notes in Physics (Springer, 2006).

[44] M. Born and R. Oppenheimer, Annalen der Physik **389**, 457 (1927).

[45] J. C. Tully, Theoretical Chemistry Accounts **103**, 173 (2000).

[46] L. Thomas, Proceedings of the Cambridge Philosophical Society **23**, 542548 (1927).

[47] E. Fermi, Rend. Accad. Naz. Lincei **6**, 602607 (1927).

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

[49] P. Dirac, Proceedings of the Cambridge Philosphical Society **26**, 376 (1930).

[50] D. M. Ceperley and B. J. Alder, Physical Review Letters **45**, 566 (1980).

[51] J. P. Perdew *et al.*, Physical Review B **46**, 6671 (1992).

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

[53] S. J. Vosko, L. Wilk, and M. Nusair, Canadian Journal of Physics **58**, 1200 (1980).

[54] J. Perdew, K. Burke, and M. Ernzerhof, Physical Review Letters **77**, 3865 (1996).

[55] S. Ghosh and R. Parr, Physical Review A **34**, 785 (1986).

[56] A. Becke and M. Roussel, Physical Review A **39**, 3761 (1989).

[57] S. Kurth, J. Perdew, and P. Blaha, International Journal of Quantum Chemistry **75**, 889 (1999).

[58] J. Perdew, S. Kurth, A. Zupan, and P. Blaha, Physical Review Letters **82**, 2544 (1999).

[59] J. Perdew, S. Kurth, A. Zupan, and P. Blaha, Physical Review Letters **82**, 5179 (1999).

[60] J. Tao, J. Perdew, V. Staroverov, and G. Scuseria, Physical Review Letters **91**, (2003).

[61] A. Becke, Journal of Chemical Physics **98**, 1372 (1993).

[62] C. Lee, W. Yang, and R. Parr, Physical Review B **37**, 785 (1988).

[63] S. Vosko, L. Wilk, and M. Nusair, Canadian Journal of Physics **58**, 1200 (1980).

[64] P. Stephens, F. Devlin, C. Chabalowski, and M. Frisch, Journal of Physical Chemistry **98**, 11623 (1994).

[65] A. Gorling, The Journal of Chemical Physics **123**, 062203 (2005).

[66] J. Talman and W. Shadwick, Physical Review A **14**, 36 (1976).

[67] J. Krieger, L. Yan, and G. Iafrate, Physical Review A **45**, 101 (1992).

[68] E. K. U. Gross, J. F. Dobson, and M. Petersilka, *Topics in Current Chemistry - vol 181* (Springer-Verlag, 1996).

[69] R. Van Leeuwen, International Journal of Modern Physics B **15**, 1969 (2001).

[70] K. Burke, J. Werschnik, and E. Gross, Journal of Chemical Physics **123**, (2005).

[71] P. Elliott, K. Burke, and F. Furche, arXiv:cond-mat/0703590v1 (2007).

[72] E. Runge and E. K. U. Gross, Physical Review Letters **52**, 997 (1984).

[73] M. Payne, M. Teter, D. Allan, T. Arias, and J. Joannopoulos, Reviews of Modern Physics **64**, 1045 (1992).

[74] J. Andzelm and E. Wimmer, Journal of Chemical Physics **96**, 1280 (1992).

[75] K. Cho, T. Arias, J. Joannopoulos, and P. Lam, Physical Review Letters **71**, 1808 (1993).

[76] J. Chelikowsky, N. Troullier, K. Wu, and Y. Saad, Physical Review B **50**, 11355 (1994).

[77] E. Hernandez, M. Gillan, and C. Goringe, Physical Review B **55**, 13485 (1997).

[78] C. Skylaris, P. Haynes, A. Mostofi, and M. Payne, Journal of Chemical Physics **122**, (2005).

[79] B. Delley, Journal of Chemical Physics **92**, 508 (1990).

[80] T. L. Beck, Reviews of Modern Physics **72**, 1041 (2000).

[81] F. Jensen, *Introduction to Computational Chemistry*, 2nd ed. (Wiley, 2006).

[82] S. F. Boys and F. Bernardi, Molecular Physics **19**, 553 (1970).

[83] S. Huzinaga and M. Klobukowski, Chemical Physics Letters **120**, 509 (1985).

[84] D. Sánchez-Portal, E. Artacho, and J. M. Soler, Journal of Physics: Condensed Matter **8**, 3859 (1996).

[85] G. Wannier, Physical Review **52**, 0191 (1937).

[86] W. Kohn, Physical Review **115**, 809 (1959).

[87] J. Cloizeaux, Physical Review **129**, 554 (1963).

[88] D. R. Hamann, M. Schlüter, and C. Chiang, Physical Review Letters **43**, 1494 (1979).

[89] L. Kleinman, Physical Review B **21**, 2630 (1980).

[90] G. B. Bachelet and M. Schlüter, Physical Review B **25**, 2103 (1982).

[91] N. Troullier and J. L. Martins, Physical Review B **43**, 1993 (1991).

[92] L. Kleinman and D. M. Bylander, Physical Review Letters **48**, 1425 (1982).

[93] P. E. Blöchl, Physical Review B **41**, 5414 (1990).

[94] O. F. Sankey and D. J. Niklewski, Physical Review B **40**, 3979 (1989).

[95] G. Lippert, J. Hutter, P. Ballone, and M. Parrinello, Journal of Physical Chemistry **100**, 6231 (1996), International Conference on Molecular Quantum Mechanics - Methods and Applications, in Memory of Samuel Francis Boys and in Honor of Isaiah Shavitt, Cambridge, England, Sep 03-07, 1995.

[96] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, 1992).

[97] P. Ordejón, D. Drabold, M. Grumbach, and R. Martin, Physical Review B **48**, 14646 (1993).

[98] P. Ordejón, D. A. Drabold, R. M. Martin, and M. P. Grumbach, Physical Review B **51**, 1456 (1995).

[99] F. Mauri and G. Galli, Physical Review B **50**, 4316 (1994).

[100] F. Mauri, G. Galli, and R. Car, Physical Review B **47**, 9973 (1993).

[101] L. Greengard and V. Rokhlin, Journal of Computational Physics **73**, 325 (1987).

[102] M. Merrick, K. Iyer, and T. Beck, Journal of Physical Chemistry **99**, 12478 (1995).

[103] S. Kenny, A. Horsfield, and H. Fujitani, Physical Review B **62**, 4899 (2000).

[104] M. Challacombe, Journal of Chemical Physics **110**, 2332 (1999).

[105] X. Li, R. Nunes, and D. Vanderbilt, Physical Review B **47**, 10891 (1993).

[106] A. Palser and D. Manolopoulos, Physical Review B **58**, 12704 (1998).

[107] A. Niklasson, Physical Review B **66**, (2002).

[108] T. Ozaki, Physical Review B **67**, (2003).

[109] T. Ozaki, Physical Review B **74**, (2006).

[110] T. Zhu, W. Pan, and W. Yang, Physical Review B **53**, 12713 (1996).

[111] T. Zhu, W. Pan, and W. Yang, Theor. Chem. Acc **96**, 2 (1997).

[112] O. Warschkow, J. Dyke, and D. Ellis, Journal of Computational Physics **143**, 70 (1998).

[113] W. Pan, T. Lee, and W. Yang, Journal of Computational Chemistry **19**, 1101 (1998).

[114] D. Shaw and A. St-Amant, Journal of Theoretical & Computational Chemistry **3**, 419 (2004).

[115] P. Vashishta, R. Kalia, and A. Nakano, Journal of Physical Chemistry B **110**, 3727 (2006).

[116] A. Nakano *et al.*, Computational Materials Science **38**, 642 (2007).

[117] P. Vashishta, R. K. Kalia, A. Nakano, B. E. Homan, and K. L. McNesby, Journal of Propulsion and Power **23**, 688 (2007).

[118] F. Shimojo, R. K. Kalia, A. Nakano, and P. Vashishta, Physical Review B **77** (2008).

[119] F. Shimojo, A. Nakano, R. K. Kalia, and P. Vashishta, Applied Physics Letters **95** (2009).

[120] Z. Zhao, J. Meza, and L.-W. Wang, Journal of Physics - Condensed Matter **20** (2008), CECAM Workshop on Linear-Scaling Ab Initio Calculation - Applications and Future Directions, Lyon, France, Sep 03-06, 2007.

[121] X. He and K. Merz Jr., Journal of Chemical Theory and Computation **6**, 405 (2010).

[122] R. Alizadegan, K. J. Hsia, and T. J. Martinez, Journal of Chemical Physics **132** (2010).

[123] M. Kobayashi and H. Nakai, International Journal of Quantum Chemistry **109**, 2227 (2009).

[124] M. Kobayashi and H. Nakai, Journal of Chemical Physics **129** (2008).

[125] T. Akama, A. Fujii, M. Kobayashi, and H. Nakai, Molecular Physics **105**, 2799 (2007).

[126] M. Sironi, A. Genoni, M. Civera, S. Pieraccini, and M. Ghitti, Theoretical Chemistry Accounts **117**, 685 (2007).

[127] W. Kohn, Chemical Physics Letters **208**, 167 (1993).

[128] M. Geller and W. Kohn, Physical Review B **48**, 14085 (1993).

[129] R. Mulliken, Journal of Chemical Physics **36**, 3428 (1962).

[130] C. Lee and W. Yang, Journal of Chemical Physics **96**, 2408 (1992).

[131] L. S. Blackford *et al.*, *ScaLAPACK Users' Guide* (Society for Industrial and Applied Mathematics, 1997), ISBN 0-89871-397-8.

[132] C.-K. Yang, J. Zhao, and J. P. Lu, Phys. Rev. B **66**, 041403 (2002).

[133] D. Meagher, Computer Graphics and Image Processing **19**, 129 (1982).

[134] D. Meagher, Computer Graphics World **7**, 75 (1984).

[135] J. Bentley, Communications of the ACM **18**, 509 (1975).

[136] J. Bentley and J. Friedman, Computing Surveys **11**, 397 (1979).

[137] J. Bentley, IEEE Transactions on Software Engineering **5**, 333 (1979).

[138] L. Trefethen and D. Bau III, *Numerical Linear Algebra* (SIAM: Society for Industrial and Applied Mathematics, 1997).

[139] J. Boughton and P. Pulay, Journal of Computational Chemistry **14**, 736 (1993).

[140] J. Pipek and P. Mezey, Journal of Chemical Physics **90**, 4916 (1989).

[141] D. Caldwell, P. Redington, and H. Eyring, Proceedings of the National Academy of Sciences of the United States of America **76**, 3042 (1979).

[142] S. F. Boys, *Quantum Theory of Atoms, Molecules and the Solid State* (New York Academic, 1966).

[143] C. Edmiston and K. Ruedenberg, *Quantum Theory of Atoms, Molecules and the Solid State* (New York Academic, 1966).

[144] T. Andrea, W. Swope, and H. Andersen, Journal of Chemical Physics **79**, 4576 (1983).

[145] C. Brooks, B. Pettitt, and M. Karplus, Journal of Chemical Physics **83**, 5897 (1985).

[146] C. Brooks, Journal of Chemical Physics **86**, 5156 (1987).

[147] R. Loncharich and B. Brooks, Proteins-Structure Function and Genetics **6**, 32 (1989).

[148] E. Burchart, V. Verheij, H. Vanbekkum, and B. Vandegraaf, Zeolites **12**, 183 (1992).

[149] K. Tasaki, S. McDonald, and J. Brady, Journal of Computational Chemistry **14**, 278 (1993).

[150] K. Lau, H. Alper, T. Thacher, and T. Stouch, Journal of Physical Chemistry **98**, 8785 (1994).

[151] A. van Duin *et al.*, Journal of Physical Chemistry A **107**, 3803 (2003).

[152] A. Klamt and G. Schuurmann, Journal of the Chemical Society-Perkin Transactions 2 , 799 (1993).

[153] H. Senn, P. Margl, R. Schmid, T. Ziegler, and P. Blochl, Journal of Chemical Physics **118**, 1089 (2003).

[154] B. Delley, Molecular Simulation **32**, 117 (2006).

[155] J. Gale and A. Rohl, Molecular Simulation **33**, 1237 (2007).

[156] T. Akama, M. Kobayashi, and H. Nakai, International Journal of Quantum Chemistry **109**, 2706 (2009).

[157] U. Essmann *et al.*, Journal of Chemical Physics **103**, 8577 (1995).

[158] D. Cerutti, R. Duke, T. Darden, and T. Lybrand, Journal of Chemical Theory and Computation **5**, 2322 (2009).

[159] L. Greengard, Science **265**, 909 (1994).

[160] A. Warshel and M. Levitt, Journal of Molecular Biology **103**, 227 (1976).

[161] P. Kollman and D. Hayes, Journal of the American Chemical Society **103**, 2955 (1981).

[162] O. Tapia, J. Andres, J. Aullo, and C. Branden, Journal of Chemical Physics **83**, 4673 (1985).

[163] M. Field, P. Bash, and M. Karplus, Journal of Computational Chemistry **11**, 700 (1990).

[164] P. Bash *et al.*, BioChemistry **30**, 5826 (1991).

[165] V. Luzhkov and A. Warshel, Journal of Computational Chemistry **13**, 199 (1992).

[166] V. Thery, D. Rinaldi, J. Rivail, B. Maigret, and G. Ferenczy, Journal of Computational Chemistry **15**, 269 (1994).

[167] R. Stanton, D. Hartsough, and K. Merz, Journal of Computational Chemistry **16**, 113 (1995).

[168] S. Humbel, S. Sieber, and K. Morokuma, Journal of Chemical Physics **105**, 1959 (1996).

[169] J. Gao, *Methods and Applications of Combined Quantum Mechanical and Molecular Mechanical Potentials*, Reviews in Computational Chemistry Vol. Volume 7 (Wiley-VCH, 2007).

[170] U. Singh and P. Kollman, Journal of Computational Chemistry **7**, 718 (1986).

[171] J. Gao, Journal of Physical Chemistry B **101**, 657 (1997).

[172] P. Cummins and J. Gready, Journal of Computational Chemistry **18**, 1496 (1997).

[173] P. Cortona, Physical Review B **44**, 8454 (1991).

[174] T. Wesolowski and A. Warshel, Journal of Physical Chemistry **97**, 8050 (1993).

[175] T. Wesolowski and A. Warshel, Journal of Physical Chemistry **98**, 5183 (1994).

[176] A. Mazur, Journal of Physical Chemistry B **102**, 473 (1998).

[177] R. Mulliken, Journal of Chemical Physics **23**, 1833 (1955).

[178] G. Chen and S. Mukamel, Journal of Physical Chemistry **100**, 11080 (1996).

[179] S. Yokojima and G. Chen, Chemical Physics Letters **292**, 379 (1998).

[180] S. Yokojima and G. Chen, Physical Review B **59**, 7259 (1999).

[181] C. Yam, S. Yokojima, and G. Chen, Physical Review B **68**, (2003).

[182] C. Yam, S. Yokojima, and G. Chen, Journal of Chemical Physics **119**, 8794 (2003).

[183] W. Liang, S. Yokojima, and G. Chen, Journal of Chemical Physics **110**, 1844 (1999).

[184] W. Liang, S. Yokojima, and G. Chen, Journal of Chemical Physics **113**, 1403 (2000).

[185] A. Niklasson and M. Challacombe, Physical Review Letters **92**, (2004).

[186] A. Niklasson, V. Weber, and M. Challacombe, Journal of Chemical Physics **123**, (2005).

[187] V. Weber, A. Niklasson, and M. Challacombe, Physical Review Letters **92**, (2004).

[188] J. Kussmann and C. Ochsenfeld, Journal of Chemical Physics **127**, (2007).

[189] Y. Mochizuki *et al.*, Chemical Physics Letters **418**, 418 (2006).

[190] T. Ishikawa *et al.*, Chemical Physics Letters **430**, 361 (2006).

[191] S. Coriani *et al.*, Journal of Chemical Physics **126**, (2007).

[192] B. Walker, A. M. Saitta, R. Gebauer, and S. Baroni, Physical Review Letters **96**, 113001 (2006).

[193] C. M.E., *Recent Advances in Density Functional Methods Part I* (World Scientific, Singapore, 1995).

[194] M. Casida, Journal of Chemical Physics **122**, (2005).

[195] T. Touma, M. Kobayashi, and H. Nakai, Chemical Physics Letters **485**, 247 (2010).

[196] T. Ozaki, Physical Review B **64**, art. no. (2001).

[197] B. Jansik, S. Host, P. Jorgensen, J. Olsen, and T. Helgaker, Journal of Chemical Physics **126**, (2007).

[198] S. Mukamel, *Principles of Nonlinear Optics and Spectroscopy* (Oxford University Press, 1995).

[199] A. Castro and M. Marques, *Time-Dependent Density Functional Theory*, The Lecture Notes in Physics, Vol. 706 (Springer, 2006), Propagators for the Time-Dependent Kohn-Sham Equations.

[200] A. Castro, M. Marques, and A. Rubio, Journal of Chemical Physics **121**, 3425 (2004).

[201] A. Tsolakidis, D. Sánchez-Portal, and R. Martin, Physical Review B **66**, (2002).

[202] J. Crank and P. Nicolson, Proceedings of the Cambridge Philosophical Society **43**, 50 (1947).

[203] J. Crank and P. Nicolson, Advances in Computational Mathematics **6**, 207 (1996).

[204] D. Sánchez-Portal, personal communication, 2009.

[205] R. Kosloff and C. Cerjan, Journal of Chemical Physics **81**, 3722 (1984).

[206] C. Marston, G. Balintkurti, and R. Dixon, Theoretica Chimica Acta **79**, 313 (1991).

[207] M. Monnerville, P. Halvick, and J. Rayez, Chemical Physics **159**, 227 (1992).

[208] T. Burnus, M. Marques, and E. Gross, Physical Review A **71**, (2005).

[209] A. Vibok and G. Balintkurti, Journal of Physical Chemistry **96**, 8712 (1992).

[210] A. Vibok and G. Balintkurti, Journal of Chemical Physics **96**, 7615 (1992).

[211] A. N. Hussain and G. Roberts, Phys. Rev. A **63**, 012703 (2000).

[212] A. Vibok and G. Halasz, Physical Chemistry Chemical Physics **3**, 3048 (2001).

[213] J. Harris, Phys. Rev. B **31**, 1770 (1985).

[214] W. M. C. Foulkes and R. Haydock, Phys. Rev. B **39**, 12520 (1989).

[215] A. Torralba *et al.*, Journal of Physics - Condensed Matter **20**, (2008).

[216] D. Case *et al.*, University of California, San Francisco (2006).

[217] I. Kratochvilova *et al.*, Journal of Physical Chemistry B **114**, 5196 (2010).

[218] R. Dixon and P. Kollman, Journal of Computational Chemistry **18**, 1632 (1997).

[219] M. Aida, G. Corongiu, and E. Clementi, International Journal of Quantum Chemistry **42**, 1353 (1992).

[220] R. Santamaria, G. Cocho, L. Corona, and E. Gonzlez, Chemical Physics **227**, 317 (1998).

[221] A. Bende, F. Bogr, and J. Ladik, Chemical Physics Letters **463**, 211 (2008).

[222] K. Park *et al.*, Proceedings of the National Academy of Sciences of the United States of America **103**, 10186 (2006).

[223] H. Hayashi, A. Cote, H. Furukawa, M. O'Keeffe, and O. Yaghi, Nature Materials **6**, 501 (2007).

[224] R. Banerjee *et al.*, Science **319**, 939 (2008).

[225] O. Yaghi *et al.*, Nature **423**, 705 (2003).

[226] H. Chae *et al.*, Nature **427**, 523 (2004).

[227] Q. Li *et al.*, Science **325**, 855 (2009).

[228] D. Britt, H. Furukawa, B. Wang, T. Glover, and O. Yaghi, Proceedings of the National Academy of Sciences of the United States of America **106**, 20637 (2009).

[229] S. Keskin, J. Liu, R. Rankin, J. Johnson, and D. Sholl, Industrial & Engineering Chemistry Reseacrh **48**, 2355 (2009).

[230] J. Liu, R. Rankin, and J. Johnson, Molecular Simulation **35**, 60 (2009).

[231] K. Walton *et al.*, Journal of the American Chemical Society **130**, 406 (2008).

[232] T. Yildirim and M. Hartman, Physical Review Letters **95**, (2005).

[233] H. Wu, W. Zhou, and T. Yildirim, Journal of the American Chemical Society **129**, 5314 (2007).

[234] Y. Liu, H. Liu, Y. Hu, and J. Jiang, Journal of Physical Chemistry B **113**, 12326 (2009).

[235] Q. Yang and C. Zhong, Journal of Physical Chemistry B **109**, 11862 (2005).

[236] Q. Yang and C. Zhong, Journal of Physical Chemistry B **110**, 655 (2006).

[237] G. Garberoglio, A. Skoulidas, and J. Johnson, Journal of Physical Chemistry B **109**, 13094 (2005).

[238] M. Zhou, Q. Wang, L. Zhang, Y. Liu, and Y. Kang, Journal of Physical Chemistry B **113**, 11049 (2009).

[239] I. Baburin, S. Leoni, and G. Seifert, Journal of Physical Chemistry B **112**, 9437 (2008).

[240] H. Frost and R. Snurr, Journal of Physical Chemistry C **111**, 18794 (2007).

[241] D. Liu, C. Zheng, Q. Yang, and C. Zhong, Journal of Physical Chemistry C **113**, 5004 (2009).

[242] R. Rankin, J. Liu, A. Kulkarni, and J. Johnson, Journal of Physical Chemistry C **113**, 16906 (2009).

[243] Q. Yang and C. Zhong, Journal of Physical Chemistry B **110**, 17776 (2006).

[244] A. Rappe, C. Casewit, K. Colwell, W. Goddard, and W. Skiff, Journal of the American Chemical Society **114**, 10024 (1992).

[245] D. R. Bowler and T. Miyazaki, Journal of Physics-Condensed Matter **22** (2010).

[246] S. van Gisbergen, J. Snijders, and E. Baerends, Journal of Chemical Physics **109**, 10644 (1998).

[247] M. Ferrero, M. Rerat, R. Orlando, R. Dovesi, and I. J. Bush, Coupled perturbed kohn-sham calculation of static polarizabilities of periodic compounds - art. no. 012016, in *Ab Initio Simulation of Crystalline Solids: History and Prospects - Contributions in Honor of Cesare Pisani*, edited by R. Dovesi, R. Orlando, and C. Roetti, , Journal of Physics Conference Series Vol. 117, p. 12016, 2008, Conference on Ab Initio Simulation of Crystalline Solids, Torino, ITALY, Sep 08-09, 2008.

[248] H. Sekino and R. Bartlett, Journal of Chemical Physics **85**, 976 (1986).

[249] S. Karna and M. Dupuis, Journal of Computational Chemistry **12**, 487 (1991).

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.