Techniques and Applications of Quantum Monte Carlo

A dissertation submitted for the degree of

Doctor of Philosophy

at the University of Cambridge

Paul Richard Charles Kent

Robinson College, Cambridge

August 1999

Preface

This dissertation describes work undertaken between October 1996 and August 1999 in the Theory of Condensed Matter group at the Cavendish Laboratory, Cambridge, under the supervision of Dr. Richard Needs.

Except where reference is made to the work of others, this dissertation is the result of my own work and contains nothing which is the outcome of work done in collaboration. This thesis has not been submitted in whole or in part for any degree or diploma at this or any other university.

Paul Richard Charles Kent
August 1999

Acknowledgements

The research in this thesis would have taken far longer to complete without the encouragement from many others. It is a delight to acknowledge those who have supported me over the last three years.

I would like to thank my supervisor, Dr. Richard Needs, for his guidance and relaxed, thoughtful insight. I thank Dr. Guna Rajagopal for his hospitality and for his remarkable demonstrations of the process known as "negotiation".

This work was supported by an EPSRC studentship and substantial allocations of supercomputer time, without either of which this work would not have been possible.

I am particularly thankful for the help and advice of Dr. Randolph Hood and Dr. Andrew Williamson during my first few months at the Cavendish, without whom the learning curve would have been very much steeper.

The days would have passed far more slowly without the support of my friends, both at the Cavendish and around Cambridge, whom I thank for putting up with my idiosyncrasies and for providing such a rich source of conversation, education and entertainment.

I am indebted to those at the Cavendish Laboratory who welcomed me at such short notice, and to the many members of staff at Bath University who encouraged me on this journey.

Finally, I wish to thank my parents for their love and encouragement, without whom I would never have enjoyed so many opportunities.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Quantum Mechanics

The formulation and development of quantum theory in the first half of the 20^{th} century has led to a revolution in our understanding of fundamental physics. Quantum theory has demonstrated a surprising accuracy and predictive power, and the importance of quantum theory in the pure and applied sciences in virtually unchallenged.

The relevant equations to be solved are clearly the equations of quantum mechanics , if we are to attempt to model real processes and real materials. Unfortunately, Schrödinger's equation, the most fundamental equation of wavefunction based quantum mechanics, cannot be solved analytically for all but the most trivial of systems, most of which are not relevant to the world at large.

In order to make useful progress the equations must be solved numerically.

1.2 Electronic structure calculations

Electronic structure calculations - numerical solutions of Schrödinger's equation for a specific system - are distinct from other forms of modelling because they are first-principles in nature. That is, except through the choice of the researcher, the calculations contain no external parameters other than a most basic description of the system. Calculations of this nature enable the speculative study of systems, potentially without reference to experiment. Where a given physical property is physically inaccessible - such as the binding energy of an atom or molecule deeply embedded in a complex host - the availability of reliable qualitative data is a powerful stimulus.

However, the numerical solution of Schrödinger's equation remains a difficult task. Exact solutions of the equation are, in general, only solvable in times scaling exponentially with system size. This scaling precludes exact calculations for all but the smallest and simplest of systems. Approximations may be introduced to reduce the equations to a form that can be solved in polynomial time, but at the penalty of losing some degree of accuracy and predictive power. The treatment of electron-electron interactions is the principle source of difficulty: the physical and chemical properties of a system depend principally on the interaction of the electrons with each other and with the atomic cores. These interactions cannot easily be separated out or treated without approximation.

The most successful electronic structure methods in current use, those of density functional theory and quantum chemistry, have been applied to a wide range of systems relevant to the real world. In practice, the density functional and quantum chemical approaches involve approximations for the electron-electron interactions, limiting the achievable accuracy.

In this thesis, stochastic methods for the solution of Schrödinger's equation are developed and applied to real systems. Quantum Monte Carlo (QMC) methods treat electron-electron interactions almost without approximation and with a computational cost scaling cubically with system size. Thieir accuracy enables an unprecedented degree of confidence to be placed in the results obtained. A number of technical developments are made in the earlier chapters of the thesis, prior to two significant applications: the computation of the one-body density matrix of silicon and several related quantities, and a study of the energetic stability of a series of carbon clusters. The use of Quantum Monte Carlo methods was essential in both of these applications, particularly in the study of carbon clusters where no other currently applicable computational method is sufficiently accurate.

1.3 Outline of thesis

The thesis is broadly arranged into three sections:

Electronic structure methods are introduced and details of Quantum Monte Carlo methods are given in chapters 2 - 4. In chapter 2, density functional and several quantum chemical methods are reviewed and their advantages and disadvantages presented. Chapter 3 introduces the two most important QMC methods: variational and fixed-node diffusion Monte Carlo. Several key results are derived. The computational realisation of the methods is described in chapter 4.

In chapters 5 and 6, two QMC techniques are developed and their utility verified. In chapter 5, many techniques for optimising many-body wavefunctions are tested and an efficient optimisation procedure proposed. The treatment of finite size effects, due to the supercell approximation used for periodic systems, is examined in chapter 6. A modified form of Coulomb interaction is shown to significantly reduce finite size effects.

Finally, two specific applications of QMC are presented: In chapter 7, the one-body density matrix of the valence electrons of bulk silicon is computed using a correlated wavefunction. The band structure of silicon is also determined using an extension of Koopmans' theorem. In chapter 8, the energetic (zero temperature) stability of small carbon clusters, including fullerenes, are examined. QMC calculations are used to identify the smallest energetically stable fullerene and analyse the performance of currently popular density functionals.

In chapter 9, the results obtained are summarised and an outline given for future work in the field.

Chapter 2

Electronic structure methods

2.1 Introduction

Quantum Monte Carlo techniques provide a direct and potentially efficient means for solving the many-body Schrödinger equation of quantum mechanics. Finding accurate solutions of this equation is one of the central problems in physics and chemistry.

The simplest Quantum Monte Carlo technique, variational Monte Carlo (VMC), is based on a direct application of Monte Carlo integration to the calculation of multidimensional integrals of expectation values such as the total energy. Monte Carlo methods are statistical and a key result is that the value of integrals computed using Monte Carlo converges faster than by using conventional methods of numerical quadrature, once the problem involves more than a few dimensions. Statistical methods therefore provide a practical means of solving the full many-body Schrödinger equation by direct integration, making only limited and well-controlled approximations.

Comparisons are made with several well-developed approaches for solving the many-body Schrödinger equation in this chapter. Numerous quantum chemical methods and density functional approaches have been developed for solving the "electronic structure problem". A review of several methods currently in general use is given along with an outline of their relative advantages and shortcomings. Finally, the advantages of Quantum Monte Carlo techniques are outlined prior to their full description in chapter 3.

2.2 The Hamiltonian

The time independent Schrödinger equation for a system of N particles interacting via the Coulomb interaction is

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

where

$$\hat{H} = \sum_{i=1}^{N} \left(-\frac{\hbar^2}{2m_i} \nabla_i^2 \right) + \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{Z_i Z_j}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} \tag{2.2}$$

and Ψ is an N-body wavefunction, $\mathbf{r}$ denotes spatial positions and Z the charges of the individual particles. E denotes the energy of either the ground or an excited state of the system.

Most physical problems of interest consist of a number of interacting electrons and ions. The total number of particles, N, is usually sufficiently large that an exact solution cannot be found. Controlled and well understood approximations are sought to reduce the complexity to a tractable level. Once the equations are solved, a large number of properties may be calculated from the wavefunction. Errors or approximations made in obtaining the wavefunction will be manifest in any property derived from the wavefunction. Where high accuracy is required, considerable attention must be paid to the derivation of the wavefunction and any approximations made.

2.2.1 The Born-Oppenheimer approximation

A common and very reasonable approximation used in the solution of equation 2.1 is the Born-Oppenheimer Approximation. [1] In a system of interacting electrons and nuclei there will usually be little momentum transfer between the two types of particles due to their greatly differing masses. The forces between the particles are of similar magnitude due to their similar charge. If one then assumes that the momenta of the particles are also similar, then the nuclei must have much smaller velocities than the electrons due to their far greater mass. On the time-scale of nuclear motion, one can therefore consider the electrons to relax to a ground-state given by the Hamiltonian equation 2.2 with the nuclei at fixed locations. This separation of the electronic and nuclear degrees of freedom is known as the Born-Oppenheimer approximation.

This approximation will be used for the remainder of this thesis. It is important to note that this approximation does not limit the techniques described to systems of fixed ions: in principle, once the electronic configuration is known, the nuclear degrees of freedom could also be solved for, giving rise to nuclear motion.

In practice Newtonian mechanics using forces calculated via quantum mechanics is often sufficient to solve for the motion of the nuclei, however, these aspects go beyond the scope of the thesis so that from now on a simpler version of the many-body Hamiltonian, equation 2.2, is used

$$\hat{H} = \sum_i -\frac{1}{2} \nabla_i^2 + \sum_i \sum_\alpha \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{d}_\alpha|}$$

$$+ \quad \frac{1}{2}\sum_{i}\sum_{j\neq i}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \frac{1}{2}\sum_{\alpha}\sum_{\beta\neq\alpha}\frac{Z_\alpha Z_\beta}{|\mathbf{d}_\alpha - \mathbf{d}_\beta|} \quad . \tag{2.3}$$

The opportunity has been taken to separate the interacting particles into electrons and ions. The terms in the Hamiltonian are now expressed in terms of N electrons of charge -1 at positions $\mathbf{r}_i$ and ions of charge Z_α at positions $\mathbf{d}_\alpha$. This simplified electronic Hamiltonian remains very difficult to solve. No analytic solutions exist for general systems with more than one electron.

Note that this equation has been written in atomic units ($e = m_e = \hbar = 4\pi\epsilon_0 = 1$) which are more convenient for quantum mechanical problems and will be used for the remainder of the equations in this thesis.

2.3 Hartree Fock theory

Hartree Fock theory is one the simplest approximate theories for solving the many–body Hamiltonian. It is based on a simple approximation to the true many–body wavefunction: that the wavefunction is given by a single Slater determinant of N spin-orbitals

$$\Psi = \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_1(\mathbf{x}_2) & \cdots & \psi_1(\mathbf{x}_N) \\ \psi_2(\mathbf{x}_1) & \psi_2(\mathbf{x}_2) & \cdots & \psi_2(\mathbf{x}_N) \\ \vdots & \vdots & & \\ \psi_N(\mathbf{x}_1) & \psi_N(\mathbf{x}_2) & \cdots & \psi_N(\mathbf{x}_N) \end{vmatrix} \quad , \tag{2.4}$$

where the variables $\mathbf{x}$ include the coordinates of space and spin. This simple ansatz for the wavefunction Ψ captures much of the physics required for accurate solutions of the Hamiltonian. Most importantly, the wavefunction is antisymmetric with respect to an interchange of any two electron positions. This property is required by the Pauli exclusion principle, i.e.

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_N) = -\Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N) \quad . \tag{2.5}$$

This wavefunction may be inserted into the Hamiltonian, equation 2.3, and an expression for the total energy derived. [2–4] Applying the theorem that the value of a determinant is unchanged by any non-singular linear transformation, we may choose the ψ to be an orthonormal set. We now introduce a Lagrange multiplier ϵ_i to impose the condition that the ψ are normalised, and minimise with respect to the ψ

$$\frac{\delta}{\delta\psi}\left[< \hat{H} > - \sum_j \epsilon_j \int |\psi_j|^2 d\mathbf{r}\right] = 0 \quad . \tag{2.6}$$

An enormous simplification of the expressions for the orbitals ψ results. They reduce to a set of one-electron equations of the form

$$-\frac{1}{2}\nabla^2\psi_i(\mathbf{r}) + V_{ion}(\mathbf{r})\psi_i(\mathbf{r}) + U(\mathbf{r})\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}) \quad , \tag{2.7}$$

where $U(\mathbf{r})$ is a non-local potential and the local ionic potential is denoted by V_{ion}. The one-electron equations resemble single-particle Schrödinger equations.

The full Hartree-Fock equations are given by

$$\begin{aligned} \epsilon_i\psi_i(\mathbf{r}) &= \left(-\frac{1}{2}\nabla^2 + V_{ion}(\mathbf{r})\right)\psi_i(\mathbf{r}) + \sum_j \int d\mathbf{r}' \frac{|\psi_j(\mathbf{r}')|^2}{|\mathbf{r}-\mathbf{r}'|}\psi_i(\mathbf{r}) \\ &- \sum_j \delta_{\sigma_i\sigma_j} \int d\mathbf{r}' \frac{\psi_j^*(\mathbf{r}')\psi_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\psi_j(\mathbf{r}) \quad . \end{aligned} \tag{2.8}$$

The right hand side of the equations consists of four terms. The first and second give rise are the kinetic energy contribution and the electron-ion potential. The third term, or Hartree term, is the simply electrostatic potential arising from the charge distribution of N electrons. As written, the term includes an unphysical self-interaction of electrons when $j = i$. This term is cancelled in the fourth, or exchange term. The exchange term results from our inclusion of the Pauli principle and the assumed determinantal form of the wavefunction. The effect of exchange is for electrons of like-spin to avoid each other. Each electron of a given spin is consequently surrounded by an "exchange hole", a small volume around the electron which like-spin electron avoid.

The Hartree-Fock approximation corresponds to the conventional single-electron picture of electronic structure: the distribution of the N electrons is given simply by the sum of one-electron distributions $|\psi|^2$. This allows concepts such as labelling of electrons by angular momenta (" a 3d electron in a transition metal"), but it must be remembered that this is an artifact of the initial ansatz and that in some systems modifications are required to these ideas.

Hartree-Fock theory, by assuming a single-determinant form for the wavefunction, neglects correlation between electrons. The electrons are subject to an *average* non-local potential arising from the other electrons, which can lead to a poor description of the electronic structure. Although qualitatively correct in many materials and compounds, Hartree-Fock theory is insufficiently accurate to make accurate quantitative predictions.

Basis set expansion

In the preceding section, the single-particle Hartree-Fock equations were presented (equation 2.8). Numerical solutions usually are found by expanding the orbitals in a basis set.

$$\psi_i = \sum_k^M c_{ik}\phi_k \tag{2.9}$$

The unknown Hartree-Fock orbitals, ψ, are written as a linear expansion in M known basis functions ϕ. Inserting equation 2.9 into equation (2.8) leads to a set of matrix equations for the expansion coefficients, c_{ik}. The problem of solving the Hartree-Fock equations is reduced to a linear algebra problem, which may be solved by techniques such as iterative diagonalisation. [5, 6]

In practice, a basis of plane waves in periodic systems or localised Gaussians in finite systems, is commonly used. The basis set expansion represents an additional limitation of the techniques: unless the basis set expansion has sufficient freedom to encompass the exact solutions for the Hartree-Fock orbitals, ϕ, a compromise solution with a higher Hartree-Fock energy will be found[1]. In practical applications, convergence of the basis set must be studied to verify that the expansion is sufficiently complete.

2.4 Post Hartree-Fock techniques

The limitations of the Hartree-Fock method may be reduced by going beyond the ansatz of a single-determinant wavefunction. There are two broad categories of such approaches: those based on perturbation theory and those based on the variational principle. Among the latter approaches is the configuration interaction method which will be covered here because the focus of this thesis is on obtaining accurate many-body wavefunctions.[2]

2.4.1 Configuration interaction

Configuration interaction (CI) methods are one of the conceptually simplest methods for solving the many-body Hamiltonian. Although theoretically elegant, in principle exact, and relatively simple to implement, in practice *full* CI can be applied to only the smallest of systems.

[1] The solution found for the orbitals will be the “best fit” solution in a least-squares sense.

[2] Methods to apply perturbative corrections to configuration interaction have been developed. Applying perturbative corrections to a configuration interaction–derived wavefunction can be an efficient way of improving accuracy at lower cost than performing a higher level configuration interaction–based calculation. Uses of these techniques are quite specialised and give a small non–variational correction to the CI energy.

The basis for CI methods is the simple observation that an *exact* many-body wavefunction, Ψ, may be written as a linear combination of Slater determinants, D_k,

$$\Psi = \sum_{k=0}^{\infty} c_k D_k \quad , \tag{2.10}$$

where the D_k fully span the Hilbert space of the wavefunction. The determinants can be any complete set of N-electron antisymmetric functions but are typically constructed from Hartree-Fock orbitals such that D_0 is the ground-state Hartree-Fock determinant.

The Hartree-Fock "reference determinant" D_0 is by definition the best single-determinant approximation to the exact wavefunction Ψ. In most electronic systems, the Hartree-Fock energy accounts for the majority of the exact total energy, and the missing correlation energy is small. If the coefficients c_k are normalised then typically $c_0 \approx 1$ and all remaining c_k are very small. A very large number of configurations is required to yield energies and wavefunctions approaching the exact many-body wavefunction. In practice the expansion must be limited on physical grounds, as the total number of determinants is

$$k_{max} = \frac{M!}{N!(M-N)!} \quad , \tag{2.11}$$

where the length of the expansion k_{max} is given in terms of the number of electrons, N, and the number of basis states, M, in the expansion ($M >> N$).

The scientific problem in adapting the CI method into a practical one is to obtain the best wavefunction, and hence lowest CI energy, with the shortest expansion length. A typical approach would be to truncate the expansion after only double or quadruple excitations from the reference determinant, where an excitation consists of replacing a ground state occupied orbital by an unoccupied one. These levels of truncation are the CI singles-doubles (CISD) and CI singles-doubles-triples-quadruples (CISDTQ) methods. A formidable number of terms are still left in the expansion. Accurate applications of the methods are consequently limited due to their computational cost.

When performed within a finite reference space, an additional problem with the method becomes apparent: the methods lack "size-extensivity" and do not perform equally well in systems of differing size. As the size of system increases, the proportion of the electronic correlation energy contained within a fixed reference space (such as all single and double excitations) decreases. The lack of size-extensivity results in a non-cancellation of errors when systems of different sizes are compared, resulting in difficulties when interaction or bonding energies are required.

Despite these limitations, CI represents a controlled (and variational) improvement to the ground-state wavefunction, and may therefore be used in the determinantal parts of trial wavefunctions in QMC. Details of this approach are given in chapter 3.

2.4.2 Other methods

The lack of size extensivity and substantial cost of the CI method has led to the development of several related methods.

The coupled-cluster (CC) method [3,7] is one of the most important practical advances over the CI method. Although non-variational, it resolves the problem of size extensivity, and is often very accurate, but more expensive than (limited) CI. The CC method assumes an exponential ansatz for the wavefunction

$$\Psi_{CC} = \exp(\hat{T})\Psi_{HF} \quad , \tag{2.12}$$

where the coupled-cluster wavefunction is given by an excitation operator acting on a reference wavefunction, usually the Hartree-Fock determinant $\Psi_{HF} = D_0$. The operator $\hat{T}$ is generates k-fold excitations from a reference state

$$\hat{T} = \sum_k \hat{T}_k = \hat{T}_1 + \hat{T}_2 + \ldots \quad , \tag{2.13}$$

so that, for example,

$$\hat{T}_2 \Psi_{HF} = \sum t_{ij}^{ab} D_{ij}^{ab} \quad . \tag{2.14}$$

The operator $\hat{T}_2$ generates excitations from pairs of occupied states, ij, to pairs of virtual states, ab, from a reference Hartree-Fock determinant D. The expansion coefficients t_{ij}^{ab} are determined self-consistently.

A "coupled-cluster doubles" wavefunction is written as

$$\Psi_{CCD} = \exp(T_2)\Psi_{HF} = (1 + \hat{T}_2 + \hat{T}_2^2/2 + \hat{T}_2^3/3! + \ldots)\Psi_{HF} \quad , \tag{2.15}$$

$$\Psi_{CCD} = D_0 + \sum t_{ij}^{ab} D_{ij}^{ab} + \frac{1}{2} \sum \sum t_{ij}^{ab} t_{kl}^{cd} D_{ijkl}^{abcd} + \ldots \quad . \tag{2.16}$$

The CC expansion is usually terminated after all double excitations or all quadruple excitations have been included. It may be shown [3] that this expansion is size consistent. By including many excitation terms in the expansion, CC methods are computationally very expensive relative to HF calculations. Formally, CC singles-doubles scales as the sixth power of the number of basis states included in the expansion, and calculations including up to quadruple excitations scale as the tenth power of the number of states.

The key limitation of the CC methods are their rapid increase in computational cost with system size. This presently limits the methods to small molecular systems.

2.4.3 Limitations

Hartree-Fock is a simple theory which satisfies the commonly known features of fermionic wavefunctions. The theory generates wavefunctions that are antisymmetric with respect to the exchange of two electron positions and includes exchange between like–spin electrons. The cost of a Hartree-Fock calculation formally scales with the cube of the number of basis functions, but depending on implementation the scaling can be between linear and quartic with system size. It is insufficiently accurate for quantitative predictions of the properties of many compounds. By neglecting electron correlation, interaction energies are typically very poor. A Hartree-Fock wavefunction is a well-controlled approximation to the many-body wavefunction, and for this reason Hartree-Fock continues to be widely used: it is often *predictably* accurate or inaccurate, and therefore useful for determining qualitative information such as trends in a structural parameter with system size.

Almost all post Hartree-Fock methods share the combined limitations of a poor scaling with system size and a strong basis set dependence. In practice, post Hartree-Fock methods typically scale with the fourth or higher power of the number of basis states included in the calculation, limiting their application to small systems. The basis states depend on the underlying basis set used in their numeric expansion, equation 2.9, and it is commonly found that use of improved methods requires an improved basis set, further increasing the cost of calculation. CI and CC–based methods effectively transform the electron-correlation problem into a basis set problem, where the basis set is the set of molecular orbitals derived from a Hartree-Fock (or similar) calculation.

2.5 Density functional theory

Density functional theory (DFT) is a powerful, formally exact theory (see Refs. [3, 8, 9] and references within). It is distinct from quantum chemical methods in that it is a non-interacting theory and does not yield a correlated N-body wavefunction. In the Kohn-Sham [10] DFT, the theory is a one-electron theory and shares many similarities with Hartree-Fock. DFT has come to prominence over the last decade as a method *potentially* capable of very accurate results at low cost. In practice, approximations are required to implement the theory, and a significantly variable accuracy results. Calibration studies are therefore required to establish the likely accuracy in a given class of systems.

2.5.1 The Hohenberg-Kohn theorem

The Hohenberg-Kohn theorem [11] states that if N interacting electrons move in an external potential $V_{ext}(\mathbf{r})$, the ground-state electron density $n_0(\mathbf{r})$ minimises the functional

$$E[n] = F[n] + \int n(\mathbf{r})V_{ext}(\mathbf{r})d\mathbf{r} \tag{2.17}$$

where F is a universal functional of n and the minimum value of the functional E is E_0 the exact ground-state electronic energy.

The proof of equation 2.17 is straightforward. It is a proof only of existence; additional theory is required before a method can be implemented.

Levy [12] gave a particularly simple proof of the Hohenberg-Kohn theorem which is as follows:

A functional O is defined as

$$O\left[n(\mathbf{r})\right] = \min_{|\Psi>\rightarrow n(\mathbf{r})} \left\langle \Psi|\hat{O}|\Psi \right\rangle \quad , \tag{2.18}$$

where the expectation value is found by searching over all wavefunctions, Ψ, giving the density $n(\mathbf{r})$ and selecting the wavefunction which minimises the expectation value of $\hat{O}$.

$F[n(\mathbf{r})]$ is defined by

$$F\left[n(\mathbf{r})\right] = \min_{|\Psi>\rightarrow n(\mathbf{r})} \left\langle \Psi|\hat{F}|\Psi \right\rangle \quad , \tag{2.19}$$

so that

$$\hat{F} = \sum_i -\frac{1}{2}\nabla_i^2 + \frac{1}{2}\sum_{i\neq j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad . \tag{2.20}$$

Considering an N-electron ground state wavefunction Ψ_0 which yields a density $n(\mathbf{r})$ and minimises $< \Psi|\hat{F}|\Psi >$, then from the definition of the functional E

$$E[n(\mathbf{r})] = F[n(\mathbf{r})] + \int n(\mathbf{r})V_{\text{ext}}(\mathbf{r})d\mathbf{r} =< \Psi|\hat{F} + V_{\text{ext}}|\Psi > \quad . \tag{2.21}$$

The Hamiltonian is given by $\hat{F} + V_{\text{ext}}$, and so $E[n(\mathbf{r})]$ must obey the variational principle,

$$E[n(\mathbf{r})] \geq E_0 \quad . \tag{2.22}$$

This completes the first part of the proof, which places a lower bound on $E[n(\mathbf{r})]$.

From the definition of $F[n(\mathbf{r})]$ equation 2.19 we obtain

$$F[n_0(\mathbf{r})] \leq < \Psi_0|\hat{F}|\Psi_0 >, \tag{2.23}$$

since Ψ_0 is a trial wavefunction yielding $n_0(\mathbf{r})$. Combining $\int n(\mathbf{r})V_{\text{ext}}(\mathbf{r})d\mathbf{r}$ with the above equation gives

$$E[n_0(\mathbf{r})] \leq E_0 \quad , \tag{2.24}$$

which in combination with equation 2.22 yields the key result

$$E[n_0(\mathbf{r})] = E_0 \quad , \tag{2.25}$$

completing the proof.

2.5.2 The Kohn-Sham equations

Kohn and Sham [10] derived a coupled set of differential equations enabling the ground state density $n_0(\mathbf{r})$ to be found.

Kohn and Sham separated $F[n(\mathbf{r})]$ into three distinct parts, so that the functional E becomes

$$E\left[n(\mathbf{r})\right] = T_s\left[n(\mathbf{r})\right] + \frac{1}{2}\int\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + E_{XC}[n(\mathbf{r})] + \int n(\mathbf{r})V_{\text{ext}}(\mathbf{r})d\mathbf{r} \quad , \tag{2.26}$$

where $T_s\left[n(\mathbf{r})\right]$ is defined as the kinetic energy of a non-interacting electron gas with density $n(\mathbf{r})$,

$$T_s\left[n(\mathbf{r})\right] = -\frac{1}{2}\sum_{i=1}^{N}\int \psi_i^*(\mathbf{r})\nabla^2\psi_i(\mathbf{r})d\mathbf{r} \quad , \tag{2.27}$$

and *not* the kinetic energy of the real system. Equation 2.26 also defines the exchange-correlation energy functional $E_{XC}[n]$. Introducing a normalisation constraint on the electron density, $\int n(\mathbf{r})d\mathbf{r} = N$, we obtain

$$\begin{aligned} \frac{\delta}{\delta n(\mathbf{r})}\left[E[n(\mathbf{r})] - \mu\int n(\mathbf{r})d\mathbf{r}\right] &= 0 \\ \Rightarrow \frac{\delta E[n(\mathbf{r})]}{\delta n(\mathbf{r})} &= \mu \quad . \end{aligned} \tag{2.28}$$

Equation 2.28 may now be rewritten in terms of an effective potential, $V_{\text{eff}}(\mathbf{r})$,

$$\frac{\delta T_s\left[n(\mathbf{r})\right]}{\delta n(\mathbf{r})} + V_{\text{eff}}(\mathbf{r}) = \mu \quad , \tag{2.29}$$

where

$$V_{\text{eff}}(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + \int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' + V_{XC}(\mathbf{r}) \tag{2.30}$$

and

$$V_{XC}(\mathbf{r}) = \frac{\delta E_{XC}[n(\mathbf{r})]}{\delta n(\mathbf{r})} \quad . \tag{2.31}$$

Crucially, *non-interacting* electrons moving in an external potential $V_{eff}(\mathbf{r})$ would result in the same equation 2.29. To find the ground state energy, E_0, and the ground state density, n_0, the one electron Schrödinger equation

$$\left(-\frac{1}{2}\nabla_i^2 + V_{\text{eff}}(\mathbf{r}) - \epsilon_i\right)\psi_i(\mathbf{r}) = 0 \tag{2.32}$$

should be solved self-consistently with

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\psi_i(\mathbf{r})|^2 \quad , \tag{2.33}$$

and equations 2.29 and 2.30. A self-consistent solution is required due to the dependence of $V_{\text{eff}}(\mathbf{r})$ on $n(\mathbf{r})$.

The above equations provide a theoretically exact method for finding the ground state energy of an interacting system provided the form of E_{XC} is known. Unfortunately, the form of E_{XC} is in general unknown and its exact value has been calculated for only a few very simple systems. In electronic structure calculations E_{XC} is most commonly approximated within the local density approximation or generalised-gradient approximation.

2.5.3 The local density approximation

In the local density approximation (LDA), the value of $E_{XC}[n(\mathbf{r})]$ is approximated by the exchange–correlation energy of an electron in an homogeneous electron gas of the same density $n(\mathbf{r})$, i.e.

$$E_{XC}^{LDA}[n(\mathbf{r})] = \int \epsilon_{XC}(n(\mathbf{r}))n(\mathbf{r})d\mathbf{r} \quad . \tag{2.34}$$

The most accurate data for $\epsilon_{XC}(n(\mathbf{r}))$ is from Quantum Monte Carlo calculations. [13][3] The LDA is often surprisingly accurate and for systems with slowly varying charge densities generally gives very good results. The failings of the LDA are now well established: it has a tendency to favour more homogeneous systems and over-binds molecules and solids. In weakly bonded systems these errors are exaggerated and bond lengths are too short. In good systems where the LDA works well, often those mostly consisting of *sp* bonds, geometries are good and bond lengths and angles are accurate to within a few percent. Quantities such as the dielectric and piezoelectric constant are approximately 10% too large.

The principle advantage of LDA-DFT over methods such as Hartree-Fock is that *where the LDA works well* (correlation effects are well accounted for) many experimentally rele-

[3]Published in 1980, this is now one of the oldest, most widely referenced papers in the field of numerical electronic structure.

vant physical properties can be determined to a useful level of accuracy. Difficulties arise where it is not clear whether the LDA is applicable. For example, although the LDA performs well in bulk group-IV semiconductors it is not immediately clear how well it performs at surfaces of these materials.

2.5.4 Limitations

Despite the remarkable success of the LDA, its limitations mean that care must be taken in its application. For systems where the density varies slowly, the LDA tends to perform well, and chemical trends are well reproduced. In strongly correlated systems where an independent particle picture breaks down, the LDA is very inaccurate. The transition metal oxides XO (X=Fe,Mn,Ni) are all Mott insulators, but the LDA predicts that they are either semiconductors or metals. The LDA has been applied to high T_c superconductors, but finds several to be metallic, when in reality they are insulating at 0K. [3]

The LDA finds the wrong ground state for in many simpler cases. For example, the LDA finds the wrong ground state for the titanium atom. The LDA does not account for van der Waals bonding, and gives a very poor description of hydrogen bonding. These phenomena are essential for most of biochemistry: the structure of DNA of depends critically on hydrogen bonding, as do the changes in the structure of most molecules on solvation.

The success of the LDA has been shown by QMC calculations to result from a real-space cancellation of errors in the LDA exchange and correlation energies. [14, 15] This is illustrated in figure 2.1, where the exchange and correlation energy densities of silicon are compared with an accurate QMC calculation. The cancellation represents a difficulty when improvements to the LDA are attempted, as an improvement in only the exchange or correlation contributions may give worse results.

An obvious approach to improving the LDA is to include gradient corrections, by making E_{XC} a functional of the density and its gradient:

$$E_{XC}^{GGA}[n(\mathbf{r})] = \int \epsilon_{XC}(n(\mathbf{r}))n(\mathbf{r})d\mathbf{r} + \int F_{XC}[n(\mathbf{r}), |\nabla n(\mathbf{r})|]d\mathbf{r} \quad , \tag{2.35}$$

where F_{XC} is a correction chosen to satisfy one or several known limits for E_{XC}.

Clearly, there is no unique recipe for F_{XC}, and several dozen functionals have been proposed in the literature. They do not always represent a systematic improvement over the LDA and results must be carefully compared against experiment.[4] The development of improved functionals is currently a very active area of research and although incremental improvements are likely, it is far from clear whether the research will be successful in

[4] An illustration of this point is found in chapter 8, where different functionals give significantly different energetic orderings compared with QMC calculations for carbon clusters up to C_{32}.

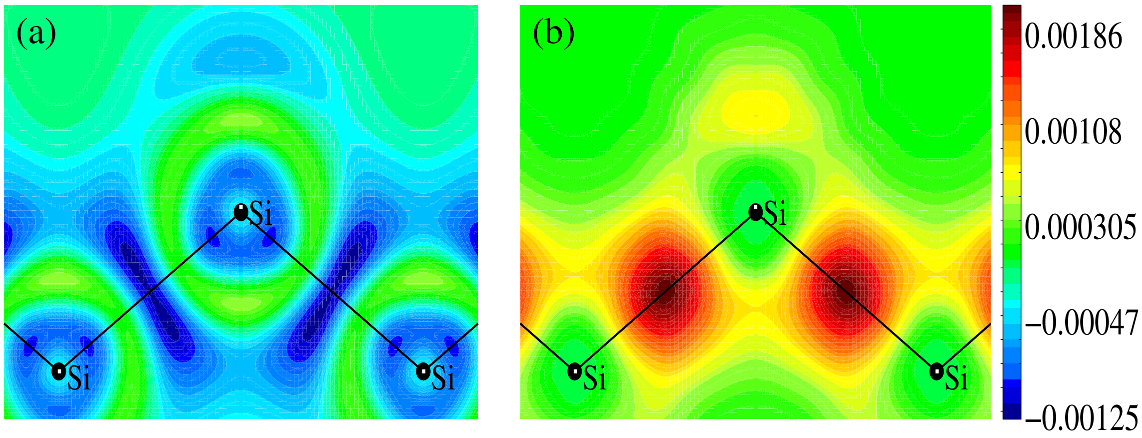

Figure 2.1: Contour plots of the LDA exchange and correlation energy densities compared with VMC calculations for Si in the diamond structure in the (110) plane: (a) $e_x^{\mathrm{VMC}}(\mathbf{r}) - e_x^{\mathrm{LDA}}(\mathbf{r})$; (b) $e_c^{\mathrm{VMC}}(\mathbf{r}) - e_c^{\mathrm{LDA}}(\mathbf{r})$. The chain of atoms and bonds are represented schematically. The contours are in atomic units. Note that in the bonding regions between silicon atoms, the error in the exchange energy density tends to cancel with the error in the correlation energy density. Figure courtesy R. Q. Hood. [15]

providing the substantial increase in accuracy desired.

2.6 Summary

The main problem of the Hartree-Fock and DFT methods is the underlying treatment of electron correlation. In the Hartree-Fock method, electron correlations beyond a mean field picture are entirely neglected, whereas in DFT they are included approximately via a functional $E_{XC}[n]$. DFT methods consequently require careful calibration to establish their accuracy (or inaccuracy) on a case by case basis. Both methods provide a relatively inexpensive route to performing computational physics, chemistry and materials science, provided only trends and not highly accurate quantitative predictions are required.

Post Hartree-Fock methods are potentially very accurate, but their poor scaling with system size limits their usefulness, restricting their application to small molecules and excluding most topics in condensed matter.

As will be discussed in chapter 3, Quantum Monte Carlo methods combine favourable scaling with system size with a very accurate treatment of electron correlation. This renders them ideal for correlated studies of large molecules and condensed matter systems when high accuracy is required.

An additional novel and very practical feature of Quantum Monte Carlo methods is that they provide a direct measure of the likely accuracy obtained. This feedback represents a fundamental advance over current density functional and quantum chemical methodology where the accuracy obtained is both unknown and variable. This is clearly advantageous when new systems are investigated, where there is little or no reference data.

Chapter 3

Quantum Monte Carlo methods

3.1 Introduction

In this chapter, the principal ideas behind Quantum Monte Carlo methods are outlined. The theory of Monte Carlo integration is presented, and overviews given of the variational Monte Carlo method and the more sophisticated and accurate diffusion Monte Carlo method. These methods are both applications of Monte Carlo integration and form the basis for most of the calculations in this thesis.

Efficient implementation of these methods is critical. Inefficiencies increase computational overheads and effectively limit the size of system and statistical accuracy that may be obtained. Although some details and suggestions for efficient calculations are given, full details of application of the methods to electronic structure problems and their implementation are presented in the following chapter.

3.2 Monte Carlo methods

Monte Carlo methods were first developed as a method for estimating integrals that could not be evaluated analytically. Although many statistical techniques are now included in the category of "Monte Carlo methods" [16, 17], the method used in this thesis is principally Monte Carlo integration.

3.2.1 Monte Carlo integration

A straightforward application of Monte Carlo is the evaluation of definite integrals. Consider the one dimensional integral

$$E = \int_a^b f(x)dx \quad . \tag{3.1}$$

By application of the mean value theorem of calculus, the integral may be approximated by

$$E_N = \frac{(b-a)}{N} \sum_{i=1}^{N} f(x_i) \quad , \tag{3.2}$$

where the points x_i fully cover the range of integration. In the limit of a large number of points N, E_N tends to the exact value E.

A conventional choice for the points x_i would be a uniform grid. More accurate methods of numeric quadrature, such as Simpson's rule and Gaussian quadrature [18] use a weighted average of the points:

$$E_N = \frac{(b-a)}{N} \frac{\sum_{i=1}^{N} w_i f(x_i)}{\sum_{i=1}^{N} w_i} \quad . \tag{3.3}$$

These methods are highly effective for low dimensional integrals. The computational cost of evaluating an integral, to a fixed accuracy, of dimensionality d increases as N^d. An alternate and more efficient approach is to select the points x_i randomly, from a given probability distribution by Monte Carlo methods.

If points are selected at random over the interval $[a, b]$, the Monte Carlo estimate of the integral, equation 3.1, becomes

$$E_N = (b-a)\overline{f} = \frac{(b-a)}{N} \sum_{i=1}^{N} f(x_i) + O(1/\sqrt{(N)}) \quad , \tag{3.4}$$

where $\overline{f}$ denotes the mean value of f over the set of sampled points $\{x_i\}$. By the central limit theorem, [19] the set of all possible sums over different $\{x_i\}$ will have a Gaussian distribution. The standard deviation σ_N of the different values of E_N is a measure of the uncertainty in the integral's value

$$\sigma_N = \sqrt{\frac{\frac{b-a}{N} \sum_{i=1}^{N} f^2(x_i) - E_N^2}{N-1}} \quad . \tag{3.5}$$

The probability that E is within $E_N \pm \sigma_N$ is ≈ 0.68 and the probability of being with $2\sigma_N$ is ≈ 0.95. This error decays as $1/\sqrt{N}$ *independent* of the dimensionality of the integral, unlike grid based methods which have a strong dimensional dependence.

3.2.2 Importance sampling

Simple schemes for Monte Carlo integration can suffer from low efficiency. Many functions of interest have significant weight in only a few regions. For example, most of the contributions to an integral of a simple Gaussian are located near the central peak. In a simple Monte Carlo integration scheme, points are sampled uniformly, wasting considerable effort sampling the tails of the Gaussian. Techniques for overcoming this problem act to increase the density of points in regions of interest and hence improve the overall efficiency. These techniques are called importance sampling.

Points are sampled from a non-uniform distribution $w(x)$, where $w(x)$ is always positive and chosen to approximate $f(x)$ over the region of interest. The integral may now be evaluated by selecting points from the probability distribution $p(x)$,

$$p(x) = \frac{w(x)}{\int_a^b w(x)dx} \quad , \tag{3.6}$$

so that the integral may be evaluated

$$E = \int_a^b g(x)p(x)dx \approx \frac{1}{N}\sum_{i=1}^{N} g(x_i) \quad , \tag{3.7}$$

where $g(x) = f(x)/p(x)$ and the points x_i are generated from the probability distribution $p(x)$.

Provided $f(x)$ is well described by $w(x)$, $g(x)$ remains close to unity, so that the standard deviation defined in equation 3.5 is greatly reduced. (The best possible choice for $p(x)$ is $|f(x)/g(x)|$.) The integration procedure is therefore more efficient, although the error still decays as $1/\sqrt{N}$.

3.2.3 The Metropolis algorithm

In this thesis, the Metropolis algorithm [20] is used in the importance sampling of multi-dimensional integrals. The Metropolis algorithm generates a random walk of points distributed according to a required probability distribution. From an initial "position" in phase or configuration space, a proposed "move" is generated and the move either accepted or rejected according to the Metropolis algorithm. By taking a sufficient number of trial steps all of phase space is explored and the Metropolis algorithm ensures that the points are distributed according to the required probability distribution.

The Metropolis algorithm may be derived by considering the probability density, ρ at two points in configuration space X and X'. Configuration space is specified by the integration variables and limits. For one dimensional integration, $X = \{x\}$ for $x \in [a, b]$. Averaged over many trial steps, the average number of samples at X and X' should be equal, $\rho(X) = \rho(X')$.

The probability for a trial move from X to X' defined $T(X \to X')$ and we require

$$T(X \to X') = T(X' \to X) \quad . \tag{3.8}$$

If the probability of accepting a move from X to X' is $P(X \to X')$ then the total probability of accepting a move from X to X' is $P(X \to X')T(X \to X')$. Therefore, at equilibrium

$$\rho(X)P(X \to X')T(X \to X') = \rho(X')P(X' \to X)T(X' \to X) \quad . \tag{3.9}$$

The Metropolis algorithm corresponds to choosing

$$P(X \to X') = \min\left\{1, \frac{\rho(X')}{\rho(X)}\right\} \quad . \tag{3.10}$$

For the Metropolis algorithm to be valid, it is essential that the random walk is ergodic, that is any point X' in configuration space may be reached from any other point X. In some applications of the Metropolis algorithm, parts of configuration space may be difficult to reach. Long simulations or a modification of the algorithm are then necessary. Many methods have been developed to cope with this "slowing down", [16] but for the applications presented here the Metropolis algorithm has been found both adequate and sufficient for the problems investigated.

3.3 Variational Monte Carlo

Variational Monte Carlo (VMC) is based on a direct application of Monte Carlo integration to explicitly correlated many-body wavefunctions. The variational principle of quantum mechanics, derived in the following section, states that the energy of a trial wavefunction will be greater than or equal to the energy of the exact wavefunction. Optimised forms for many-body wavefunctions enable the accurate determination of expectation values.

3.3.1 The variational principle

The variational principle of quantum mechanics may be derived by expanding a normalized trial wavefunction, ψ_T, in terms of the exact normalized eigenstates of the Hamiltonian,

$$\psi_T = \sum_{i=0}^{\infty} c_i \psi_i \quad , \tag{3.11}$$

where the expansion coefficients, c_i, are normalised

$$\sum_{i=0}^{\infty} |c_i|^2 = 1 \quad . \tag{3.12}$$

The expectation of the many-body Hamiltonian, $\hat{H}$, may be evaluated

$$
\begin{aligned}
< \psi_T|\hat{H}|\psi_T > &= < \sum_i c_i\psi_i|\hat{H}|\sum_j c_j\psi_j > \\
&= \sum_i\sum_j c_i^* c_j < \psi_i|\hat{H}|\psi_j > \\
&= \sum_i |c_i|^2\epsilon_i \quad , \qquad (3.13)
\end{aligned}
$$

where

$$
\epsilon_i = < \psi_i|\hat{H}|\psi_i > \quad . \qquad (3.14)
$$

The expectation value of a trial wavefunction with the Hamiltonian must therefore be greater than or equal to the true ground state energy.

Variational calculations depend crucially on the form of trial wavefunction used. By selecting trial wavefunctions on physically motivated grounds, accurate wavefunctions may be obtained. Commonly, wavefunctions obtained from a Hartree-Fock or similar calculations are used and additional parameters added, building-in additional physics such as known limits and derivatives of the many-body wavefunction. The additional variational freedom is then exploited to further optimize the wavefunction. These techniques are briefly described in the remainder of this section, and in more detail in chapters 4 and 5.

3.3.2 Monte Carlo integration

Before the expectation value of a trial wavefunction with the many-body Hamiltonian may be computed, the integral must be transformed into a form suitable for Monte Carlo integration.

Trial wavefunctions, ψ_T, are dependent on the set of N electron positions, $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N\}$. The expectation value is given by

$$
E = \frac{\int \psi_T^* \hat{H} \psi_T d\mathbf{R}}{\int \psi_T^* \psi_T d\mathbf{R}} \quad , \qquad (3.15)
$$

which may be rewritten in an importance sampled form in terms of the probability density $|\psi_T|^2$:

$$
E = \frac{\int |\psi_T|^2 \frac{\hat{H}\psi_T}{\psi_T} d\mathbf{R}}{\int |\psi_T|^2 d\mathbf{R}} \quad . \qquad (3.16)
$$

The Metropolis algorithm samples configurations - sets of electron positions $\mathbf{R}$ - from the probability distribution $|\psi_T|^2$, and the variational energy is obtained by averaging the "local energy" E_L over the set of configurations $\{\mathbf{R}\}$:

$$E_L(\mathbf{R}) = \frac{\hat{H}\psi_T(\mathbf{R})}{\psi_T(\mathbf{R})} \tag{3.17}$$

and

$$E = \frac{1}{N}\sum E_L(\mathbf{R}_i) \quad . \tag{3.18}$$

3.3.3 The local energy

The local energy, $E_L(\mathbf{R})$, equation 3.17 is one of the central quantities in QMC methods. It occurs in both the variational and diffusion Monte Carlo algorithms and its properties are exploited to optimise trial wavefunctions.

The local energy has the useful property that for an exact eigenstate of the Hamiltonian, the local energy is constant. For a general trial wavefunction the local energy is not constant and the variance of the local energy is a measure of how well the trial wavefunction approximates an eigenstate. The spatially averaged variance of the local energy is therefore a quantity suitable for optimisation, and methods exploiting this observation are presented in chapter 5.

Determination of the local energy is one of the most computationally costly operations performed in QMC calculations. Application of the Hamiltonian to the trial wavefunction requires computation of the second derivatives of the wavefunction and the calculation of the electron-electron and electron-ion potentials. Efficient methods for the evaluation of $E_L(\mathbf{R})$ are given in chapter 4.

3.3.4 Trial wavefunctions

The choice of trial wavefunction is critical in VMC calculations. All observables are evaluated with respect to the probability distribution $|\Psi_T(\mathbf{R})|^2$. The trial wavefunction, $\Psi_T(\mathbf{R})$, must well approximate an exact eigenstate for all $\mathbf{R}$ in order that accurate results are obtained. Improved trial wavefunctions also improve the importance sampling, reducing the cost of obtaining a certain statistical accuracy.

Quantum Monte Carlo methods are able to exploit trial wavefunctions of arbitrary forms. Any wavefunction that is physical and for which the value, gradient and laplacian of the wavefunction may be efficiently computed can be used.

The power of Quantum Monte Carlo methods lies in the flexibility of the form of the trial wavefunction. In early studies of bosonic ^{4}He by McMillan [21] the wavefunction was taken to be a Jastrow or two-body correlation function, [22]

$$\psi = \exp\left[\sum_{i<j}^{N} -u(r_{ij})\right] \quad . \tag{3.19}$$

The function u was chosen to miminise the energy of the system under consideration, by choosing u to increase the probability of particles being at a distance that minimises their interaction energy. Variations on this idea have been successfully applied to fermionic systems by multiplying a determinantal wavefunction by a two-body or higher body correlation functions. [23, 24] Well chosen correlation functions include correlation effects more efficiently than CI-based approaches.

It is important that the trial wavefunction satisfies as many known properties of the exact wavefunction as possible. A determinantal wavefunction is correctly anti-symmetric with respect to the exchange of any two electrons. An additional *local* set of constraints which may be readily imposed are for electron-electron and electron-nucleus coalescence. These constraints are the “cusp conditions” [25], and are a constraint on the derivatives of the wavefunction. For particle-particle coalescence, it may be shown that [25]

$$\left.\frac{d\Psi}{dr}\right|_{r=0} = \zeta\Psi|_{r=0} \quad , \tag{3.20}$$

where Ψ is the many-body wavefunction and r is either an electron-electron or electron-ion separation. ζ is $-Z$ for coalescence of an electron with a nucleus of charge Z, and $1/2$ or $1/4$ for spin parallel and anti-parallel electron-electron coalescence respectively.

A commonly used and simple form of Jastrow factor suitable for solids is

$$u(r) = \frac{A}{r}(1 - e^{-\frac{r}{F}}) \quad , \tag{3.21}$$

where F is parameterised in terms of A and chosen to satisfy the electron-electron cusp conditions,

$$\left.\frac{du}{dr}\right|_{r=0} = \begin{cases} -\frac{1}{4} & \text{for parallel spins} \\ -\frac{1}{2} & \text{for anti-parallel spins} \end{cases} \quad , \tag{3.22}$$

and the complete trial wavefunction is then taken to be

$$\Psi(\mathbf{R}) = D(\mathbf{R}) \exp\left[\sum_{i<j}^{N} -u(r_{ij})\right] \quad , \tag{3.23}$$

where D is a determinant of Hartree-Fock or DFT orbitals. The variational freedom in the value of A is used to optimise the trial wavefunction.

The development of better approximations to exact many-body wavefunctions, and practical methods for obtaining them will remain an important area of research due to the improved levels of accuracy and efficiency that result. The Jastrow factors used in this thesis contain of the order of 10-200 parameters and details of parameter selection and optimisation are given in the next chapter.

3.3.5 The VMC algorithm

The VMC algorithm consists of two distinct phases. In the first a walker consisting of an initially random set of electron positions is propagated according to the Metropolis algorithm, in order to equilibrate it and begin sampling $|\Psi|^2$. In the second phase, the walker continues to be moved, but energies and other observables are also accumulated for later averaging and statistical analysis.

To simplify the notation, $\Psi(\mathbf{r}')$ is used to indicate the wavefunction with a single electron moved to point $\mathbf{r}'$ and all other electrons unmoved:

1. Equilibration phase:
   (a) Generate initial configuration using random positions for the electrons
   (b) For every electron in the configuration:
      i. Propose a move from $\mathbf{r}$ to $\mathbf{r}'$
      ii. Compute $w = |\Psi(\mathbf{r}')/\Psi(\mathbf{r})|^2$
      iii. Accept or reject move according to Metropolis probability $\min(1, w)$
   (c) Repeat configuration moves until equilibrated
2. Accumulation phase:
   (a) For every electron in the configuration:
      i. Propose a move from $\mathbf{r}$ to $\mathbf{r}'$
      ii. Compute $|\Psi(\mathbf{r}')/\Psi(\mathbf{r})|^2$
      iii. Accumulate the contribution to the local energy, and other observables, at $\mathbf{r}'$ and at $\mathbf{r}$, weighted by the Metropolis acceptance and rejection probabilities respectively. [26]
      iv. Accept or reject move according to Metropolis acceptance probability
   (b) Repeat configuration moves until sufficient data accumulated

In this algorithm, the electrons are moved individually and not as a whole configuration. This improves the efficiency of the algorithm in larger systems, where configuration moves require increasingly small steps to maintain the acceptance ratio. Observables are also accumulated on a per electron basis, weighted by the acceptance and rejection probabilities:

$$< O >= \frac{1}{M}\sum_{i=1}^{M}\left[p_i O_i(\mathbf{r}') + (1-p_i)O_i(\mathbf{r})\right] \quad , \tag{3.24}$$

where p_i indicates the Metropolis acceptance probability of electron i and $O_i(\mathbf{r})$ its contribution to the value of the observable, and there are M moves in the simulation. This formula, which can be applied to any observable, has improved statistics compared with averages computed with reference to the unmoved (or moved) positions alone. [26]

3.4 Diffusion Monte Carlo

Diffusion Monte Carlo (DMC) is a projector or Green's function based method for solving for the ground state of the many-body Schrödinger equation. In principle the DMC method is exact, although in practice, several well-controlled approximations must be introduced for calculations to remain tractable.

3.4.1 Outline of method

The DMC method is based on rewriting the Schrödinger equation in imaginary time, $\tau = it$. The imaginary time Schrödinger equation is:

$$\frac{\partial|\psi\rangle}{\partial\tau} = -\hat{H}|\psi\rangle \quad . \tag{3.25}$$

The state $|\psi\rangle$ is expanded in eigenstates of the Hamiltonian

$$|\psi\rangle = \sum_{i=0}^{\infty} c_i|\phi_i\rangle \quad , \tag{3.26}$$

where

$$\hat{H}|\phi_i\rangle = \epsilon_i|\phi_i\rangle \quad . \tag{3.27}$$

A formal solution of the imaginary time Schrödinger equation is

$$|\psi(\tau_1+\delta\tau)\rangle = e^{-\hat{H}\delta\tau}|\psi(\tau_1)\rangle \quad , \tag{3.28}$$

where the state $|\psi\rangle$ evolves from imaginary time τ_1 to a later time $\tau_1 + \delta\tau$. If the initial state $|\psi(\tau_1)\rangle$ is expanded in energy ordered eigenstates, following equation 3.26, then

$$|\psi(\delta\tau)\rangle = \sum_{i=0}^{\infty} c_i e^{-\epsilon_i \delta\tau} |\phi_i\rangle \quad . \tag{3.29}$$

Hence any initial state, $|\psi\rangle$, that is not orthogonal to the ground state, $|\phi_0\rangle$, will evolve to the ground state in the long time limit:

$$\lim_{\tau\to\infty} |\psi(\tau)\rangle = c_0 e^{-\epsilon_0 \tau} |\phi_0\rangle \quad . \tag{3.30}$$

This derivation shares many formal similarities with that given for the variational principle in section 3.3.1. However in the DMC method the imaginary time evolution results in excited states decaying exponentially fast, whereas in the VMC method any excited state contributions remain and contribute to the VMC energy.

The DMC method is a realisation of the above derivation in position space. equation 3.30 becomes

$$\lim_{\tau\to\infty} \psi(\mathbf{R}, \tau) = c_0 e^{-\epsilon_0 \tau} \phi_0(\mathbf{R}) \quad . \tag{3.31}$$

By introducing a constant offset to the energy, $E_T = \epsilon_0$, the long-time limit of equation 3.31 can be kept finite. If the Hamiltonian is separated into the kinetic energy and potential terms, the imaginary time Schrödinger equation, equation 3.25, takes on a form similar to a diffusion equation:

$$-\frac{\partial \psi(\mathbf{R}, \tau)}{\partial \tau} = \left[\sum_{i=1}^{N} -\frac{1}{2} \nabla_i^2 \psi(\mathbf{R}, \tau) \right] + (V(\mathbf{R}) - E_T) \psi(\mathbf{R}, \tau) \quad . \tag{3.32}$$

This equation is a diffusion equation where $\psi(\mathbf{R})$ may be interpreted as the density of diffusing particles (or "walkers"), and the term $(V(\mathbf{R}) - E_T)$ is a rate term describing a potential-dependent increase or decrease in the particle density. This is shown schematically in one-dimension for a simple potential in figure 3.1.

The above equation may be transformed into a form suitable for Monte Carlo methods, [23,27] but this leads to a very inefficient algorithm. The potential $V(\mathbf{R})$ is unbounded in coulombic systems and hence the rate term, $V(\mathbf{R}) - E_T$, can diverge. Large fluctuations in the particle density then result and give impractically large statistical errors. These fluctuations may be substantially reduced by the incorporation of importance sampling in the algorithm.

The above method also lacks the important constraint that for electronic systems the solution should be antisymmetric with respect to the exchange of any two electrons, i.e.

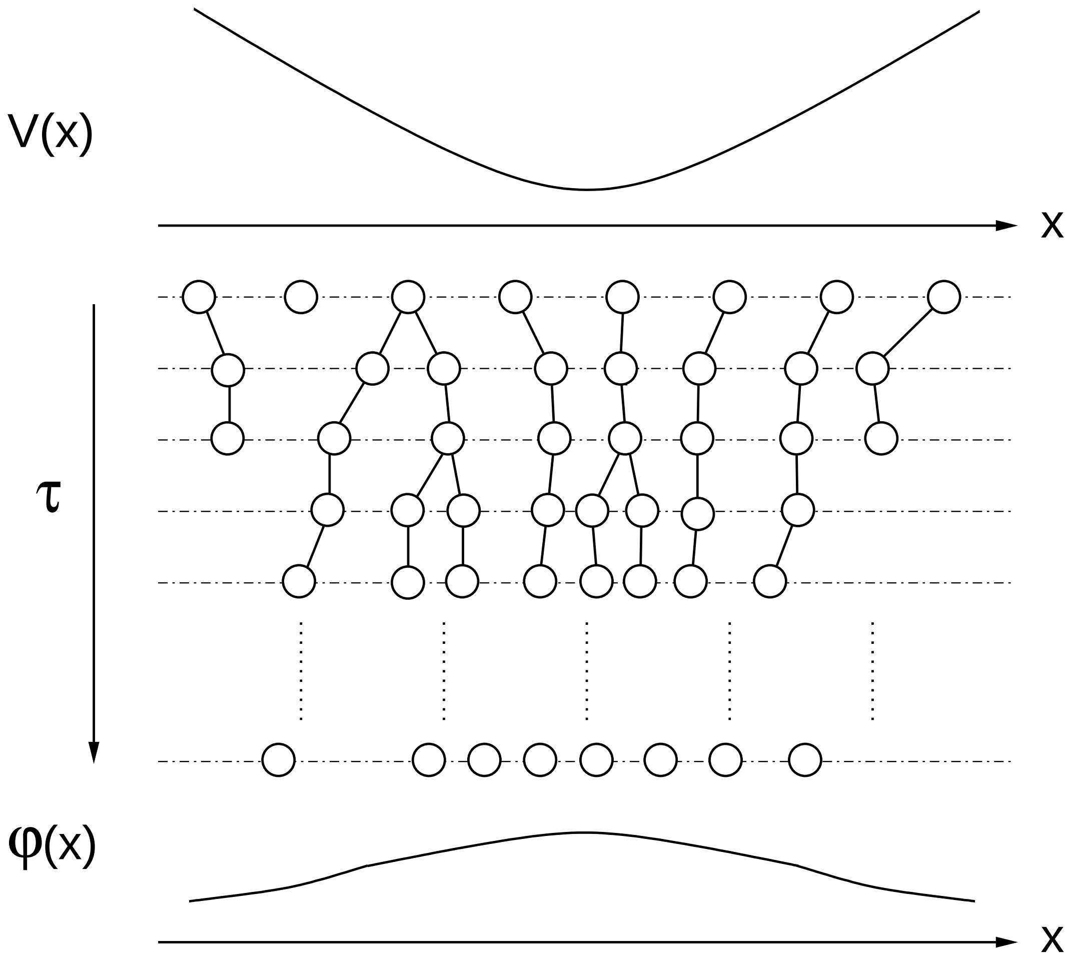

Figure 3.1: The evolution of walkers in the DMC method. A single particle in confined in a simple potential well, $V(x)$. An initially even distribution of walkers (circles) is propagated in imaginary time, τ. The distribution gradually evolves by a process of diffusion and branching to a distribution representative of the ground state wavefunction, $\psi(x)$. Note that where the potential energy is low, walkers tend to branch giving a higher density of walkers, and where the potential energy is high, walkers tend to be removed.

is fermionic. This constraint may be imposed by assuming a form for the nodal surface of the fermionic wavefunction, but at the expense of obtaining only a variational solution to the exact ground state energy. This method of constraint, the *fixed-node approximation*, and other methods of enforcing fermionic anti-symmetry are discussed in section 3.4.4.

3.4.2 Importance sampling

Importance sampling is essential for DMC methods, if the simulation is to be efficient. [27] A trial or guiding wavefunction, $\Psi_G(\mathbf{R})$, which closely approximates the ground state wavefunction is introduced. The derivation of Reynolds *et. al.* [23, 28] is followed in the remainder of this section.

A new distribution is defined

$$f(\mathbf{R},\tau) = \psi_G(\mathbf{R})\psi(\mathbf{R},\tau) \quad , \tag{3.33}$$

which is also a solution of the Schrödinger equation when $\psi(\mathbf{R},\tau)$ is a solution. The equation for non-importance sampled DMC, equation 3.32, is consequently modified:

$$-\frac{\partial f(\mathbf{R},\tau)}{\partial \tau} = \left[\sum_{i=1}^{N} -\frac{1}{2}\nabla_i^2 f(\mathbf{R},\tau)\right] - \nabla . \left[\frac{\nabla\psi(\mathbf{R})}{\psi(\mathbf{R})} f(\mathbf{R},\tau)\right] + (E_L(\mathbf{R}) - E_T) f(\mathbf{R},\tau) \quad . \tag{3.34}$$

E_T is a "trial energy" introduced to maintain normalisation of the projected solution at large τ. The term

$$\mathbf{F}(\mathbf{R}) = \frac{\nabla\psi(\mathbf{R})}{\psi(\mathbf{R})} \quad , \tag{3.35}$$

is commonly referred to as the "quantum force" [29]. The local energy, $E_L(\mathbf{R})$, equation 3.17 is computed with respect to the guiding wavefunction $\psi_G(\mathbf{R})$.

The right hand side of the importance sampled DMC equation (equation 3.34) consists, from left to right, of diffusion, drift and rate terms. The problematic potential dependent rate term of the non-importance sampled method is replaced by a term dependent on the difference between the local energy of the guiding wavefunction and the trial energy. The trial energy is initially chosen to be the VMC energy of the guiding wavefunction, and is updated as the simulation progresses. Use of an optimised guiding function minimises the difference between the local and trial energies, and hence minimises fluctuations in the distribution f. A wavefunction optimised using VMC is ideal for this purpose, and in practice VMC provides the best method for obtaining wavefunctions that accurately approximate ground state wavefunctions locally. The guiding wavefunction may be also constructed to minimise the number of divergences in $E_L(\mathbf{R})$, unlike the non-importance sampled method where divergences in the coulomb interactions are always present.

3.4.3 Transformation to integral form

To be of use, the importance sampled DMC equation equation 3.34 must be transformed into a form suitable for Monte Carlo integration. The transformation is more complex than for VMC, which simply required the insertion of the factor $|\psi(\mathbf{R})|^2/\int|\psi(\mathbf{R})|^2 d\mathbf{R}$ into the conventional formulas for quantum mechanical expectation values.

A Green's function, $G(\mathbf{R}',\mathbf{R};\tau)$, that is a solution of equation 3.34 is desired, i.e, a spatial representation of the imaginary time propagator, $\exp[-\tau(\hat{H} - E_T)]$:

$$f(\mathbf{R}', \tau_0 + \tau) = \int G(\mathbf{R}', \mathbf{R}; \tau) f(\mathbf{R}, \tau_0) d\mathbf{R} \tag{3.36}$$

The Green's function may be approximated to second order in τ by factorising the propagator into branching and diffusion parts: [23, 28]

$$G(\mathbf{R}', \mathbf{R}; \tau) \approx \tilde{G}(\mathbf{R}', \mathbf{R}; \tau) \equiv \tilde{G}_b \tilde{G}_d \quad . \tag{3.37}$$

The branching part of the Green's function, $\tilde{G}_b$, is a function of the "surplus local energy":

$$\tilde{G}_b(\mathbf{R}', \mathbf{R}; \tau) = \exp[(-\frac{1}{2}(E_L(\mathbf{R}') + E_L(\mathbf{R})) - E_T)\tau] \quad . \tag{3.38}$$

The diffusion part is obtained by assuming that the quantum force, $\mathbf{F}$, is constant over the move $\mathbf{R} \rightarrow \mathbf{R}'$:

$$\tilde{G}_d(\mathbf{R}', \mathbf{R}; \tau) = (2\pi\tau)^{-3N/2} \exp[-(\mathbf{R} - \mathbf{R}' - \tau\mathbf{F}(\mathbf{R}')/2)^2/2\tau] \quad . \tag{3.39}$$

Although accurate to second order in τ, this approximation is significant in all-electron calculations, where $\mathbf{F}$ varies rapidly in core regions. Additionally, this component of the short-time approximation Green's function, $\tilde{G}$, is not Hermitian. "Detailed balance" is imposed with the conventional Metropolis probability

$$p = \min(1, W(\mathbf{R}', \mathbf{R}; \tau)) \quad , \tag{3.40}$$

where

$$W(\mathbf{R}', \mathbf{R}; \tau) = \frac{|\Psi_G(\mathbf{R}')|^2 \tilde{G}(\mathbf{R}', \mathbf{R}; \tau)}{|\Psi_G(\mathbf{R})|^2 \tilde{G}(\mathbf{R}, \mathbf{R}'; \tau)} \quad , \tag{3.41}$$

3.4.4 The fixed-node approximation

The DMC method discussed so far lacks a constraint crucial for the simulation of fermionic systems. In general, the ground state wavefunction of a fermionic system has nodes and consequently regions of positive and negative sign. These regions are essential for a wavefunction to be anti-symmetric with respect to the interchange of any two fermions. The methods discussed so far rely on a probability or particle distribution which by definition is positive definite. If the constraint of antisymmetry is not imposed, the application of the DMC method results in a bosonic solution of lower energy than the true fermionic ground state energy being obtained. The most successful attempt to address to this "nodal problem" is the fixed-node approximation. It is used in all of the DMC calculations presented in this thesis.

The fixed-node approximation, due to Anderson, [30] was the first solution of the nodal problem. The approximation consists of imposing the nodes of a reference wavefunction. This approximation is variational: the fixed-node DMC energy is an upper bound to the exact ground state energy. [31] In Anderson's original implementation, walkers attempting to cross nodes were deleted. This has been shown to introduce a bias that is linear in the time step, [29] and all current implementations reject moves that cross nodes. A bias of second order in the time step results, which is the same order as the short-time approximation to the Green's function.

The accuracy of the fixed-node approximation is wholly dependent upon the nodes of the reference wavefunction. In practice the nodes of an HF or DFT based wavefunction are found to be very accurate, giving energies well below those of the best VMC wavefunctions. This is fortunate as currently there are no general methods for systematically improving the DMC energy that scale polynomially with the size of the system.

3.4.5 Improving the fixed–node energy

Two methods for *systematically* improving the fixed-node energy exist:

- The release-node method [31] is a simple refinement of the fixed-node method. An antisymmetric wavefunction is obtained by averaging the difference between populations of negative and positive walkers, initially "equilibrated" within the fixed-node approximation. Moves that cross node are then permitted and the asymptotic release-node energy obtained. This method is impractical for more than a few electrons as the populations and hence statistical errors grow exponentially. The fixed-nodes must lie close to the exact nodes and the system must rapidly relax to the exact fermionic solution for the method to be practical, even in small systems. The most successful application of this technique is for the electron gas. [13]

- The nodes of CI-based wavefunctions can be expected to more accurately represent the exact nodes than HF-based wavefunctions, as correlation effects are partly included and a CI wavefunction is a better approximation to the exact wavefunction. Therefore a systematic, but inefficient, method of improving the nodes of the reference wavefunction is to perform large CI calculations and use the resulting wavefunction (with many determinants) in DMC. This method is impractical for systems of moderate size for the same reasons that the CI approach is in general impractical (see section 2.4.1). For small systems this approach is practical, but has not been extensively employed in the literature.

A third method of improving the nodes exists. Calculations by Kwon [32, 33] have shown that small improvements to DMC energies may be obtained for the 2D and 3D electron gases, at high densities, by replacing the positions of the electrons in the wavefunctions by "quasiparticle coordinates" [34] parameterised in terms of the positions of (in principle) all the electrons:

$$\mathbf{x}_i = \mathbf{r}_i + \sum_{j \neq i}^{N} \eta(r_{ij})(\mathbf{r}_i - \mathbf{r}_j) \quad . \tag{3.42}$$

η is a correlation or *backflow* [34] function containing variable parameters. Presently it is unclear whether the use of quasiparticle coordinates will be of use in real electronic systems. These methods carry considerable overhead, as the full wavefunction must be calculated when a single electron is moved. Although these methods may prove useful in small systems, the use of quasiparticle coordinates in large systems is likely to restricted, as the cost of moving all of the electrons is now $O(N^4)$ due to the additional cost of wavefunction recalculation (see chapter 4).

Additionally, these methods are perturbative in the sense that they modify an underlying reference nodal set. Although the use of quasiparticle coordinates may result in small improvements in real electronic systems, where good quality nodal sets are already available, their use cannot be considered a *solution* to the nodal problem. They are however, one of the few ideas present in the literature that scales reasonably well with system size.

3.4.6 The DMC algorithm

The DMC algorithm used for most of the calculations in this thesis is given below. It is similar to those of Refs. [23, 28, 29]. As with VMC there are distinct equilibration and accumulation phases, but both are almost identical as energies are required at every move in the calculation in the evaluation of the configuration branching factor.

$\Psi(\mathbf{r}')$ is used to indicate the wavefunction with a single electron moved to point $\mathbf{r}'$ and all other electrons unmoved:

1. Initialise an ensemble of N_c configurations, which should be uncorrelated and distributed according to the probability density of the guiding function $|\Psi_G|^2$. Initialise the trial energy E_T to the average VMC energy of the ensemble.

2. For every configuration j:

   (a) For every electron i in the current configuration:

      i. Propose a move according to

$$\mathbf{r}'_i = \mathbf{r}_i + \tau \mathbf{F}(\mathbf{r}_i) + \eta \quad , \tag{3.43}$$

where $\mathbf{F}$ is the "quantum force" of equation 3.35, and η is a Gaussian random vector of mean zero and variance τ per Cartesian component.

ii. Apply the fixed-node approximation, checking that $\Psi_G(\mathbf{r}')$ is of the same sign as $\Psi_G(\mathbf{r})$. If a node has been crossed, reject the move of electron i and consider the next electron.

iii. Compute the weight for the move

$$W(\mathbf{r}', \mathbf{r}) = \frac{|\Psi_G(\mathbf{r}')|^2 \tilde{G}(\mathbf{r}', \mathbf{r}; \tau)}{|\Psi_G(\mathbf{r})|^2 \tilde{G}(\mathbf{r}, \mathbf{r}'; \tau)} \quad , \tag{3.44}$$

where $\tilde{G}$ is the short-time approximation to the Green's function (section 3.4.3).

iv. Accept the move according to the Metropolis probability $\min(1, W(\mathbf{r}', \mathbf{r}))$

(b) Compute the branching factor for the moved configuration j:

$$P_B = \exp\left(-\tau(\frac{1}{2}\left[E_L(\mathbf{R}') + E_L(\mathbf{R})\right] - E_T)\right) \tag{3.45}$$

(c) Accumulate the energy and any observables weighted by the branching factor P_B

(d) Make $\mathrm{int}(P_B + u)$ copies of the configuration, where u is a uniformly distribution random number of $[0, 1]$.

3. Repeat the configuration moves for a number of steps (a block), typically $O(100 - 1000)$.

4. Update the trial energy E_T, to bring it closer to that of the current ensemble, by using the average energy for the last block.

5. Renormalise the number of walkers in the ensemble to the target number N_c by randomly creating or deleting walkers.

6. Repeat until sufficient data is obtained.

During equilibration, a slight variant of the above algorithm is used, where the trial energy is updated more frequently. While equilibrating, there is tendency for the number of walkers to increase exponentially as the excited eigenstates in the trial wavefunction are propagated out. The trial energy is replaced by a reference energy updated after every move of all configurations:

$$E_{\mathrm{REF}} = E_T - \frac{C_{\mathrm{EREF}}}{\tau} \ln\left(\frac{N_w}{N_c}\right) \quad , \tag{3.46}$$

where C_{EREF} is a system dependent parameter set to control the fluctuations in the current number of configurations N_w compared with the target number of configurations N_c.

Several improvements to the above algorithm have been proposed. [29] These modify the Green's function where the short time approximation is expected to be least accurate (adjacent to atomic cores and near nodes), and also take account of the modified timestep due to the rejection of some of the single electron moves. These improvements give a significant reduction in the timestep error in all electron calculations. In pseudopotential calculations the improvements are much smaller. Small efficiency improvements are possible by assigning weights to each walker, and only branching or deleting walkers when the weights extend beyond a defined range. Provided sufficiently small timesteps are used with sufficiently accurate wavefunctions, these different algorithms give the same ground state energy.

3.5 Summary

The VMC and DMC methods have been described.

The VMC method provides a way to obtain accurate many–body wavefunctions for electronic systems and to statistically evaluate most observables using these wavefunctions. By computing all integral statistically, potentially any form of wavefunction is usable.

The variance of the local energy of the wavefunction is a measure of its accuracy. However, although the variational principle guarantees that energies are variational, no guarantees exist for other observables. The VMC method provides no direct insight into the parameterisation of a wavefunction, so that wavefunctions must be designed using physical insight. Hence although precise measurements of many–body observables is possible, there is only limited confidence in their accuracy for quantities other than the total energy.

The fixed-node DMC method is a near-exact method for finding the fermionic ground state of a Hamiltonian. The key approximation in the fixed-node DMC method is the use of an approximate nodal surface from a guiding wavefunction. In practice, the nodes of HF or DFT wavefunctions are very accurate. Good quality guiding wavefunctions are required to perform DMC calculations and optimised VMC wavefunction are sufficient.

DMC and VMC may be seen as complementing each other: DMC providing a very reliable estimate of the ground state energy, and the cheaper VMC potentially providing analytic many–body wavefunctions of similar accuracy.

The VMC and DMC methods are elegant, potentially very powerful methods for solving the Schrödinger equation. They must be judged by their performance in actual calculations on realistic systems: the methods must be sufficiently accurate and computationally affordable in systems meriting high accuracy investigation or in systems not readily accessed by other methods. Detailed descriptions of their implementation are described in the next chapter, and applications of the methods are presented in the following chapters.

Chapter 4

Implementation

4.1 Introduction

In this chapter, the practical realisation of the VMC and DMC methods is described. Although the theoretical presentation of QMC methods in chapter 3 describes the methods and their physics in mathematical terms, further developments are required in order to perform accurate and efficient calculations using reasonable computer power.

Solid-state and quantum chemical investigations require convenient parameterised forms for trial functions that are suitable for optimisation. Forms of trial functions are briefly reviewed and a practical scheme for wavefunction optimisation is presented.[1] The incorporation of pseudopotentials into the many-body framework and the use of the supercell approximation for extended systems is described. Data from actual calculations are given to help clarify and further illustrate the methods.

The details of implementation, although many in number, are important for the later applications in this thesis. In particular, efficient algorithms and implementation is essential for the calculations in chapters 7 and 8, where systems involving 216 and up to 128 valence electrons are simulated. At the time of writing, these calculations could be considered large even if electron-correlations were treated at a *density functional* level and are sufficiently large that quantum chemical methods cannot be applied.

4.2 History and background

At the time of writing, the uptake and use of QMC methods for electronic structure is quite limited. The reasons for the limited uptake provide a useful insight into the technical challenges of QMC methods in their present form.

[1] This topic is studied further in chapter 5.

Since Ceperley and Alder's QMC calculations for the Homogeneous Electron Gas (HEG) [13] in 1980 there has been a gradual increase in the use of QMC methods. [24] Successful application of QMC methods to inhomogeneous systems - those including atoms - is considerably more difficult than for the HEG.

The computational cost of QMC methods scale with the third power of number of electrons in the system. However, in practice it has been established that the *prefactor* for realistic calculations is very large. Hence, the barrier to performing calculations has been the large prefactor or *intrinsic variance* of the wavefunctions used in the calculation as well as the number of particles (electrons). The development of improved, more efficient wavefunction forms and methods for their optimisation has led to a considerable reduction in the prefactor.

A hidden scaling of QMC methods is their scaling with the atomic number, Z, of the atoms included in the simulation. It has been argued [23, 35, 36] that the scaling of computer time, T, with Z is,

$$T \sim Z^{5.5-6.5} \quad . \tag{4.1}$$

This scaling is severe, and results from the small size of the atomic cores requiring a different sampling (such as a different timestep) than valence regions. A further limitation results from the large kinetic energies and strong potentials in core regions. These can result in large fluctuations in the local energy and it is difficult to design trial functions that significantly decrease the fluctuations. The wavefunctions in section 4.3 are able to reduce but not eliminate the fluctuations.

Fortunately, for most properties, the atomic cores are almost unaffected by the valence (outermost) electrons and their environment. This observation can be used to remove the core electrons from the simulation, greatly reducing the problems of time and energy scales. The removal is achieved by the incorporation of a pseudopotential. This procedure, which is also common in DFT and quantum chemical calculations,[2] is described in section 4.5.

[2]Even within DFT and quantum chemical methods, the use of pseudopotentials is often unavoidable. Pseudopotentials allow the use of far smoother wavefunctions, greatly reducing the number of basis functions required. In the case of a plane-wave basis, most commonly employed in solid-state calculations, the use of pseudopotentials is almost essential. A hidden pseudopotential-like approximation is frequently made in quantum chemical calculations: that of neglecting selected core molecular orbitals from correlation calculations. This "frozen-core" approximation is almost essential for large-scale quantum chemical calculations.

4.3 Wavefunctions

4.3.1 Form of trial wavefunction

The trial wavefunctions used in this thesis all take the form of a sum of products of Slater determinants multiplied by a correlating Jastrow factor:

$$\Psi(\mathbf{R}) = \sum_n a_n D_n^{\uparrow}(\mathbf{r}_1^{\uparrow}, \ldots, \mathbf{r}_N^{\uparrow}) D_n^{\downarrow}(\mathbf{r}_1^{\downarrow}, \ldots, \mathbf{r}_N^{\downarrow}) \exp\left[J\right] \quad . \tag{4.2}$$

The general trial wavefunction, Ψ, consists of a sum of products of up- and down-spin determinants. Separate determinants are used for the $N^{\uparrow}$ up-spin and $N^{\downarrow}$ down-spin electrons, consisting of sets of single-particle orbitals:

$$D_n^{\uparrow} = \begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_1(\mathbf{r}_2) & \cdots & \phi_1(\mathbf{r}_{N^{\uparrow}}) \\ \phi_2(\mathbf{r}_1) & \phi_2(\mathbf{r}_2) & \cdots & \phi_2(\mathbf{r}_{N^{\uparrow}}) \\ \vdots & \vdots & & \\ \phi_{N^{\uparrow}}(\mathbf{r}_1) & \phi_{N^{\uparrow}}(\mathbf{r}_2) & \cdots & \phi_{N^{\uparrow}}(\mathbf{r}_{N^{\uparrow}}) \end{vmatrix} \quad . \tag{4.3}$$

The use of separate determinants for differing spins results in a trial function that is not antisymmetric with respect to interchange of opposite spin electrons, but gives the correct expectation values for spin-independent operators. [23] The Jastrow factor, J, in the most general case, is a function of all electron and atom positions. Jastrow functions are discussed in the next section.

The determinantal part of the wavefunction is usually obtained from a self-consistent method such as LDA-DFT or HF. [37,38] For most of the calculations in this thesis a single product of up and down-spin determinants of orbitals, ϕ, is used to form the wavefunctions. In principle, configuration-interaction or multi-configuration calculations can be used to provide multiple determinants, including the coefficients a_n,[3] but many applications of QMC have demonstrated that a high accuracy may usually be obtained using only a single determinant.

The orbitals, ϕ, are expanded in a computationally convenient basis-set. In principle the orbitals could be expanded in a basis that is optimal for QMC. The key criterion would be the speed of evaluation of the value and first and second derivatives of the orbitals at arbitrary points in space. However, to be obtainable from within a self-consistent scheme

[3] Multiple determinants are particularly effective in systems with small gaps. For large gap systems, multiple determinants provide only a small (inefficient) improvement in the wavefunction. In the Be atom, a trial function consisting of ≈ 0.85 of the $1s^2 2s^2$ ground state and ≈ 0.15 of the low lying $1s^2 2p^2$ excited states and optimised Jastrow factor is found to obtain significantly more correlation energy ($\approx 20\%$) than a wavefunction consisting of the ground state determinant and an optimised Jastrow factor. In chapter 8, a 43 determinant wavefunction is used for a C_{26} ring.

the basis must be convenient for solving the single-particle equations. Consequently, radial numerical grids, Gaussian functions [39] and plane-waves [40] are typically used for atomic, molecular and continuum systems respectively.[4]

A further consideration is that the orbitals in the Slater determinants must be purely real functions. In supercell (periodic boundary condition) calculations, this is achieved by choosing sets of bloch orbitals such that simple linear combinations of the orbitals yield real functions.

4.3.2 Form of Jastrow factor

The Jastrow factor [22] is the primary method for including electron correlation beyond the mean-field level in QMC simulations. In VMC it is the only method. Although a Jastrow factor is not in principle required for DMC calculations, a HF wavefunction is usually insufficiently accurate for use as a guiding wavefunction. A good quality Jastrow factor should incorporate the physics of electron correlation in a compact and rapidly computable form that is convenient for optimisation. The Jastrow factor is therefore one of the keys to efficient QMC simulations.

The number of forms of Jastrow factor that have enjoyed repeated use in electronic structure applications is limited. In this section, three forms are presented, two of which are used in subsequent calculations. A thorough study of the efficiency of different forms would be difficult to produce due to the computational cost involved. Alexander and Coldwell [41] reviewed 118 trial wavefunction forms for the He, Li and Be atoms, but did not consider other species, or molecular and continuum systems.

Although the quality of a Jastrow function can be judged by the VMC energies obtained for candidate systems, compromises must inevitably be made. Benchmark accuracy Jastrow factors for molecular and continuum systems typically consist of very large numbers of terms and parameters. The increase in accuracy per additional term (of a given kind) inevitably reduces. Consequently, a wavefunction obtaining 80% of the correlation energy and 30 parameters may be preferable to one obtaining 90% of the correlation energy with 200 parameters. For use as a guiding DMC wavefunction, the simpler wavefunction would be preferable due to the lower computational cost. For VMC calculations, the preference would depend on the specific application and desired final accuracy.

[4] Due to basis set incompleteness, a subsequent projection onto a basis more suited to QMC applications is numerically treacherous. However, as we are only dealing with *trial* wavefunctions, this approach remains a potential area of future research.

Mitáš' approach

Mitáš *et al.* have successfully optimised wavefunctions for atomic systems, [42] molecular carbon [43, 44] and silicon [45] systems, using a Jastrow function of the following form:

$$J = \sum_I \sum_i^N \sum_{j \neq i}^N u(r_{iI}, r_{jI}, rij) \quad , \tag{4.4}$$

$$u(r_{iI}, r_{jI}, r_{ij}) = -\frac{c}{\gamma} e^{-\gamma r_{ij}} + \sum_{k<l,m} c_{klm}[a_k(r_{iI})a_l(r_{jI}) + a_k(r_{jI})a_l(r_{iI})]b_m(r_{ij}) \quad , \tag{4.5}$$

where

$$a_k(r) = \left(\frac{\alpha_k r}{1+\alpha_k r}\right)^2 \ , \ \alpha_k = \frac{\alpha_0}{2^{k-1}} \ , \ k > 0 \tag{4.6}$$

$$b_m(r) = \left(\frac{\beta_m r}{1+\beta_m r}\right)^2 \ , \ \beta_m = \frac{\beta_0}{2^{m-1}} \ , \ m > 0 \quad , \tag{4.7}$$

and r_{iI} is the separation between the i^{th} of N electrons and the I^{th} ion, and $a_0(r) = b_0(r) = 1$. This form can satisfy the electron-electron cusp condition and has no electron-nucleus cusp. This is desired as the orbitals in the determinant satisfy, or can be made to satisfy, the electron-nuclear cusp.

This functional form has yielded wavefunctions of sufficient accuracy to perform DMC calculations in several systems, but has not been systematically tested for accuracy and convergence of VMC energies.

Although successfully used in several calculations, the functional form has two undesirable features. The correlation function, u, is long ranged implying that the sum over electrons and ions is not readily cutoff and will increase in cost with system size. For certain values of γ, the long range of the exponential term requires the use of Ewald summation (section 4.6.1) in supercell calculations. As this sum involves all pairs of electrons, this operation would become costly.

Additionally, the function is not linear in the α_k parameters, which is desirable to avoid full recalculation of the function as the parameters are changed during optimisation. This increases the computational cost of the optimisation (see section 4.7.1).

Williamson's approach

Williamson *et al.* [46, 47] developed a homogeneous Jastrow function designed for solid-state periodic boundary condition applications, which avoids potentially costly sums. This was used in conjunction with a one-body "χ" term that had been used in Fahy's early bulk calculations. [26]

$$J = \sum_{i=1}^{N} \chi(\mathbf{r}_i) - \sum_{i=1}^{N} \sum_{j \neq i}^{N} u(r_{ij}) \quad , \tag{4.8}$$

where $\mathbf{r}_i$ is the absolute position of electron i, and r_{ij} is a minimum-image [17] electron-electron distance. The one-body term consists of a plane-wave expansion

$$\chi(\mathbf{r}_i) = \sum_{\mathbf{G}} \chi(\mathbf{G}) e^{i\mathbf{G}.\mathbf{r}_i} \quad , \tag{4.9}$$

where reciprocal lattice vectors are denoted by $\mathbf{G}$. The number of independent coefficients $\chi(\mathbf{G})$ is reduced by imposing symmetry.

The electron-electron dependent part of the Jastrow function consists of cusp-satisfying part and a separate power expansion, both of which are constrained to be short ranged. For the cusp-satisfying part, the following form was chosen:

$$u_0(r) = \frac{A}{r} \left(1 - \exp(-\frac{r}{F})\right) \exp\left(-\frac{r^2}{L_0^2}\right) \quad , \tag{4.10}$$

where F is chosen so that the cusp condition is obeyed and L_0 is chosen so that u_0 is effectively zero at the Wigner-Seitz cell edge in supercell calculations. A is a variational parameter, and L_0 is fixed at approximately 30% of the Wigner-Seitz cell radius. A power expansion in Chebyshev polynomials, over a pre-determined range L, was used in a second term:

$$\begin{aligned} f(r) &= B(\tfrac{L}{2} + r)(L - r)^2 && 0 \leq r \leq L \\ &+ r^2 (L - r)^2 \textstyle\sum_{l=0}^{M} c_l T_l(\overline{r}) && \\ &= 0 && r > L \end{aligned} \quad , \tag{4.11}$$

where B and c_l are variational coefficients and T_l denotes the lth Chebyshev polynomial. The expansion range, $(0, L)$, of the correlation function is rescaled to the orthogonality interval of the Chebyshev polynomials, $[-1, 1]$,

$$\overline{r} = \frac{2r - L}{L} \quad . \tag{4.12}$$

The parameters B, c_l are different for spin parallel and anti-parallel electron pairs, to model the additional Pauli repulsion experienced by parallel spin electrons. These functions satisfy the electron-electron cusp condition and are continuous in value and first derivative. The expansion range, L, is chosen to be less than or equal to the Wigner-Seitz radius in supercell calculations avoiding the need to perform an Ewald sum. A further advantage of the Chebyshev expansion is that unlike the terms in a simple power expansion, the terms carry equal weight over their range and consequently exhibit greater stability during optimisation.

In applications to bulk carbon and silicon, wavefunctions obtaining 80−85% of the fixed-node DMC correlation energy were obtained. The high accuracy of these wavefunctions indicates that the limited range (form) of the correlation functions does not severely limit the accuracies obtained.

Atom centred approach

For the calculations in this thesis, a development of the above approach has been used. The plane-wave χ function was replaced by atom-centred functions, expanded over a short range in a complete polynomial. By extending the idea of short ranged power expansions to atom centred functions, partially *transferable* correlation functions result. This is distinct from the plane-wave χ function which is not easily transfered to other systems. The high accuracy of the previous functions were further increased by adding parameterised terms dependent on electron-electron-ion distances. These are present in Mitáš' function and are important in all-electron calculations and systems with tightly bound electrons.

A general power expansion, in r_i and r_{ij}, about an atom over a range L must satisfy similar conditions to Williamson's function, above: the function must go to zero at L and be continuous in value and first derivative. The atom centred power expansion in electron-ion distance, r_i and electron-electron distance r_{ij} contains 5 independent series: [48, 49]

$$
\begin{aligned}
S_1 &= (r_i - L)^2 r_i^2 \sum_{l\geq 0} \alpha_l T_l(\bar{r}_i) + B(r_i - L)^2 \left(\frac{L}{2} + r_i\right) \qquad (4.13)\\
S_2 &= (r_{ij} - 2L)^2\, r_{ij}^2 \sum_{n\geq 0} \beta_n T_n(\bar{r}_{ij}) + B'\,(r_{ij} - 2L)^2 \left(\frac{2L}{2} + r_{ij}\right)\\
S_3 &= (r_i - L)^2 (r_j - L)^2 r_i^2 r_j^2 \sum_{l\geq 0}\sum_{m\geq 0} \gamma_{lm} T_l(\bar{r}_i) T_m(\bar{r}_j)\\
S_4 &= (r_i - L)^2\,(r_{ij} - 2L)^2\, r_i^2 r_{ij}^2 \sum_{l\geq 0}\sum_{n\geq 0} \epsilon_{ln} T_l(\bar{r}_i) T_n(\bar{r}_{ij})\\
S_5 &= (r_i - L)^2 (r_j - L)^2 r_i^2 r_j^2 r_{ij}^2 \sum_{l\geq 0}\sum_{m\geq 0}\sum_{n\geq 0} \omega_{lmn} T_l(\bar{r}_i) T_m(\bar{r}_j) T_n(\bar{r}_{ij}) \quad ,
\end{aligned}
$$

where it is implicitly assumed that terms are zero for $r_i > L$ or $r_{ij} > 2L$ and

$$
\overline{r}_{ij} = \frac{r_{ij} - L}{L} \quad . \qquad (4.14)
$$

The overall Jastrow factor consists of sets of the above functions supplemented by the electron-electron cusp-satisfying u_0 term, equation 4.10, and the homogeneous r_{ij}-dependent $f(r_{ij})$ term, equation 4.11. This is shown schematically in figure 4.1.

For the main applications in this thesis, only the terms S_1 were used as sufficient

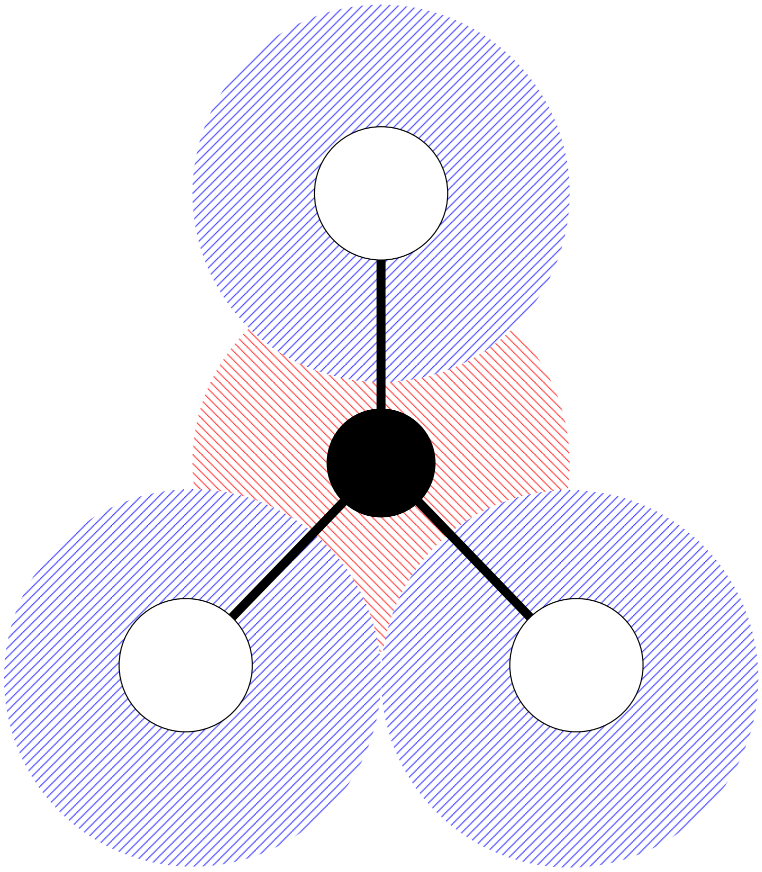

Figure 4.1: Example Jastrow factor parameterisation for CO_3 (C_{2v} symmetry). Overlapping atom-centred functions are placed on the carbon (filled circle) and oxygen (empty circle) atoms and describe electron-ion, electron-electron and electron-electron-ion correlations within a specified range (shading). The functions for the oxygen atoms are identical by symmetry. An additional longer-ranged electron-electron term (not shown) describes r_{ij}-dependent correlations near atoms and also in the valence regions.

accuracy was obtained and the full series had yet to be implemented. In tests on the ground state of an all-electron oxygen atom, using exact HF orbitals, 89% of the experimental correlation energy was obtained using all the terms, approximately 20% more than a wavefunction using only the S_1 terms. This accuracy is similar to that obtained in *pseudopotential* calculations of oxygen atoms, demonstrating the importance of the additional correlation terms when treating core electrons.

4.4 Algorithms and practicalities

In this section, the numerical implementation of QMC methods using wavefunctions based on Slater determinants is discussed. Although critical for any QMC simulation, this discussion has been delayed until now because the implementation is relatively straightforward. An implementation of VMC and DMC algorithms for molecular and solid-state systems is similar in complexity to a conventional quantum chemical HF or solid–state DFT code. However, although many of the quantities involved are the same the actual implementation is very different. For example, no integrals are calculated analytically and all evaluations are performed in real-space.

4.4.1 Calculation and update of the Slater determinants

The calculation and update of the Slater determinants in trial wavefunctions gives rise to the $O(N^3)$ scaling of QMC methods. Provided the efficient rank-1 update procedure described below is adopted, for molecular and solid–state systems of between approximately 10^1 and 10^3 electrons, the update does not require a substantial proportion of processor time.[5] For sufficiently large systems the update will dominate. Note that the $O(N^3)$ scaling is that of moving all of the electrons with respect to system size, and not that of obtaining a fixed statistical accuracy with respect system size.

Two quantities are required during a QMC calculation: the ratio of the old to new determinant when a single electron is moved, and the updated determinant when a single electron move is accepted. Although similar, the first of these operations is a $O(N)$ process, the second $O(N^2)$. There are $O(N)$ accepted single electrons moves per "configuration move" giving rise to the $O(N^3)$ scaling of QMC methods.

It has been shown [23, 26, 50] that it is possible to work entirely with the matrix of cofactors of the Slater determinant. This also applies to multi–determinant wavefunctions.

The Slater determinant has elements

$$D_{ji} = \phi_j(\mathbf{r}_i) \quad . \tag{4.15}$$

When electron i is moved from position $\mathbf{r}_i$ to $\mathbf{r}'_i$ only the elements of the determinant D_{ji}, where $j = 1, N$ are changed. (As a single electron is moved, only the respective column of the Slater determinant of the same spin changes.)

The relation

$$|\mathbf{D}|\mathbf{I} = \mathbf{D}^T\mathbf{D}^C \quad , \tag{4.16}$$

[5] The proportion of time spent is very case–dependent, but proportions between 0.5% and 25% were observed during the calculations in this thesis.

states that the value of the determinant, D, can be obtained from the transpose of the matrix $\mathbf{D}$ and its matrix of cofactors $\mathbf{D}^C$. By definition,

$$(\mathbf{D}^T)^{-1} = \frac{1}{D}\mathbf{D}^C = \overline{\mathbf{D}} \quad , \tag{4.17}$$

so that

$$D^C_{ji} = |\mathbf{D}|\overline{\mathbf{D}}_{ji} \quad . \tag{4.18}$$

From equation 4.16, we can write

$$|\mathbf{D}| = \sum_{j=1}^{N} D_{ij} D^C_{ji} \tag{4.19}$$

and

$$|\mathbf{D}'| = \sum_{j=1}^{N} D'_{ij} D^C_{ji} \quad , \tag{4.20}$$

where the primes indicate the matrices after the move. When electron i is moved, only the i^{th} column of the matrix $\mathbf{D}$ changes, and the i^{th} column of the matrix of cofactors $\mathbf{D}^C$ is unchanged. Hence,

$$q_i = \left|\frac{\mathbf{D}'}{\mathbf{D}}\right| = \sum_{j=1}^{N} D'_{ji}\overline{D}_{ji} \quad , \tag{4.21}$$

and the ratio of the old to new determinants is computed without reference to $\mathbf{D}'$ in $O(N)$ operations. The efficiency of this method is particularly important in the evaluation of the non-local pseudopotential energy, which requires this ratio to be evaluated at for each point on an integration grid (see section 4.5.1).

If the move of electron i to position $\mathbf{r}'$ is accepted, the matrix of cofactors is updated:

$$\overline{D}'_{jk} = \begin{cases} \frac{\overline{D}_{jk}}{q_i} & \text{for } k = i \\ \overline{D}_{jk} - \frac{\overline{D}_{ji}}{q_i}\left(\sum_{l=1}^{N} \overline{D}_{lk} D'_{li}\right) & \text{for } k \neq i \end{cases} \tag{4.22}$$

These rules may be derived by considering the change to the $\overline{D}$ matrices resulting from changing column i in the Slater determinant, [50]

$$\overline{D}'_{jk} = \overline{D}_{jk} + \Delta_{jk} \tag{4.23}$$

where

$$\Delta_{jk} = \delta_{ik}(\phi_j(\mathbf{r}') - \phi_j(\mathbf{r}) \quad . \tag{4.24}$$

4.4.2 Evaluation of the local energy

Evaluation of the local energy during a QMC calculation requires computation of the kinetic and potential energies. Calculation of the kinetic energy requires derivatives of the trial wavefunction. This is performed analytically and their calculation is presented below.

Calculation of the potential energies requires evaluation of the election-ion and electron-electron potentials. For pseudopotential calculations, the electron-ion potential is evaluated using the technique of section 4.5.1. The remaining Coulomb potentials are evaluated by direct summation in finite systems, and by Ewald summation (see section 4.6.1) in periodic systems.

4.4.3 Calculation of the kinetic energy

Calculation of the kinetic energy is performed by evaluating the estimator

$$\mathrm{KE} = \sum_{i=1}^{N} -\frac{1}{2}\frac{\nabla_i^2 \Psi}{\Psi} \quad , \tag{4.25}$$

where Ψ is the trial wavefunction. When averaged of the probability distribution $|\Psi^2|$, this estimator gives the exact kinetic energy of the trial wavefunction,

$$< \mathrm{KE} >= \sum_{i=1}^{N} -\frac{1}{2}\frac{< \Psi|\nabla_i^2|\Psi >}{< \Psi|\Psi >} \quad . \tag{4.26}$$

The kinetic energy contribution from a single electron, KE_i, is most conveniently evaluated by defining two additional quantities, T_i and F_i:

$$T_i = -\frac{1}{4}\nabla_i^2 \ln \Psi \tag{4.27}$$

and

$$F_i = \frac{1}{\sqrt{2}}\nabla_i \ln \Psi \quad . \tag{4.28}$$

Then the kinetic energy is given by

$$\mathrm{KE}_i = 2T_i - F_i^2 = -\frac{1}{2}\frac{\nabla_i^2 \Psi}{\Psi} \quad . \tag{4.29}$$

The use of logarithms is particularly convenient for Slater-Jastrow wavefunctions (section 4.3.1). For a single-determinant wavefunction, written as a product of spin up and down determinants,

$$\begin{aligned} \nabla_i \ln \Psi &= \nabla_i \ln D^{\uparrow} + \nabla_i \ln D^{\downarrow} + \nabla_i \ln[\exp(J)] \\ &= \frac{1}{D^{\uparrow}} \nabla_i D^{\uparrow} + \frac{1}{D^{\downarrow}} \nabla_i D^{\downarrow} + \nabla_i J \quad , \end{aligned} \tag{4.30}$$

and

$$\nabla_i^2 \ln \Psi = -\left(\frac{1}{D^{\uparrow}} \nabla_i D^{\uparrow}\right)^2 + \frac{1}{D^{\uparrow}} \nabla_i^2 D^{\uparrow} - \left(\frac{1}{D^{\downarrow}} \nabla_i D^{\downarrow}\right)^2 + \frac{1}{D^{\downarrow}} \nabla_i^2 D^{\downarrow} - \nabla_i^2 J \quad . \tag{4.31}$$

Averaging the quantities T_i and F_i^2 independently provides a useful check on the simulation. In VMC, Green's relation shows that

$$< \mathrm{KE}_i >=< T_i >=< F_i^2 > \quad , \tag{4.32}$$

provided all of configuration space, $\mathbf{R}$, has been sampled. [23, 26] The quantities T_i and F_i^2 usually have worse statistics than the kinetic energy and are not used to determine the kinetic energy directly.[6]

4.4.4 Calculation of the potential energy

Calculation of the potential energy terms in QMC simulations is straightforward. The electron-electron and electron-ion potential in periodic systems deserve special mention due to the use of the Ewald summation ([51] and section 4.6.1).

The Ewald sum consists of separate real and reciprocal space sums. For computational efficiency, an optimised Ewald sum is used in which only a single real space term is taken. The sum is therefore biased towards the reciprocal space part which is more rapidly evaluated. An optimised form for the real and reciprocal space potentials [52] minimises errors.

4.4.5 Parallelisation of VMC

Almost all of the calculations in this thesis were performed on parallel machines. Extension of the VMC algorithm of section 3.3.5 to parallel machines requires very few changes, as the algorithm is naturally parallel.

[6] In pseudopotential calculations, T_i and F_i^2 were found to have worse statistics than the kinetic energy. However, in all–electron VMC calculations of the oxygen atom and water monomer using exact and near-exact HF orbitals respectively, the kinetic energy was found to have worse statistical errors by a factor of approximately 2. No mention of this was found in the literature.

To parallelise the VMC algorithm it is sufficient to note that provided care is taken to ensure each calculation uses a different random number sequence, the algorithm can be parallelised by performing independent VMC calculations on each node of a parallel machine. Communication between the nodes is only necessary to obtain global averages and is therefore negligible: the parallel efficiency is essentially 100%, and the calculation can theoretically exploit any number of nodes without loss of efficiency.

4.4.6 Parallelisation of DMC

The DMC algorithm of section 3.4.6 requires more changes than the VMC algorithm to be parallelised, but very high parallel efficiencies are still obtainable.

During a DMC calculation an ensemble of walkers is propagated in imaginary time. It is natural to divide the walkers evenly amongst the nodes, so that communication is only strictly necessary at the end of each block, when the population is renormalised. However, this simple modification would result in a low parallel efficiency because the calculation would become poorly load balanced as the calculation progressed. The branching and deletion of walkers would result in some nodes having more walkers than others; at the end of the block many nodes would have to wait whilst others moved their remaining walkers. A load balanced DMC algorithm requires the even distribution of walkers over all nodes, and this is accomplished by redistributing the walkers after every move.

Therefore, in the parallel DMC algorithm, walkers are initially evenly distributed over all nodes. Each node is then responsible for propagating its walkers. After each "ensemble move", the walkers are redistributed as evenly as possible across all of the nodes.

The parallel efficiency is limited by two factors:

1. The number of walkers on each node.

2. The accuracy of the guiding wavefunction.

In general, the walkers are not evenly distributed between nodes and the first limitation results from the time lost while nodes complete moving all of their walkers. The efficiency of the algorithm is given by

$$\text{efficiency}\% = 100 \times \frac{N_w}{n_{\max} \times N_{\text{nodes}}} \quad , \tag{4.33}$$

where there are N_w walkers total, a maximum of $n_{\max}$ walkers on any node and N_{nodes} total. The efficiency is maximised by keeping the number of walkers well balanced and by having a large number of walkers per node.

The accuracy of the guiding wavefunction governs the fluctuations in the walker population. Hence improved guiding wavefunctions have the combined effect of increasing the statistical accuracy of the calculations and increasing the efficiency with which they are carried out. For the calculations in this thesis, parallel efficiencies of $90-95\%$ were typical.

4.5 Pseudopotentials

The inclusion of core electrons poses special problems in QMC calculations distinct from those encountered in other electronic structure methods. The differing timescale (or equivalently distance scale) on which the electrons move compared with valence electrons requires the use of special modified sampling schemes or the use of very small timesteps which reduce the efficiency of the simulation. Large ionic and kinetic energies in the core regions are a further problem. Although accurate wavefunctions may be constructed for low Z atoms,[7] for high Z atoms accurate wavefunctions have yet to be designed and obtained.

These two difficulties conspire to focus computational effort on the core electrons, forcing attention away from the chemically interacting and physically interesting valence and bonding regions. Fortunately for many properties of interest, and for many materials, the core remains almost independent of its environment and may be substituted by a pseudopotential with negligible loss of accuracy.[8] This replacement serves to reduce the effective Z of the atoms that must be dealt with. The Z-dependence in the scaling of QMC calculations is then determined by the number of valence electrons that must be included in order to obtain satisfactory results.

4.5.1 Non-local pseudopotentials

Pseudopotentials are used routinely in DFT and quantum chemical applications and consequently there exists a large literature and continued research effort on the subject. [54–61] Improvements to the pseudopotential approximation are not an aim of this work: accurate potentials exist for carbon and silicon, the main atomic species used. The pseudopotentials user were generated within LDA-DFT. Evaluation of the pseudopotentials requires techniques particular to QMC, and these are described in the remainder of the section.

[7] Arguably, sufficiently accurate wavefunctions may be constructed for all first row elements. Ref. [53] includes wavefunctions for first-row dimers obtaining 79% or greater of the experimentally determined correlation energy.

[8] In general, atoms with a large energetic separation of core and valence levels have cores that can be accurately represented by a pseudopotential. This is not always the case however: some third row elements maintain good core-valence separation, but in certain materials, such as Ga compounds, the occupied d-levels are found to significantly influence bonding. In these cases, additional electrons must be placed in the valence and a smaller core used.

The pseudopotentials modify the many–body Hamiltonian, replacing the electron-ion coulomb terms with

$$\hat{V}_I = \hat{V}_{\rm loc} + \hat{V}_{\rm nl} \quad , \tag{4.34}$$

where $\hat{V}_I$ signifies the electron-ion dependent terms of ion I, and $\hat{V}_{\rm loc}$ and $\hat{V}_{\rm nl}$ are respectively a long-ranged local potential and a short-ranged non-local potential. The two potentials are chosen within a pseudopotential construction scheme to accurately model the valence electrons. At long-range, typically of order 2-3 a.u., the local potential returns to the Coulomb tail.

The non-local potentials are written in terms of several short-ranged angular momentum dependent potentials. The operator $\hat{V}_{\rm nl}$ is given by these potentials multiplied by the appropriate projection operators, written as integrals over the angular interval Ω':

$$\hat{V}_{\rm nl} = \sum_l \sum_{m=-l}^{+l} v_l(r) Y_{lm}(\Omega) \int_{4\pi} Y^*_{lm}(\Omega') d\Omega' \tag{4.35}$$

The local component of the pseudopotential may be directly evaluated during a Monte Carlo simulation. In supercell simulations, Ewald's method (see section 4.6.1) is required to deal with the long-range of the potential, but this this not a significant complication. In VMC, the non-local potentials must be projected onto the trial function which requires statistical evaluation of the projection operators.

Evaluation of the non-local energy

Evaluation of the non-local projection operators, and hence the non-local energy cannot be performed analytically. For general correlated wavefunctions, the projection operators depend on all inter-particle distances and are therefore not amenable to analytic or numerical integration. Fahy [26, 38] developed an exact scheme for evaluating the non-local energy in VMC and first applied the technique in simulations of bulk carbon and silicon in the diamond structure. This scheme is now used almost exclusively, approximate evaluation methods having been discarded due to insufficient accuracy.

In the Fahy scheme, evaluation of $\hat{V}_{\rm nl}\psi/\psi$ is performed stochastically during a VMC simulation. The projection is written in terms of ratios of the trial function,

$$\frac{\hat{V}_{\rm nl}\psi}{\psi} = \sum_i^N \sum_l \frac{2l+1}{4\pi} v_l(r_i) \int_{4\pi} P_l[\cos(\theta'_i)] \frac{\psi(\mathbf{r}_1, \ldots, \mathbf{r}'_i, \ldots, \mathbf{r}_N)}{\psi(\mathbf{r}_1, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_N)} d\Omega'_i \quad , \tag{4.36}$$

where the sum i runs over all N electrons at distances r_i from the ion. The orientation of the coordinate system (which is arbitrary) is chosen for each electron so that the vector $\mathbf{r}_i$ lies along the z-direction, thus eliminating the m-dependence of the integrals. [26]

The angular integral over Ω' is performed stochastically during the simulation. Optimised integration grids have been developed to exactly integrate functions of up to a certain angular momentum. [62] The variance of the estimator for the non-local energy is chosen to optimise the balance of work spent evaluating the non-local pseudopotential with the work performing other parts of the simulation. The grids are very efficient, and in calculations on bulk silicon and carbon, taking multiple samples of a grid was always found to be less efficient than using a single larger grid. This result is expected to be general; the integrals are effectively of low dimensionality and grid-based integration methods should converge more rapidly than Monte Carlo methods. In practice, 6-12 points give sufficient accuracy for first and second row elements, [26, 63] where the character of the wavefunctions is dominated by s and p angular-momenta. 6 point grids are sufficient to exactly integrate d momenta, and 12 points grids, g momenta.

The locality approximation

The non-locality of accurate pseudopotentials is a significant problem in DMC calculations. An explicit form for the many-body wavefunction is not available in DMC and it has been shown (e.g. Ref. [64]) that the matrix elements of the non-local operator for imaginary time diffusion are negative. This creates a sign problem similar to the sign problem of fermion DMC. This problem has been avoided by replacing the non-local potential operator by an approximate *local* potential determined by evaluating the full non-local operator on the trial wavefunction, as in VMC. [65, 66] This approximation has been shown to converge quadratically to the exact energy as the trial wavefunction improves. [63] In practice this approximation is very good and is not overly sensitive to details of the trial function. The "locality approximation" results in a non-variational DMC energy.

4.5.2 Core polarisation potentials

Core polarisation potentials (CPPs) are a refinement of pseudopotential theory particularly designed for use in many–body calculations. The pseudopotentials most commonly used in many–body calculations are derived from mean field calculations. This represents an approximation at some level, and to consistently achieve an accuracy better than ~ 0.1 eV on atomic energy levels, some many-body effects must be incorporated into the pseudopotential.

Direct methods of generating pseudopotentials within QMC have met with little success due to the difficulties in performing accurate all-electron calculations of sufficient statistical accuracy. [67]. Methods based on configuration interaction calculations and quasiparticle calculations have met with most success. [60, 68–70]

The accuracy of pseudopotentials may be increased by taking into account the polarisability of the core, which is one of the most important effects excluded by a conventional rigid core approach. It is has been shown that this incorporates the leading terms of "core relaxation", the change that a core undergoes in different chemical environments (Ref. [71] and references within). The formulation of CPPs considers the polarisation of the atomic core by the valence electrons, within a point-dipole picture. To avoid divergences in the potential, the effect fields experienced by valence electrons are truncated close to the core. By parameterising the truncation, the correct binding energies of valence electrons is ensured. The scheme includes some of the core-core, core-valence and valence-valence correlation effects of an all-electron, many–body calculation. [68, 71]

The many-body Hamiltonian is initially modified to include additional CPP terms resulting from the polarisation induced electric fields [68]

$$\begin{aligned} \hat{H}_{\mathrm{CPP}} = \; - & \sum_J \frac{\alpha_J}{2} \left[\left| \sum_{J' \neq J} Z_{J'} \frac{\mathbf{R}_{J'} - \mathbf{R}_J}{|\mathbf{R}_{J'} - \mathbf{R}_J|^3} \right|^2 \right. \\ & - 2 \sum_{J' \neq J} \sum_i^N \frac{(\mathbf{R}_{J'} - \mathbf{R}_J).(\mathbf{r}_i - \mathbf{R}_J)}{|\mathbf{R}_{J'} - \mathbf{R}_J|^3 |\mathbf{r}_i - \mathbf{R}_J|^3} \\ & + \sum_i^N \frac{1}{|\mathbf{r}_i - \mathbf{R}_J|^4} \\ & \left. + \sum_i^N \sum_{i' \neq i}^N \frac{(\mathbf{r}_{i'} - \mathbf{R}_J).(\mathbf{r}_i - \mathbf{R}_J)}{|\mathbf{r}_{i'} - \mathbf{R}_J|^3 |\mathbf{r}_i - \mathbf{R}_J|^3} \right] \quad , \end{aligned} \tag{4.37}$$

where α is the core polarisability, and the sums run over all ions, J, and N electrons. The terms are due to ion-ion, ion-ion-electron, electron-ion and electron-electron-ion interactions respectively. The electron dependent terms are further parameterised, by means of a cutoff function, removing the r^{-4} divergences:

$$f(x) = (1 - e^{-x^2})^2 \quad , \tag{4.38}$$

where x is a rescaled coordinate. The Hamiltonian terms become

$$\begin{aligned} \hat{H}_{\mathrm{CPP}} = \; - & \sum_J \frac{\alpha_J}{2} \left[\left| \sum_{J' \neq J} Z_{J'} \frac{\mathbf{R}_{J'} - \mathbf{R}_J}{|\mathbf{R}_{J'} - \mathbf{R}_J|^3} \right|^2 \right. \\ & - 2 \sum_{J' \neq J} \sum_i^N f\left(\frac{|\mathbf{r}_i - \mathbf{R}_J|}{\bar{r}_{\mathrm{e-e}}} \right) \frac{(\mathbf{R}_{J'} - \mathbf{R}_J).(\mathbf{r}_i - \mathbf{R}_J)}{|\mathbf{R}_{J'} - \mathbf{R}_J|^3 |\mathbf{r}_i - \mathbf{R}_J|^3} \end{aligned} \tag{4.39}$$

IP	Expt	Rel-TM	HF	HF+CPP
1st	6.82	6.862(15)	6.693(13)	6.767(16)
2nd	13.58	12.347(8)	12.218(8)	13.639(9)
3rd	27.48	27.666(4)	28.288(4)	28.413(5)
4th	43.24	44.336(0)	44.774(0)	46.228(0)

Table 4.1: Ionisation potentials of Ti for several pseudopotentials computed using DMC. Energies are in eV and statistical error bars are given in brackets. Experimental data from [72]. Rel-TM denotes a relativistic pseudopotential constructed in the Troullier-Martins scheme, [58] HF and HF+CPP denote a bare HF pseudopotential and an HF pseudopotential with core-polarisation potential terms respectively. [68] The pseudopotentials were generated in the configuration suggested by G. B. Bachelet *et al.* [56] Data courtesy Y. Lee [61]

$$+\sum_{i}^{N}\frac{\sum_l f\left(\frac{|\mathbf{r}_i-\mathbf{R}_J|}{\bar{r}_l^J}\right)\hat{P}_l}{|\mathbf{r}_i-\mathbf{R}_J|^4}$$
$$\left.+\sum_{i}^{N}\sum_{i'\neq i}^{N} f\left(\frac{|\mathbf{r}_{i'}-\mathbf{R}_J|}{\bar{r}_{\mathrm{e-e}}}\right) f\left(\frac{|\mathbf{r}_{i}-\mathbf{R}_J|}{\bar{r}_{\mathrm{e-e}}}\right)\frac{(\mathbf{r}_{i'}-\mathbf{R}_J).(\mathbf{r}_i-\mathbf{R}_J)}{|\mathbf{r}_{i'}-\mathbf{R}_J|^3|\mathbf{r}_i-\mathbf{R}_J|^3}\right] \quad ,$$

where $\hat{P}_l$ is the projection operator for angular momentum l, and the $\bar{r}_l$ and $\bar{r}_{\mathrm{e-e}}$ are parameters obtained during construction of the CPP. [68]

These terms may be directly evaluated within VMC and DMC. The non-local projection operator, $\hat{P}_l$, is evaluated at the same time as non-local component of the full pseudopotential, avoiding costly evaluations of the many-body wavefunction.

Sample data comparing DMC results for the ionisation potentials of a Ti atom for different pseudopotentials [61] is given in Table 4.1.

4.6 Supercell calculations

The supercell method is the ubiquitous approach for the study of solid-state periodic boundary condition systems. In principle, a solid may also be approximated by means of a large cluster. For sufficiently large clusters, the quantum mechanics of the centre-most atoms approximate those in a solid. However, the size of cluster required to approximate a solid is large due to the dominance of surface over "bulk" atoms in small and medium sized clusters. This consideration applies to all electronic structure methods, and the best general solution is to use a limited number of repeats of a primitive cell combined with periodic boundary conditions, as illustrated in figure 4.2.

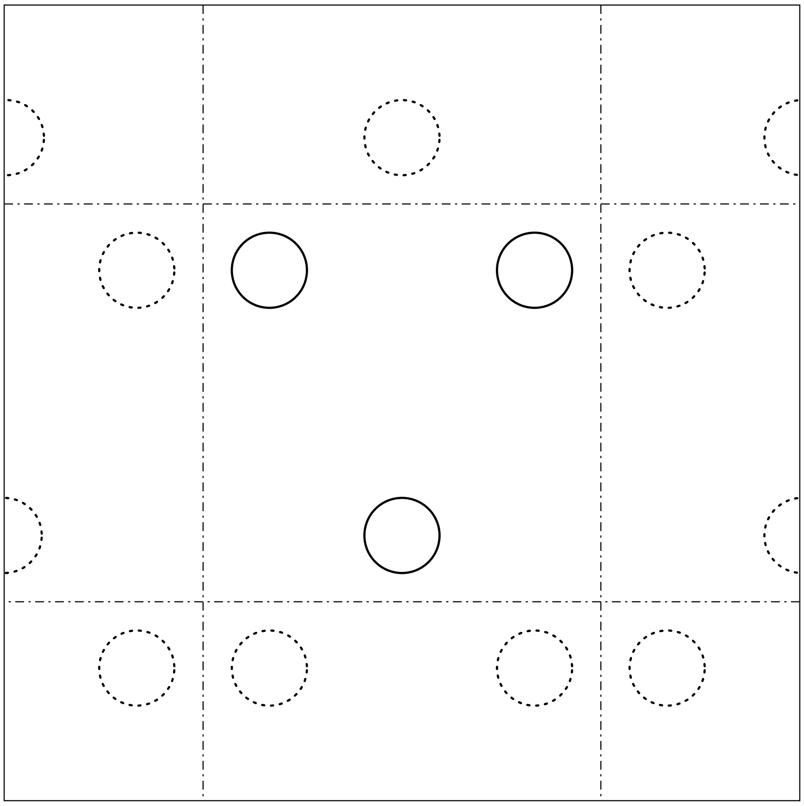

Figure 4.2: The supercell approach to periodic systems. Periodic boundary conditions are applied and all interactions are periodic with the periodicity of the supercell. The "minimum image" convention must be applied when particle-particle distances are computed: the minimum distance between particles can involve particle images in adjacent supercells.

In the supercell approach, [17, 73] an artificial periodicity is imposed on the simulation cell to better model the continuum properties of the system. Bloch's theorem may then be applied to the wavefunctions.

For perfect crystals, a limited number of repeats of the primitive cell are used. Periodic boundary conditions impose an artificial periodicity on interactions in the system. The supercell must be large enough that these spurious interactions are small. These finite size effects are examined in detail in chapter 6.

4.6.1 Long-range interactions

Within the supercell approximation, long-ranged interactions require special treatment. Any potential extending significantly beyond a single supercell, such as the coulomb potential, must correctly computed to account for the periodic boundary conditions. Neglect of the contributions to the potential from particles in neighbouring cells can lead to significant errors in energies, and the limiting behaviour of certain properties can be altered. Accurate summation of long-ranged interactions can be computationally intensive. Although approximate schemes have been developed for the rapid evaluation of long-ranged interactions, we are interested in a high final accuracy. Consequently, only accurate schemes for the evaluation of the coulomb interaction have not been used.

Ewald's method [17,51,74,75] is a near exact method for the evaluation of long-ranged interactions within periodic boundary conditions. The interaction terms forming the potential are rearranged into two series that are readily evaluated.

To illustrate the method, we consider a cubic supercell containing N particles of charge q_i. The interaction energy can be written as

$$E = \frac{1}{2} \sum_{\mathbf{n}}{}^{\prime} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{L}|} \quad , \tag{4.40}$$

where $\mathbf{L}$ are all supercell lattice vectors, and the prime indicates that for $\mathbf{L} = 0$ the self-interaction terms $i = j$ are excluded. This sum is divergent unless the system is overall charge neutral ($\sum q_i = 0$) [76] and care is required in its evaluation.

In the Ewald method, the point charge distribution is modified to give *two* convergent series, the sum of which gives the correct value for the unmodified distribution. The neutral Ewald "charge density" is given by a sum over the point charges plus a screening background.

$$\rho_{\mathrm{Ewald}}(\mathbf{r}) = \sum_{\mathbf{L}} \delta(\mathbf{r} - \mathbf{r}_i - \mathbf{L}) - \rho_{\mathrm{background}} \quad , \tag{4.41}$$

where the $\mathbf{r}_i$ are the positions the point charges and $\mathbf{L}$ is the complete set of supercell lattice vectors.

The Ewald potential is computed as a sum of two components. An array of Gaussians are used in both: in the first, the Gaussians cancel the background charge density, and in the second, the Gaussians are subtracted from the delta function charge distribution (see figure 4.3). The width of the Gaussian charges, μ, controls the rate of convergence of the two series and is usually optimised for each supercell.

$$\rho_1(\mathbf{r}) = \frac{1}{\mu\sqrt{\pi}} \sum_{\mathbf{L}} \exp(-(\mathbf{r} - \mathbf{r}_i - \mathbf{L})^2/\mu^2) - \rho_{\mathrm{background}} \tag{4.42}$$

and

$$\rho_2(\mathbf{r}) = \sum_{\mathbf{L}} \left[\delta(\mathbf{r} - \mathbf{r}_i - \mathbf{L}) - \frac{1}{\mu\sqrt{\pi}} \exp(-(\mathbf{r} - \mathbf{r}_i - \mathbf{L})^2/\mu^2) \right] \quad , \tag{4.43}$$

where μ is the width of the Gaussian charges. The potential due to ρ_1 is most easily

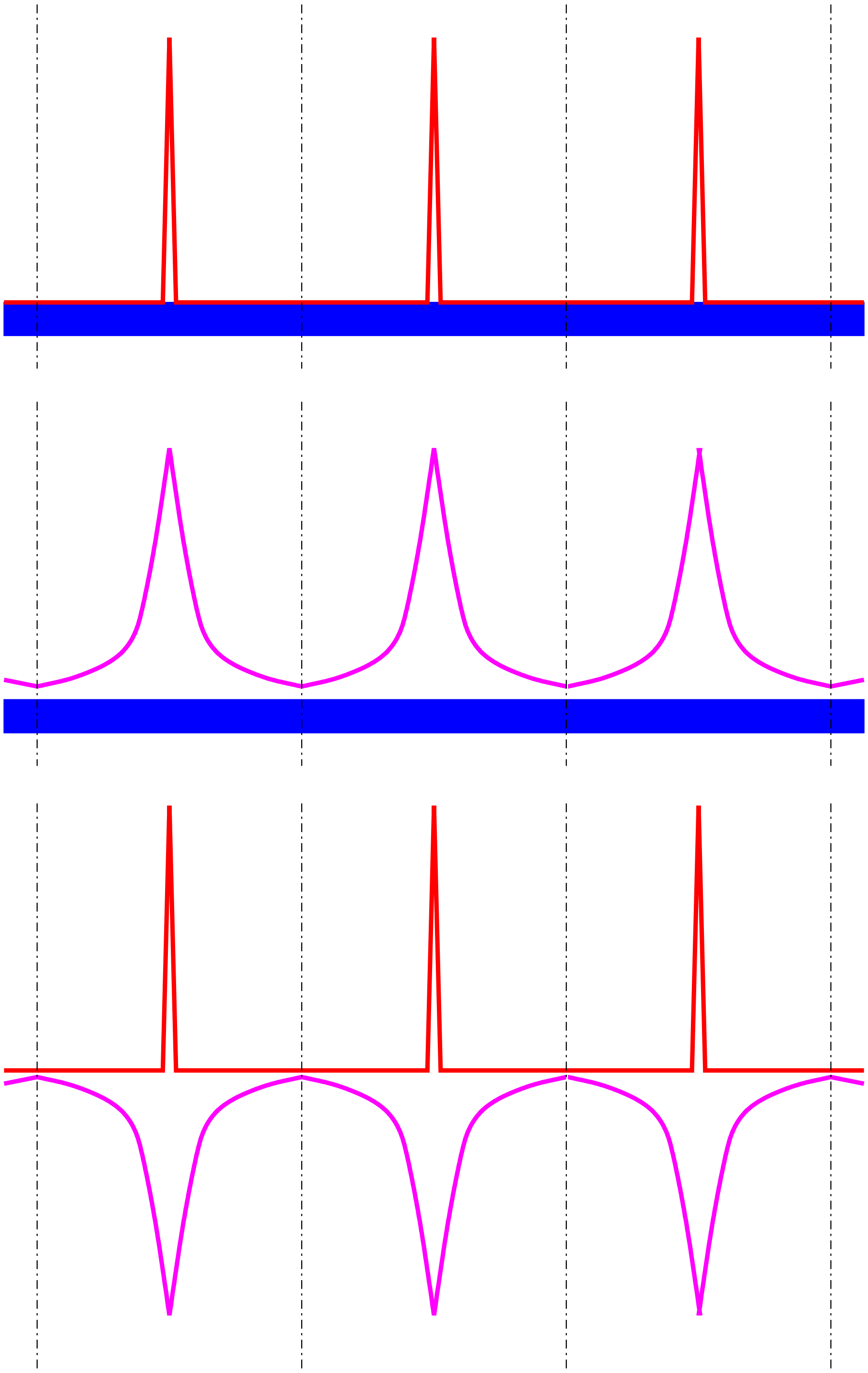

Figure 4.3: The Ewald method for periodic potentials, shown schematically in 1 dimension. A screening charge distribution is added and subtracted to the periodic point charge distribution, giving two rapidly convergent series. Top: The Ewald charge density is a sum of delta functions (red) minus a background charge density (blue). Centre: ρ_1 (equation 4.42) is given by an array of Gaussians (magenta) and a background charge density. Bottom: ρ_2 (equation 4.43) is given by an array of delta functions minus an array of Gaussians.

computed in reciprocal space: a Gaussian in real space becomes a Gaussian in reciprocal space. The potential, ϕ_2, due to ρ_2 is given in real space:

$$\phi_2(\mathbf{r}) = \sum_{\mathbf{L}} \frac{\mathrm{erfc}(|\mathbf{r}-\mathbf{r}_i-\mathbf{L}|/\mu)}{|\mathbf{r}-\mathbf{r}_i-\mathbf{L}|} - \frac{\pi\mu^2}{V^2} \quad , \tag{4.44}$$

where

$$\mathrm{erfc}(x) = 1 - \mathrm{erf}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x \exp(-x'^2)dx' \tag{4.45}$$

and the supercell volume is V. The sum runs over supercell lattice vectors, $\mathbf{L}$, and converges rapidly so that only a few vectors are needed in the sum. The overall Ewald potential is:

$$\begin{aligned} \phi_{\mathrm{Ewald}}(\mathbf{r}) &= \sum_{\mathbf{G}\neq 0} \frac{4\pi}{V\mathbf{G}^2} \exp(-\mu^2\mathbf{G}^2/4 + i(\mathbf{r}-\mathbf{r}_i).\mathbf{G}) \\ &+ \sum_{\mathbf{L}} \frac{\mathrm{erfc}(|\mathbf{r}-\mathbf{r}_i-\mathbf{L}|/\mu)}{|\mathbf{r}-\mathbf{r}_i-\mathbf{L}|} - \frac{\pi\mu^2}{V} \quad , \end{aligned} \tag{4.46}$$

where $\mathbf{G}$ are supercell reciprocal lattice vectors.

This potential could be used to compute the electron-ion potential. The Ewald sum for the electron-electron potential is similar, but the "self-image" of an electron is removed, due to the periodic array of images and Gaussian charges. Electrostatic energies are computed by multiplying the Ewald potential by the appropriate product of charges q_iq_j. In terms of speed and accuracy, for small numbers of particles (1000 or less), a well implemented version of Ewald's method has yet to be improved on. The computational cost Ewald's method, for a specified accuracy, scales as $N^{3/2}$. [77] Large numbers of particles are more efficiently dealt with using multipole expansions, based on hierarchical spatial subdivision of the charge density. [78,79] The largest supercell calculation in this thesis involves 1000 electrons (see chapter 6), and Ewald's method was applied in this case.

4.7 Wavefunction optimisation

The principles of wavefunction optimisation are straightforward: optimise (vary) a number of free parameters in a wavefunction in order to improve it. The preferred objective function used for optimisation is the variance of the local energy, $E_L = \hat{H}\Psi/\Psi$, and the reasons for this are developed in chapter 5. However, in order to evaluate *any* objective function related to the local energy, efficient methods for calculating this quantity are required. In this section, methods of evaluating the value of the wavefunction and its local energy for a *fixed* set of electron positions are presented.

During wavefunction optimisation, a representative sample of electron configurations is taken. This sample is then used for optimisation. As the parameters are varied during optimisation, the local energy is repeatedly evaluated for the fixed set of electron configurations. At a certain point, such as when parameters have been varied beyond a certain threshold, the new parameters are used to generate another set of electron configurations. This process is repeated until convergence is obtained. Typically, several thousand values of the parameters are tested during optimisation, so complete recalculation of all terms in the local energy would be very costly.

As detailed in section 4.4, the local energy consists of kinetic and potential energy terms. If the electron positions are fixed, only the kinetic energy and non-local pseudopotential energy change when the parameters are varied. These energies may be rapidly evaluated by taking advantage of the form of the Jastrow factor.

Equation 4.27, 4.28 and 4.29 show that the gradient and laplacian of the Slater determinants are also required, but the values of the actual orbitals in the determinants are not. The determinant dependent terms in the kinetic energy are independent of the Jastrow factor, with the exception of a cross term in the gradient of the determinants and Jastrow factor. This greatly simplifies the optimisation process, as it is easily made independent of the specific orbitals used.

4.7.1 Linearity of the Jastrow factor

A useful property of the Jastrow factor outlined in section 4.3.2 is the linearity of Jastrow factor with respect to the parameters. With the exception of the A parameter in the cusp-satisfying part of the function, all of the parameters are linear. For a fixed set of electron positions, the value of the Jastrow function is given by

$$J = f_0 + \sum_{k=1}^{N_p} \alpha_k f_k \quad , \tag{4.47}$$

where the N_p variable parameters, α, are multiplied by terms f_k that are dependent on the specific Jastrow factor and electron positions. The three Cartesian components of gradient and the value of the laplacian are similarly calculable in terms of products of parameters and "basis functions". By storing the f_k and similar terms for the gradient and laplacian, the value of the wavefunction, its gradient and laplacian are computable in terms of N_p, $3NN_p$ and N_p variables respectively. Recalculation of inter-particle distances and terms in the Jastrow factor is avoided, reducing the cost of the calculation by many orders of magnitude.

4.7.2 The non-local energy during wavefunction optimisation

The non-local energy is dependent on the wavefunction. During wavefunction optimisation the non-local energy changes, and should strictly be recalculated as parameter values are optimised.

For most systems, the non-local pseudopotential energy represents only a small fraction of the total energy, and the change in non-local pseudopotential energy throughout wavefunction optimisation is small. A reasonable approximation for the non-local energy is to consider it fixed throughout each cycle of wavefunction optimisation. This approximation biases the optimisation during any one cycle, but the non-local energy is recalculated when the configurations are regenerated. At worst, provided the non-local energy is only weakly dependent on the wavefunction parameters, this approximation will only slow convergence of the optimisation procedure. The approximation reduces the computational cost and memory requirements of optimisation.

The non-local energy may also be computed efficiently during wavefunction optimisation by taking advantage of the linearity of the parameters in the Jastrow factor, as is done for the evaluation of the kinetic energy.

Following equation 4.36, the non-local energy may be written

$$E_{\mathrm{nl}} = \sum_i \sum_I \sum_g \sum_l v_{Igl}(r^g_{iI}) \frac{\Psi(\mathbf{r}^g_{iI})}{\Psi(\mathbf{r}_{iI})} \quad , \tag{4.48}$$

where the sums run over electrons, i, ions, I, integration grid points, g, and angular momenta, l. $\Psi(\mathbf{r}^g_{iI})$ denotes the value of the many-body wavefunction with electron i, which is in the non–local range of ion I, with the electron at $\mathbf{r}_i$ moved to the integration grid point $\mathbf{r}^g_i$ and all other electrons unmoved. These sums may be simplified due to the short range of the non–local pseudopotential, so that only a small proportion of electrons and ions need be included. If only the linear parameters in the wavefunction are varied, the storage required for the evaluation is of order $n_p n_h n_g$, where there are n_p variable parameters, n_h electron-ion "hits" over all electrons and ions, and n_g integration grid points.

4.8 Statistics

The efficient generation and accurate analysis of data is central to all Monte Carlo simulations. In this section the statistical techniques used to compute the expected value of observables and then error bars are described. The techniques are similar to those used in other Monte Carlo and molecular dynamics methods, and the observables, though normally energies, can in principle be any quantity.

When analysing the data, two factors that must be taken into account are the equilibration (or relaxation) time of the observable, and the likely presence of serial correlations in the data. Compensating for correlations is essential in order to obtain an accurate error bar.

For the purposes of this section, a set of M sequentially observed data values x_i will be considered. The values of i correspond to simulation times $t_i = i\tau$, where τ is the Monte Carlo timestep between samples.

The mean or average value of the observable is given straightforwardly,

$$< X >= \frac{1}{M}\sum_{i=1}^{M} x_i \quad , \tag{4.49}$$

however evaluation of the error bar (the standard deviation of the mean) is more involved due to the likelyhood of serial correlations. The correlation time, τ_c is a measure of how long it takes the system to evolve between states that are independent. The correlation time is conventionally defined [16] via the autocorrelation function of the data, $\chi(t)$,

$$\chi(t) = \int [x(t')x(t'+t) - < x >^2] dt' \tag{4.50}$$

The autocorrelation function is defined such that $\chi = 0$ for completely uncorrelated data and $\chi = 1$ for completely correlated data. At long times, Markovian processes are expected to lead to an exponentially decay in χ,

$$\chi(t) \sim \exp(-t/\tau_c) \quad . \tag{4.51}$$

The autocorrelation function, equation 4.50, is estimated by [16]

$$\begin{aligned} \chi(t) = \chi(j\tau) \quad &= \quad \frac{1}{M-j}\sum_{i=0}^{M-j} x_i x_{x+j} \\ &- \quad \frac{1}{M-j}\sum_{i=0}^{M-j} x_i \times \frac{1}{M-j}\sum_{i=0}^{M-j} x_{i+j} \quad . \end{aligned} \tag{4.52}$$

Assuming the central limit theorem holds and following the conventional statistical definitions, [19] the best estimate of the standard deviation of the mean is given by

$$\sigma = \sqrt{\frac{2\tau_c}{M\tau}\left[\frac{1}{M}\sum_{i=1}^{M} x_i^2 - (\frac{1}{M}\sum_{i=1}^{M} x_i)^2\right]} \quad . \tag{4.53}$$

A simpler approach based on a reblocking of the data can be used in practice. In the reblocking scheme data is grouped into a series of block sizes, $1, 2, 4 \ldots$, and the standard deviation obtained by computing the variance of the block averages. For each block size, an estimate of the standard deviation of the mean can be shown to be [16, 75]

$$\sigma_b = \sqrt{\frac{1}{M_b - 1} \sum_{j=1}^{M_b} (< X_j^2 > - < X >_b^2)} \quad , \tag{4.54}$$

where X_j is an average over block j, $< X >_b$ the average over all blocks and M_b is the number of blocks. The standard deviation of the mean is proportional to the square root of the mean squared deviation in block averages, assuming a sufficient number of blocks and samples are included in the analysis. σ_b increases with block length b until a limiting value is reached, and this value corresponds to the true standard deviation of the mean. Where is an insufficient volume of data, a limiting value of σ_b is not reached. Other methods for computing the standard deviation of the mean, such as the jackknife and bootstrap procedures, [16] give similar results. An example of the reblocking analysis is shown in figure 4.4.

4.8.1 VMC applications

In VMC calculations an accurate estimate of the correlation length can be used to improve the efficiency of the simulation. In VMC calculations an energy measurement is costly compared with a Metropolis move. It is therefore efficient to only measure energies at timesteps approximately equal to the correlation time. The correlation time is system and wavefunction dependent but may be accurately determined using only short Monte Carlo runs. This criterion is also applicable when configurations are generated for wavefunction optimisation (see section 4.7).

The equilibration time of an observable in a VMC calculation is best determined by graphing the observable as a function of time. Equivalently, trends block averages may be used where the data is divided into short blocks. After equilibrating, the observable fluctuates around its final mean value. A further check, used in all calculations in this thesis, is to divide the complete set of data into halves or thirds. For a sufficiently long simulation, the averages of each block should agree to within error bars.

4.8.2 DMC applications

In DMC calculations an accurate estimate of the correlation length is essential as the calculations are typically one order of magnitude more expensive than VMC, partly due to

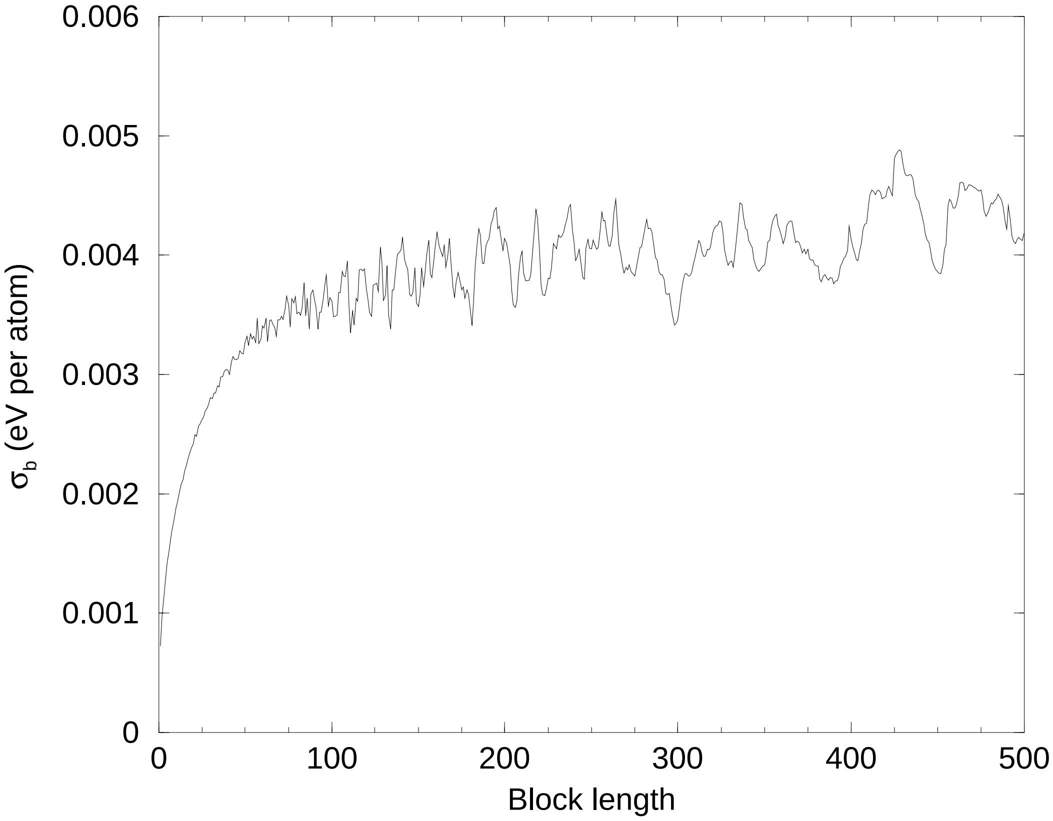

Figure 4.4: Reblocking of DMC energies. For this data, the reblocked error bar is the limiting value of $\sigma_b = 0.004$ eV. When insufficient data is reblocked, σ_b (equation 4.54) does not reach a limiting value with increasing block length.

the increased correlation length observed in DMC calculations. The increased correlation length results from the shorter timestep used in DMC calculations which are necessary to limit the errors in the short time approximation to the Green's function.

The ensemble average energy is computable at every timestep in DMC, as the individual walker energies are required to compute the branching factors. Equilibration is best judged by examining a graph of the ensemble average energy and DMC walker population against time. For observables other than the energy, it is not necessary to measure the observable at every step and the techniques used in VMC are applicable. Reblocking of the DMC data after equilibration gives accurate error bars.

In addition to the checks described for VMC calculations, comparing running averages with the overall average is a further useful method to check that there are no long term trends in the data. A long term downward trend in energy can result from a poor guiding wavefunction, which effectively increases the equilibration and correlation times for the system. The DMC walker population provides a further guide: the population is stable when the reference energy is equal to the fixed–node ground state energy. Figure 4.5 shows an example DMC calculation.

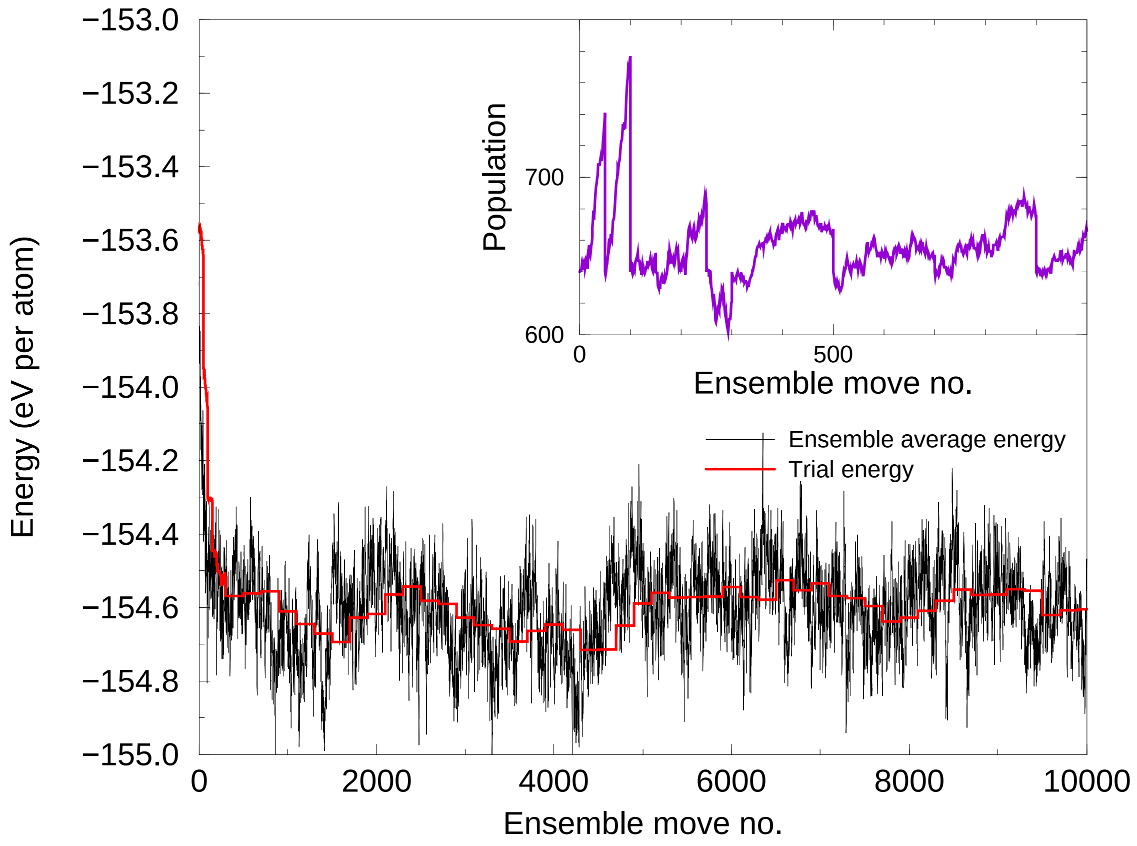

Figure 4.5: An example DMC calculation: Energies versus move number and (inset) walker population versus move number. During the initial equilibration stage, approximately 300 ensemble moves, the ensemble average energy rapidly decays to the fixed-node ground state energy and the trial energy, E_T, is updated frequently to maintain the population of the ensemble close to a target value. After the excited states have propagated out the trial energy is only updated occasionally to avoid population control errors. For this data, the DMC energy would be computed after discarding approximately the first 1000 moves. (Inset) During equilibration, the walker population increases rapidly and must be renormalised. After equilibration, the walker population remains stable around the target value of 640.

4.8.3 Systematic errors

In the previous two sections, statistical techniques for the analysis of QMC data were presented. These techniques are unable to detect any systematic errors which could be present in the simulation.

Potential systematic errors include:

- Insufficient exploration of configuration space results if a simulation is run for too short a period of time, or if the timestep chosen severely limits the number of accepted Monte Carlo moves. (The ratio of accepted to proposed moves is typically 50% in VMC and $> 99\%$ in DMC calculations.)
- Inclusion of unequilibrated data in final averaged data gives a systematic bias to the averages obtained.
- Updating the trial energy or renormalising the number of walkers in the ensemble too frequently results in population control errors, and the true ground state distribution and energy is not obtained.
- In DMC calculations, a systematic timestep error will always be present due to the approximation used for the Green's function. For every system, DMC calculations must be repeated with progressively smaller timesteps until results agree to within error bars at the *final level* of statistical accuracy desired.

Systematic errors can be avoided by careful benchmarking of new systems. When doing so it is important to consider possible changes in wavefunction quality as system size and geometry is changed. The cancellation of remaining timestep and other systematic errors in energy differences may be encouraged by choosing a fixed timestep and block size (period between population renormalisation) for all calculations. However, the calculations should always be setup such that these errors are not significant even if no cancellation takes place.

4.9 Summary

The details of implementing QMC methods were presented. The need for an efficient implementation at all levels was stressed: the Monte Carlo algorithms themselves, wavefunction evaluation, pseudopotential evaluation, wavefunction optimisation and parallelisation of the methods.

The accurate determination of physics results is critical to the entire field of simulation, and statistical methods for analysing simulation data and computing error bars were described.

All of these methods are used extensively in the following chapters.

Chapter 5

Wavefunction optimisation

5.1 Introduction

Accurate approximations to many–body wavefunctions are crucial for the success of QMC calculations. As highlighted in earlier chapters, VMC and DMC require accurate wavefunctions for both accuracy and efficiency.

In this chapter, several Monte Carlo techniques for optimizing many–body wavefunctions are investigated. The development of statistically efficient and numerically stable optimization methods is important, as the cost of optimising a many–body wavefunction is high. The most commonly used techniques in the literature are evaluated and several variants are introduced. Two reasons why variance minimisation exhibits superior numerical stability to energy minimisation are identified by investigating a realistic model system. A robust and efficient variance minimisation scheme for optimising the wavefunctions of large systems is developed.

This scheme is applied to a series of real systems in the later chapters of the thesis, where the difficulties of optimisation are far greater. These successful applications demonstrate the scheme's practicality and efficiency. Wavefunction optimisation for the carbon clusters of chapter 8 would not be possible using some of the simpler optimisation techniques that have been used in the past.

The key issues and importance of wavefunction optimisation are briefly reviewed in section 5.2. Several different choices of objective function for optimisation schemes are then examined in section 5.3. In section 5.4 and higher, possible objective functions are analysed and tested numerically. A procedure for limiting the energy of "outlying" configurations is presented in section 5.8, which in practice enables the value of objective functions to be evaluated with an improved statistical accuracy. Conclusions and a summary of the proposed method are given in section 5.10.

5.2 The importance of wavefunction optimisation

Wavefunction optimisation is one of the most critical, time consuming and important stages of a QMC calculation. Several strategies can be adopted depending on the intended use for the optimised wavefunction, but in all cases the stability and efficiency of the process is critical.

Typical solid–state QMC calculations currently use of order 10^2 parameters for 10^3–dimensional functions. The number of parameters and dimensionality of the problems tackled by QMC methods can only be expected to increase. The development of reliable and efficient methods for wavefunction optimisation is therefore important. Although many optimisation methods are present in the literature (some of these are reviewed in the next section), most have only been applied to small molecules where the number of parameters and problem dimensionality are small.

In VMC calculations, the accuracy of a trial wavefunction limits the statistical efficiency of the calculation and the final accuracy of the result obtained. It is desirable to use wavefunctions which are as accurate as possible yet may be rapidly computed. In practice, this means that the parameterised form of a many–body wavefunction should be relatively easy to compute, and that each parameter in the wavefunction should significantly improve the accuracy (quality) of the wavefunction. In extreme cases, such as when the best possible variational wavefunction is required, hundreds of parameters may be used. Optimisation of these parameters, and hence the wavefunction, can be difficult, particularly when there is only a small energetic dependency on some of the parameters. In this case, very long Monte Carlo calculations are required in order to obtain statistically significant data. It is clearly desirable, in all cases, that any optimisation method makes efficient use of data and is stable, delivering consistent improvements to the wavefunction.

In DMC calculations, the results obtained are less dependent on the trial wavefunction than in VMC calculations. Typically, the nodes of the wavefunction are not optimised and only the Jastrow part of the wavefunction are parameterised. Speed of evaluation is a little more important than in VMC calculations because derivatives of the Jastrow function are always required. The overall efficiency of a DMC calculation is determined by how well the trial wavefunction approximates an eigenstate over *all* of configuration space, as the accuracy of the local energy determines the branching of configurations.

5.3 Objective functions

In order to optimize a wavefunction we require an objective function, i.e., a quantity which is to be minimised with respect to a set of parameters, $\{\alpha\}$. The criteria that a successful objective function should satisfy for use in a Monte Carlo optimization procedure are that (i) the global minimum of the objective function should correspond to a high quality wavefunction, (ii) the variance of the objective function should be as small as possible, and (iii) the function is suitable for numerical optimisation. Ideally the global minimum of the objective function should be as sharp and deep as possible and there should be no local minima, i.e. as best adapted to numerical optimization algorithms as possible.

One natural objective function is the expectation value of the energy,

$$E_{\mathrm{V}} = \frac{\int \Phi^2(\alpha)\,[\Phi^{-1}(\alpha)\hat{H}\Phi(\alpha)]\,d\mathbf{R}}{\int \Phi^2(\alpha)\,d\mathbf{R}} \quad , \tag{5.1}$$

where the integrals are over the $3N$–dimensional configuration space. The numerator is the integral over the probability distribution, $\Phi^2(\alpha)$, of the local energy, $E_{\mathrm{L}}(\alpha) = \Phi^{-1}(\alpha)\hat{H}\Phi(\alpha)$.

The energy is in fact not the preferred objective function for wave function optimization, and the general consensus is that a better procedure is to minimise the variance of the energy, which is given by

$$A(\alpha) = \frac{\int \Phi^2(\alpha)\,[E_{\mathrm{L}}(\alpha) - E_{\mathrm{V}}(\alpha)]^2\,d\mathbf{R}}{\int \Phi^2(\alpha)\,d\mathbf{R}} \quad . \tag{5.2}$$

Optimising wavefunctions by minimising the variance of the energy is actually a very old idea, having being used in the 1930's. The first application using Monte Carlo techniques to evaluate the integrals appears to have been by Conroy, [80] but the present popularity of the method derives from the developments of Umrigar and coworkers. [29, 37]

A number of reasons have been advanced by many workers in the field for preferring variance minimisation over energy minimisation, including: (i) it has a known lower bound of zero, (ii) the resulting wavefunctions give good estimates for a range of properties, not just the energy, (iii) it can be applied to excited states, (iv) efficient algorithms are known for minimising objective functions which can be written as a sum of squares, and (v) it exhibits greater numerical stability than energy minimisation. The latter point is very significant for applications to large systems.

Minimisation of $A(\alpha)$ has normally been carried out via a correlated sampling approach in which a set of configurations distributed according to $\Phi^2(\alpha_0)$ is generated, where α_0 is an initial set of parameter values. $A(\alpha)$ is then evaluated as

$$A(\alpha) = \frac{\int \Phi^2(\alpha_0)\, w(\alpha)\, [E_{\rm L}(\alpha) - E_{\rm V}(\alpha)]^2\, d\mathbf{R}}{\int \Phi^2(\alpha_0)\, w(\alpha)\, d\mathbf{R}} \quad , \tag{5.3}$$

where the integrals contain a weighting factor, $w(\alpha)$, given by

$$w(\alpha) = \frac{\Phi^2(\alpha)}{\Phi^2(\alpha_0)} \quad . \tag{5.4}$$

$A(\alpha)$ is then minimised with respect to the parameters $\{\alpha\}$. The minimum possible value of $A(\alpha)$ is zero. This value is obtained if and only if $\Phi(\alpha)$ is an exact eigenstate of $\hat{H}$. The ensemble of configurations is normally regenerated several times with the updated parameter values so that when convergence is obtained no change in parameters $\{\alpha\}$ occurs on optimization. A variant of equation 5.3 is obtained by replacing the energy $E_{\rm V}(\alpha)$ by a fixed value, $\bar{E}$, giving

$$B(\alpha) = \frac{\int \Phi^2(\alpha_0)\, w(\alpha)\, [E_{\rm L}(\alpha) - \bar{E}]^2\, d\mathbf{R}}{\int \Phi^2(\alpha_0)\, w(\alpha)\, d\mathbf{R}} \quad . \tag{5.5}$$

Note that if $\bar{E} \leq E_0$, where E_0 is the exact ground state energy, then the minimum possible value of $B(\alpha)$ occurs when $\Phi{=}\Phi_0$, the exact ground state wavefunction. Minimisation of $B(\alpha)$ is equivalent to minimising a linear combination of $E_{\rm V}$ and $A(\alpha)$. [23] The absolute minima of both $E_{\rm V}$ and $A(\alpha)$ occur when $\Phi{=}\Phi_0$. If both of the coefficients of $E_{\rm V}$ and $A(\alpha)$ in the linear combination are positive, which is guaranteed if $\bar{E} \leq E_0$, then it follows that the absolute minimum of $B(\alpha)$ occurs at $\Phi{=}\Phi_0$. Using this method with $\bar{E} \leq E_0$ allows optimization only of the ground state wavefunction, and it has been claimed that this method yields more accurate wavefunctions than variance minimisation alone (for example, Ref. [53]).

5.4 Numerical instabilities

Although minimisation of the variance using correlated sampling methods has often been successful, in some cases the procedure can exhibit a numerical instability. Two situations where this is likely to occur have been identified. The first is when the nodes of the trial wavefunction are allowed to alter during the optimization process. A similar instability can arise when the number of electrons in the systems becomes large, which can result in an instability even if the nodes of the trial wavefunction remain fixed. The characteristic of these numerical instabilities is that during the minimisation procedure a few configurations (typically one to several tens of configurations) acquire a very large weight. The estimate of the variance is then reduced almost to zero by a set of parameters that are unphysical and which are found to give extremely poor results in subsequent QMC calculations.

When the nodes of the trial wavefunction are altered large weights are most likely to occur for configurations close to the zeros of the probability distribution $\Phi^2(\alpha_0)$. Large weights can also occur when varying the Jastrow factor if the number of electrons, N, is large. For a small change in the one–body function, $\delta\chi$, the local energy changes by an amount *proportional* to $N\delta\chi$, but the weight is multiplied by a factor which is *exponential* in $N\delta\chi$, which can result in very large or very small weights if N is large. A similar argument holds for changes in the two–body term, which shows an even more severe potential instability because the change in the two–body term scales like N^2. For high order correlations - three-body and higher terms - the potential instabilities increase in severity. This difficulty in optimising high order correlations indicates a problem with current methods of wavefunction optimisation: the optimisation process becomes increasingly fragile as higher orders of correlation are added. Until this problem is resolved it is likely that wavefunctions including high order correlations will be restricted to small systems.

The instability due to the weights has been noticed by many researchers. In principle one could overcome this instability by using more configurations, but the number required is normally impractically large. Various ways of dealing with this instability have been devised. One method is to limit the upper value of the weights [53] or to set the weights equal to unity. [46, 81] Schmidt and Moskowitz [81] set the weights equal to unity in calculations for small systems in which the nodes were altered. An alternative approach is to draw the configurations from a modified probability distribution which is positive definite, so that the weights do not get very large. [82] For the wavefunctions in this thesis, which include calculations for large systems of up to 1000 electrons, [83] the weights were set equal to unity while optimising the Jastrow factor.

When using the correlated sampling approach, whether or not the weights are modified, better results are obtained by periodically regenerating a new set of configurations chosen from the distribution $\Phi^2(\alpha)$, where $\{\alpha\}$ is the updated parameter set. This helps the convergence of the minimisation procedure by reducing any bias due to the small sample of configurations used in the optimisation procedure. One can also restrict the allowed variation in the parameters $\{\alpha\}$ before regenerating a new set of configurations, but this can slow the convergence. It has been found [46] that setting the weights to unity allowed the parameters to be altered by a larger amount before the configurations must be regenerated. Typically, after three or four regenerations, the parameters converge to stable values giving a small variance and low energy in subsequent VMC calculations.

These strategies can often overcome the numerical instability. For QMC methods to be applied to large systems with many inequivalent atoms, wavefunctions with many variable parameters are required. It is likely that the determinantal part of the wavefunction as well as the Jastrow factor will eventually be optimised on a regular basis. This has only recently been attempted for solids [84] in a limited form. It is also desirable that excited states as well as ground states be optimised. In order to accomplish these aims, improved optimization techniques are essential. In the remainder of this chapter, a deeper understanding of the issues of numerical stability is developed by analysing different energy and variance minimisation techniques.

5.5 Analysis of objective functions

In this section, different objective functions are considered and their suitability for wavefunction optimisation is evaluated. First, the procedure of setting the weights to unity is considered. This gives a new objective function, $C(\alpha)$, where

$$C(\alpha) = \frac{\int \Phi^2(\alpha_0)\, [E_{\mathrm{L}}(\alpha) - E_{\mathrm{V}}(\alpha)]^2 \, d\mathbf{R}}{\int \Phi^2(\alpha_0)\, d\mathbf{R}} \quad . \tag{5.6}$$

This objective function has the property that its absolute minimum is zero and that this value is obtained if and only if $\Phi(\alpha)$ is an exact eigenstate of $\hat{H}$, i.e., the minima have the same properties as those of $A(\alpha)$. A similar analysis can be applied to the case where the weights are subject to an upper limit, and we will refer to all such expressions with modified weights as variants of C.

It is interesting to consider the behaviour of the expectation value of the energy calculated without weighting. The minima of

$$E_C = \frac{\int \Phi^2(\alpha_0)\, [\Phi^{-1}(\alpha)\hat{H}\Phi(\alpha)] \, d\mathbf{R}}{\int \Phi^2(\alpha_0)\, d\mathbf{R}} \quad , \tag{5.7}$$

are *not* located at the eigenstates of $\hat{H}$ unless $\Phi(\alpha_0)$ is an eigenstate. The energy E_C is therefore not a satisfactory objective function. Given this result it is easy to see that if we replace the energy $E_{\mathrm{V}}(\alpha)$ in equation 5.6 by $\bar{E}$, then unless $\bar{E}$ is equal to an eigenenergy then the minima of this new variance–like objective function are not located at the eigenstates. Therefore, if the weights are to be limited or set to unity, then one should use the reference energy E_{V} in the variance rather than a value of $\bar{E}$ which is below E_{V}.

The above analysis applies for wavefunctions with sufficient variational freedom to encompass the exact wavefunction. In practical situations we are unable to find exact wavefunctions and it is important to consider the effect this has on the optimization process. Although the objective functions $A(\alpha)$ and $C(\alpha)$ are unbiased in the sense that the exact ground state wavefunction corresponds to an absolute minimum, $C(\alpha)$ is biased in the sense that for a wavefunction which cannot be exact the optimized parameters will not exactly minimise the true variance. This may be considered a "weak bias" because it disappears as the wavefunction tends to the exact one. In practice this is not a problem because in minimising $C(\alpha)$ the configurations are regenerated several times with the updated distribution until convergence is obtained, so that minimisation of $A(\alpha)$ and $C(\alpha)$ turns out to give almost identical parameter values. On the other hand, the unweighted energy, E_C, shows a "strong bias" in the sense that the nature of its stationary points are very different from those of the properly weighted energy. In section 5.7 it is shown that the exact wavefunction may not even correspond to a minimum of E_C. The ability to alter the weights while not affecting the positions of the minima is an important advantage of variance minimisation over energy minimisation. This is one of the factors which leads to the greater numerical stability of variance minimisation over energy minimisation.

5.6 Further effects of finite sampling

In the previous section the numerical instability arising from the weighting factors was described. The origin of this problem lies in approximating the integrals by the average of the integrand over a finite set of points in configuration space. There is another important issue connected with the approximation of finite sampling, which is whether the positions of the minima of the objective function are altered by the finite sampling itself.

Consider the objective function $A(\alpha)$, in the case where the trial wavefunction has sufficient variational freedom to encompass the exact wavefunction. Approximating equation 5.3 by an average over the set $\{\mathbf{R}_i\}$ containing N_s configurations drawn from the distribution $\Phi^2(\alpha_0)$ gives

$$A^{N_s} = \frac{\sum_i^{N_s} w(\mathbf{R}_i;\alpha)[E_{\mathrm{L}}(\mathbf{R}_i;\alpha) - E_{\mathrm{V}}(\{\mathbf{R}_i\};\alpha)]^2}{\sum_i^{N_s} w(\mathbf{R}_i;\alpha)} \quad . \tag{5.8}$$

The eigenstates of $\hat{H}$ give $A^{N_s} = 0$ for *any* size of sample because $E_{\mathrm{L}} = E_{\mathrm{V}}$ for an eigenstate. Clearly this result also holds for $C(\alpha)$. This behaviour contrasts with that of the variational energy, E_{V}. Consider a finite sampling of the variational energy of equation 5.1, where the configurations are distributed according to $\Phi^2(\alpha_0)$ and properly weighted,

$$E_{\mathrm{V}}^{N_s} = \frac{\sum_i^{N_s} w(\mathbf{R}_i;\alpha)E_{\mathrm{L}}(\mathbf{R}_i;\alpha)}{\sum_i^{N_s} w(\mathbf{R}_i;\alpha)} \quad . \tag{5.9}$$

The global minima of $E_{\mathrm{V}}^{N_s}$ are not guaranteed to lie at the eigenstates of $\hat{H}$ for a finite sample. The fact that the positions of the global minima of $A(\alpha)$ and $C(\alpha)$ are robust to finite sampling is a second important advantage of variance minimisation over energy minimisation.

5.7 Tests of minimisation procedures

It is important that any theoretical improvements in minimisation procedures are tested in actual calculations. In order to test the different procedures at an affordable computational cost, a non–interacting test system was selected. In a non–interacting system the exact wavefunction for the system is known, which helps in the analysis of minimisation procedures. We believe that the conclusions drawn from this systems are general because in more realistic calculations the problems simply become more serious.

5.7.1 The model system

A non–interacting model of the valence states of silicon in the diamond structure was chosen, using periodic boundary conditions to simulate the solid. This system is representative of the systems considered in current calculations, provided any instabilities in actual optimisations are not due to electron–electron interactions.

The model system consistent of an fcc simulation cell containing 16 atoms and 64–electrons. The electrons were subject to a local potential which is described by two Fourier components, V_{111} = -0.1 a.u., and V_{220} = -0.06 a.u., chosen to give a reasonable description of the valence bandstructure of silicon. These values are in reasonable agreement with empirical pseudopotential form factors for silicon [85]. Overall this model gives a reasonable description of the valence states of silicon and retains the essential features for testing the optimization techniques.

Near exact single–particle orbitals were obtained by diagonalising the Hamiltonian in a plane–wave basis set containing all waves up to a kinetic energy cutoff of 15 a.u. This basis set is still incomplete, but the square root of the variance of the energy is about 0.02 eV per atom, which is negligible for current purposes. A variational parameter, α, was added to the wavefunction in the form of a χ function with the full symmetry of the diamond structure:

$$\chi(\mathbf{r}) = \alpha \left(\sum_{\mathbf{G}} P_{\mathbf{G}} e^{i\mathbf{G}.\mathbf{r}} \right) \quad , \tag{5.10}$$

where $\mathbf{G}$ labels the 8 reciprocal lattice vectors of the [111] star and $P_{\mathbf{G}}$ is a phase factor associated with the non–symmorphic symmetry operations. The many–body wavefunction is given by

$$\psi = D^{\uparrow} D^{\downarrow} \exp \left[\sum_{i=1}^{N} \chi(\mathbf{r}_i) \right] \quad . \tag{5.11}$$

The exact value of the parameter, α, is, of course, zero. To model the situation where the wavefunction does not possess the variational freedom to encompass the exact one we used a smaller basis set cutoff of 2.5 a.u. The variational energy from this wavefunction is 0.35 eV per atom above the exact value, which is typical of the values we encounter in our solid state calculations. The optimal value of α for this inexact wavefunction is very close to zero.

This model exhibits all the numerical problems encountered in optimization procedures. In practical situations one may have more electrons and more parameters to optimize, which makes the numerical instabilities more pronounced. In order to analyse the behaviour in detail the variance of the difference objective functions was analysed. Unfeasibly large numbers of electron configurations were required to obtain accurate values of the variance of the objective functions for wavefunctions with more electrons and variable parameters than used in the model system. For these systems, improvements in optimisation procedures will be even more advantageous.

5.7.2 Generation of test ensemble

A sample of 0.96×10^6 statistically independent electronic configurations was generated and used to calculate the quantities involved in the various optimization schemes. In practical QMC applications, a typical number of configurations used for optimisation might be 10^4, but it was necessary to use a much larger number in these tests to obtain sufficiently accurate values of the different objective functions and, in particular, their variances. In a practical application an objective function, say $C(\alpha)$, is evaluated using, say, 10^4 configurations. The quantities of interest are then $C(\alpha)$ and its variance calculated as averages over blocks of 10^4 configurations.

The configurations were generated by a Metropolis walk distributed according to Φ^2, using the (inexact) reduced basis–set wave function. An optimization procedure typically starts with non–optimal parameter values which are improved during the optimization procedure. Results are presented for configurations generated with the non–optimal value of $\alpha_0 = 0.03$, which gives results typical of the starting value for an optimization, and $\alpha_0 = 0$, which is the final value from a successful optimization procedure. The qualitative behaviour was found to be not strongly influenced by the value of α_0.

5.7.3 Results

Energy minimisation schemes are considered first. In figure 5.1 the weighted and unweighted mean energies, E_{V} and E_C, and their variances as a function of α are shown, with configurations generated from $\alpha_0 = 0.03$ (figure 5.1a) and $\alpha_0 = 0$ (figure 5.1b).

The unweighted mean energy has a *maximum* at α_0, i.e., the value from which the configurations were generated. This result can be understood as follows. Consider a wave function of the form

$$\Phi = D^{\uparrow} D^{\downarrow} \exp\left[\sum_k \alpha_k J_k\right] \quad , \tag{5.12}$$

where the α_k are parameters, the J_k are correlation functions and the $D^{\uparrow}$, $D^{\downarrow}$ are up and down-spin determinants. The mean unweighted energy (equation 5.7) can be written as

$$E_C(\{\alpha_k\}) = (\alpha_k - \alpha_{k0})\, G_{kl}\, (\alpha_l - \alpha_{l0}) \quad + \quad \text{constant} \quad , \tag{5.13}$$

where α_{k0} are the parameter values from which the configurations are generated and

$$G_{kl} = \frac{-\frac{1}{2}\int \Phi^2(\{\alpha_{k0}\}) \sum_i \nabla_i J_k \nabla_i J_l \, d\mathbf{R}}{\int \Phi^2(\{\alpha_{k0}\})\, d\mathbf{R}} \quad . \tag{5.14}$$

The G_{kl} and the constant term depend on the α_{k0} but not on the α_k. When there is only a single parameter, G is negative, so that E_C is a quadratic function with a maximum at $\alpha_k = \alpha_{k0}$. When there is more than one parameter the stationary point of the quadratic can be a maximum, minimum or saddle point, which is not acceptable behaviour for an objective function. The weights may be altered in other ways, such as limiting their upper value, but if the weights are altered the minima of the energy are moved, which is a "strong bias" in the objective function. Therefore, if an energy minimisation method is used, weights *must* be included.

The distributions of the weights and local energies of the test ensemble are now examined. In figure 5.2 the distributions of the weights for $\alpha_0 = 0.03$ and $\alpha = 0$ and for $\alpha_0 = 0$ and $\alpha = 0.03$ are plotted, while in figure 5.3 the corresponding distributions of the local energies are shown. The distributions of the weights resemble Poisson distributions, but the square roots of the variances are significantly greater than the means, so there are more configurations at large weights than for a Poisson distribution with the same mean. The local energies follow normal distributions relatively well. As expected, the distributions of the local energies is wider for the $\alpha_0 = 0.03$ wavefunction. Closer inspection reveals that the distribution of the local energies is not exactly normal because the actual distributions have "fat tails". The outlying energies result from outlying values of the kinetic energy. The standard deviations are $\sigma = 0.964$ a.u. and $\sigma = 0.726$ a.u., for $\alpha_0 = 0.03$ and 0, respectively. The expected percentage of configurations beyond 3σ from the mean of a normal distribution is 0.27%, but the actual percentages are 0.443% and 0.608% for $\alpha_0 = 0.03$ and 0, respectively. Although these outlying local energies give a negligible contribution to the mean energy, calculated with or without weighting, and only a very small contribution

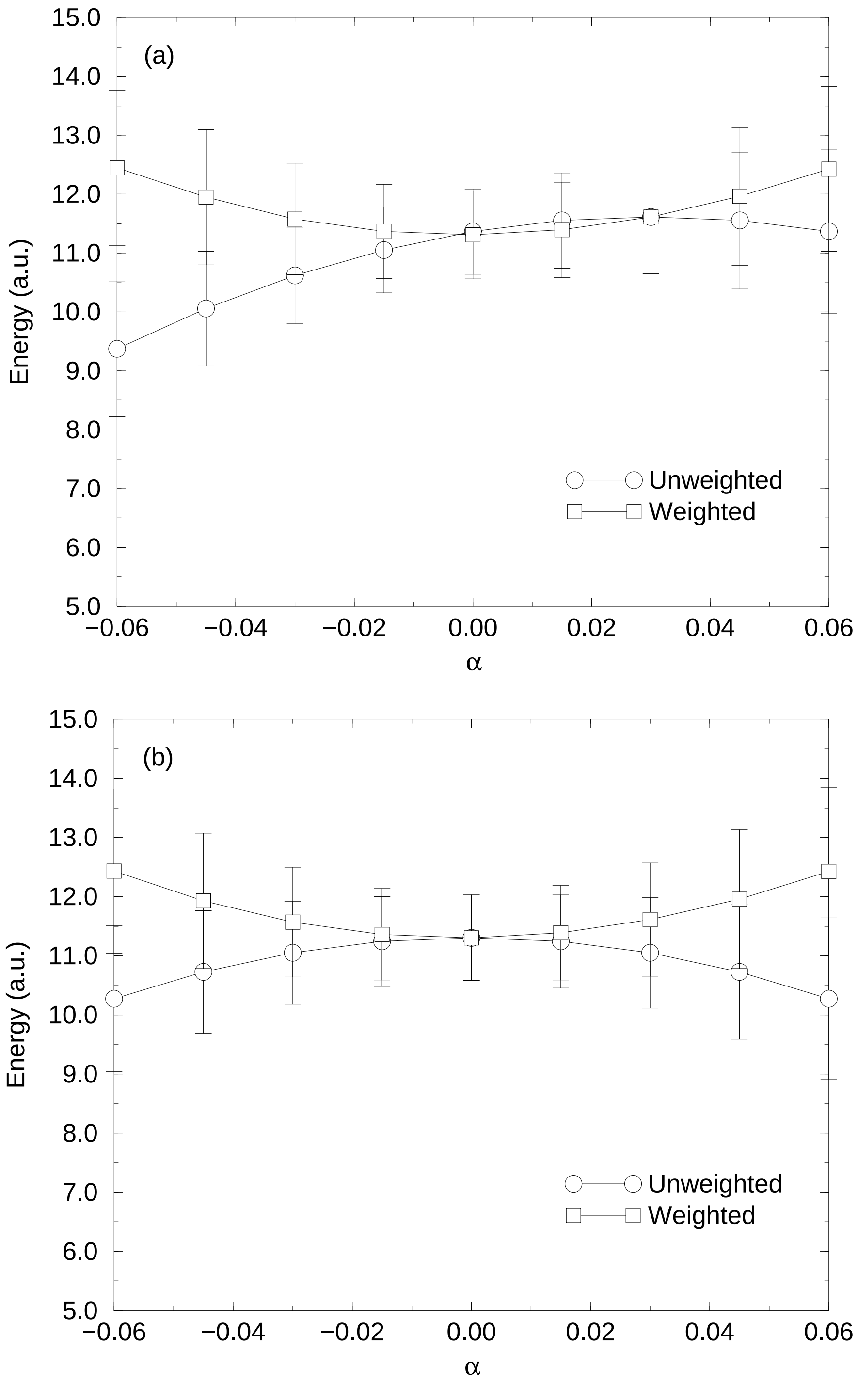

Figure 5.1: Weighted and unweighted mean energies and standard deviations, shown as error bars, for configurations generated with (a) α_0=0.03 and (b) α_0=0.

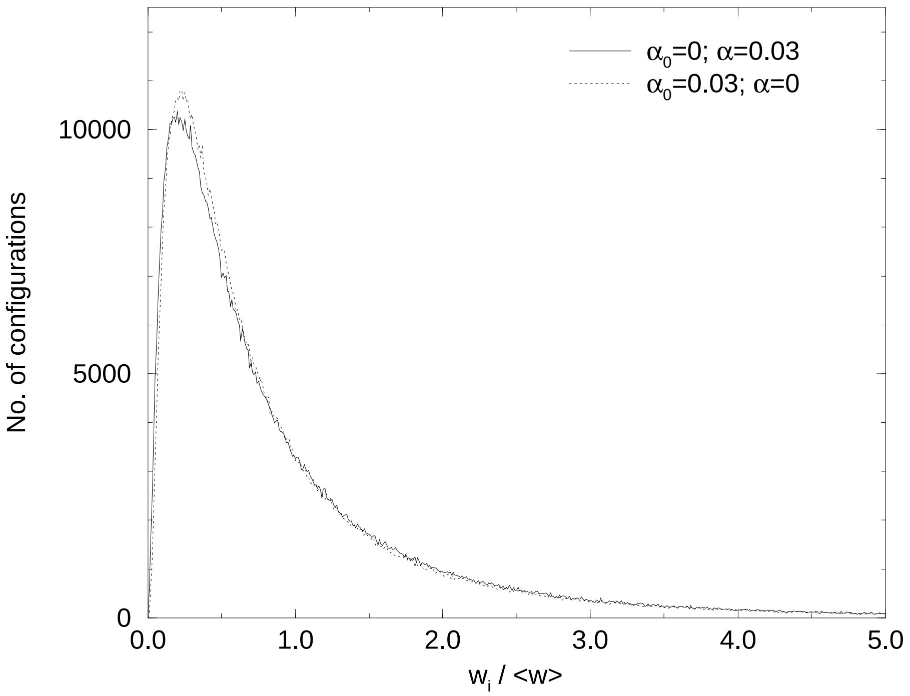

Figure 5.2: Distributions of weights for configurations generated with α_0=0.03 and α_0=0, evaluated with α=0 and α=0.03, respectively.

to the values of the variance–like objective functions, $A(\alpha)$, $B(\alpha)$, and $C(\alpha)$, they give significant contributions to the variances of the variance–like objective functions.

5.8 Limitation of outlying energies

It is highly undesirable for an objective function to have a large variance. A larger variance implies that a greater number of configurations is required to determine the objective function to a given accuracy. As noted above, only the variances and not the means of $A(\alpha)$, $B(\alpha)$, and $C(\alpha)$ are significantly affected by the outlying configurations. A decision was made to limit the outlying local energies. An alternative would be to delete the outlying configurations, but this introduces a greater bias and is not as convenient in correlated–sampling schemes. The limiting must be done by the introduction of an arbitrary criterion, which was implemented as follows. First the standard deviation of the sampled local energies, σ, is calculated. Limiting values for the local energy are then calculated as those beyond which the total expected number of configurations based on a normal distribution is less than Δ, where

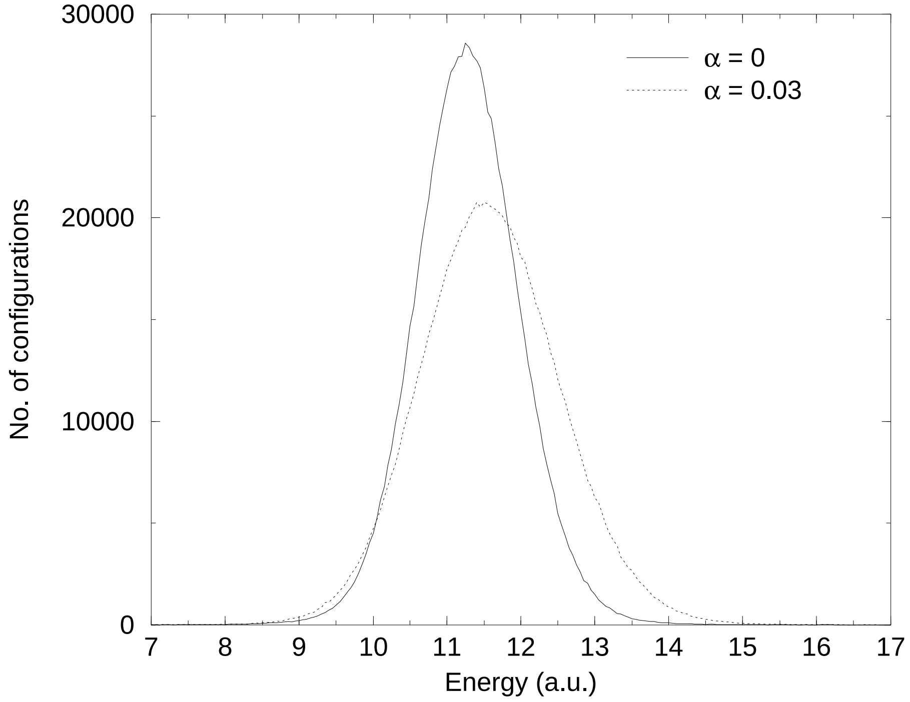

Figure 5.3: Distributions of the local energies for α=0 and α=0.03.

$$\Delta = N_s \times 10^{-p} \quad , \tag{5.15}$$

N_s is the total number of configurations and p is typically chosen to be 8, although varying p from 4 to 12 makes no significant difference to the results.

A factor of N_s was included, rather than limiting the energies beyond a given number of standard deviations to incorporate the concept that as more configurations are included, the sampling is improved. In the limit of perfect sampling, $N_s \to \infty$, the objective functions are unchanged. For the model silicon system, the percentage of configurations having their local energies limited by this procedure, with $p = 8$, is only 0.024% and 0.047% for $\alpha_0 =$ 0.03 and 0, respectively. These values corresponds to those configurations lying beyond 5.7 standard deviations from the mean. The effect of limiting the outlying local energies is illustrated in figure 5.4. In figures 5.4a and c the mean values of the objective functions C and A versus α for configurations generated with $\alpha_0 = 0.03$, with values of the limiting power, p, in equation 5.15, of 4, 8, 12 and infinity (no limiting) are shown, while in figures 5.4b and d their variances are plotted.

The mean values of C are hardly affected by the limiting, while those of A are only

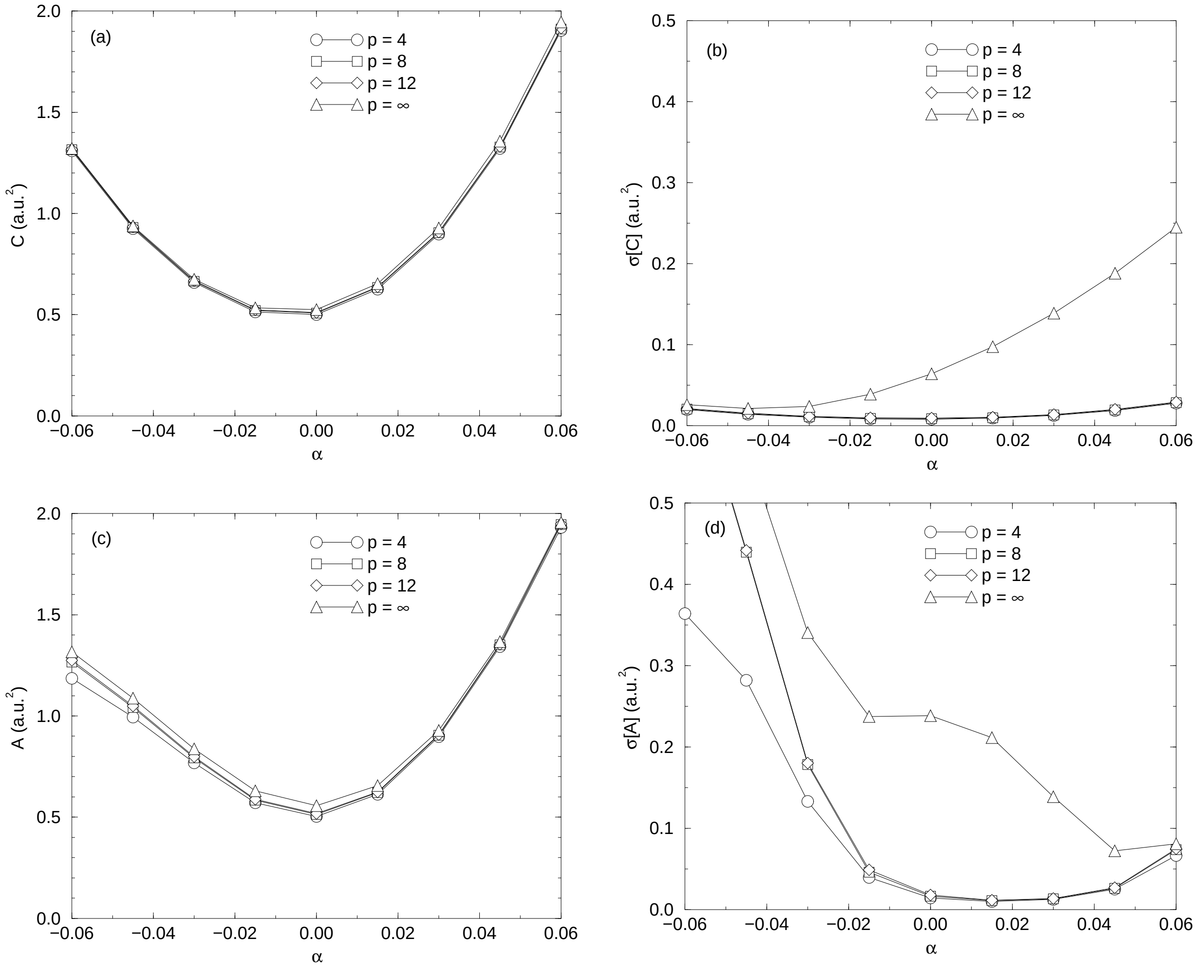

Figure 5.4: The effect of limiting outlying energies on (a,b) objective function C (the unweighted variance) and (c,d) objective function A (the weighted variance) with α_0=0.03. Outlying energies are limited as in equation 5.15 with the values of p shown.

slightly altered. The smaller variances of C and A obtained by limiting the values of the local energy are very clear. If the local energies are not limited then the variances of the objective function are not very accurately determined, even with the large samples of 0.96×10^6 configurations. Similar results hold for configurations generated with $\alpha_0 = 0$. This simple procedure can greatly reduce the variance of the variance–like objective functions without significantly affecting their mean values. Limiting the local energies is even more advantageous when small numbers of configurations are used. Limiting the local energies in this way gives significantly better numerical behaviour for all the variance–like objective functions and therefore all data shown in figures 5.5–5.7 have been limited with p=8, unless explicitly stated otherwise.

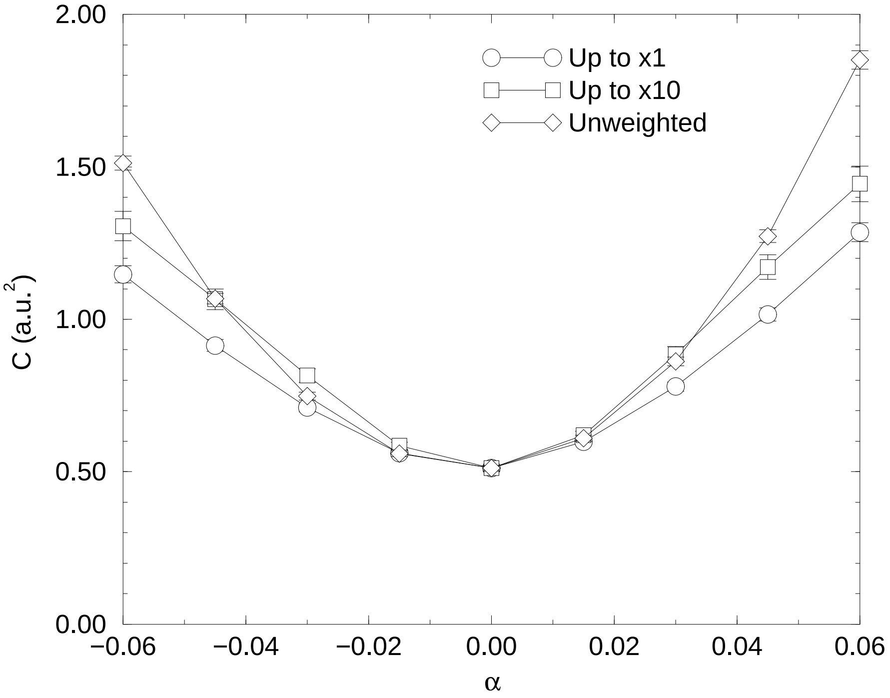

Figure 5.5: The unweighted objective function C generated with $\alpha_0 = 0$ and with limiting of the weights.

Limiting the values of the weights is a crucial part of the variance minimisation procedure for large systems. Comparison of figures 5.4b and d shows that the variance of the unweighted objective function $C(\alpha)$ is smaller than that of the weighted objective function, $A(\alpha)$, for all values of α, provided one limits the local energies. The variances close to the minimum are similar but away from the minimum the variance of A increases much more rapidly than that of C. The smaller variance of C indicates the superior numerical stability of the unweighted function. Qualitatively similarly behaviour occurs for configurations generated with $\alpha_0 = 0$. A commonly used alternative to setting the weights equal to unity is to limit the maximum value of the weights. In figure 5.5 data for objective function C with the largest value of the weights limited to multiples of 1 and 10 times the mean weight is shown, along with data for the weights set to unity. In this graph the standard deviations of the objective functions are plotted as error bars. Figure 5.5 shows that the variance of C is reduced as the weights are more strongly limited, but the lowest variance is obtained by setting the weights to unity. In addition, when the weights are limited the curvature of the objective function is reduced, which makes it more difficult to locate the minimum. Setting the weights to unity therefore gives the best numerical stability.

5.9 Other objective functions

The effect of using the objective function $B(\alpha)$ (equation 5.5), in which the variational energy, E_{V}, is replaced by a fixed reference energy, $\bar{E}$, which is chosen to be lower than the exact energy is considered. In figure 5.6 objective function $B(\alpha)$ versus α is shown for configurations generated with $\alpha_0 = 0$. The overall shapes of the curves are hardly changed as $\bar{E}$ is decreased, although the variance of the objective function slowly increases.

If $\bar{E}$ is chosen to be too low then a significant amount of energy minimisation is included and the numerical stability deteriorates. Using a value of $\bar{E}$ slightly below E_{V} appears to offer no significant advantages.

A direct comparison of the different variance–like objective functions is made in figure 5.7. The behaviour of the following objective functions are displayed: (i) A, (ii) B with $\bar{E} = E_{\mathrm{V}} - 0.3750$ a.u., (iii) C, and (iv) a variant of C with the maximum value of the weights limited to 10 times the mean weight. Limiting outlying values of the local energy improves the behaviour of all the objective functions, so in each case we have limited them according to equation 5.15 with p=8. The mean values of the objective functions are plotted in figure 5.7a, which shows them to behave similarly, with the positions of the minima being almost indistinguishable. However, the curve for the variant of C with limited weights is somewhat flatter, which is an undesirable feature. The standard deviations of the objective functions are plotted in figure 5.7b, and here the differences are more pronounced. The unweighted variance, C, has the smallest variance, which is slightly smaller than that of the variant of C with strongly limited weights. The variances of the objective functions which include the full weights increase rapidly away from α=0. This rapid increase is highly undesirable and could lead to numerical instabilities.

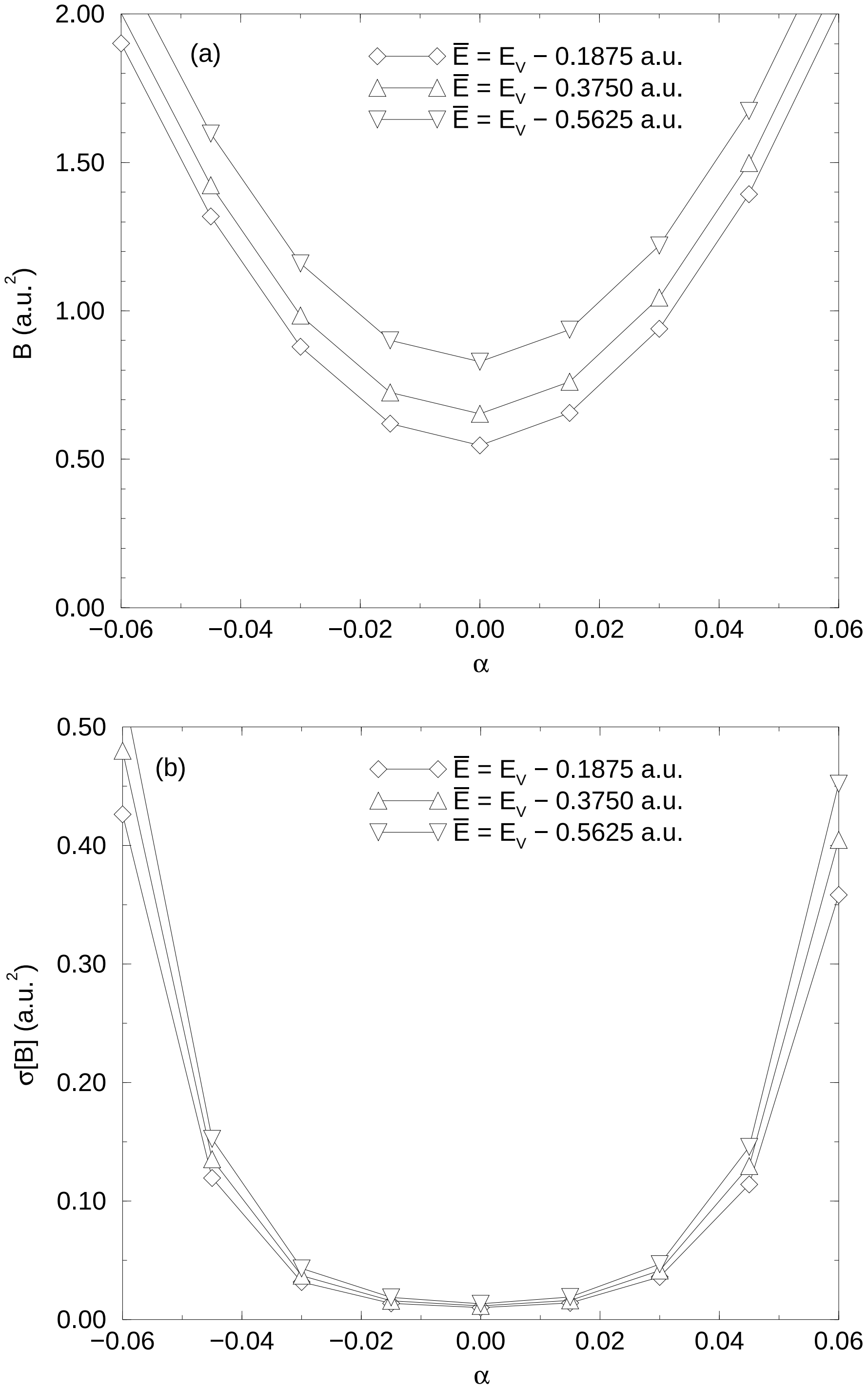

Figure 5.6: Objective function B versus α with $\alpha_0 = 0$ and $\bar{E} < E_{\mathrm{V}}$.

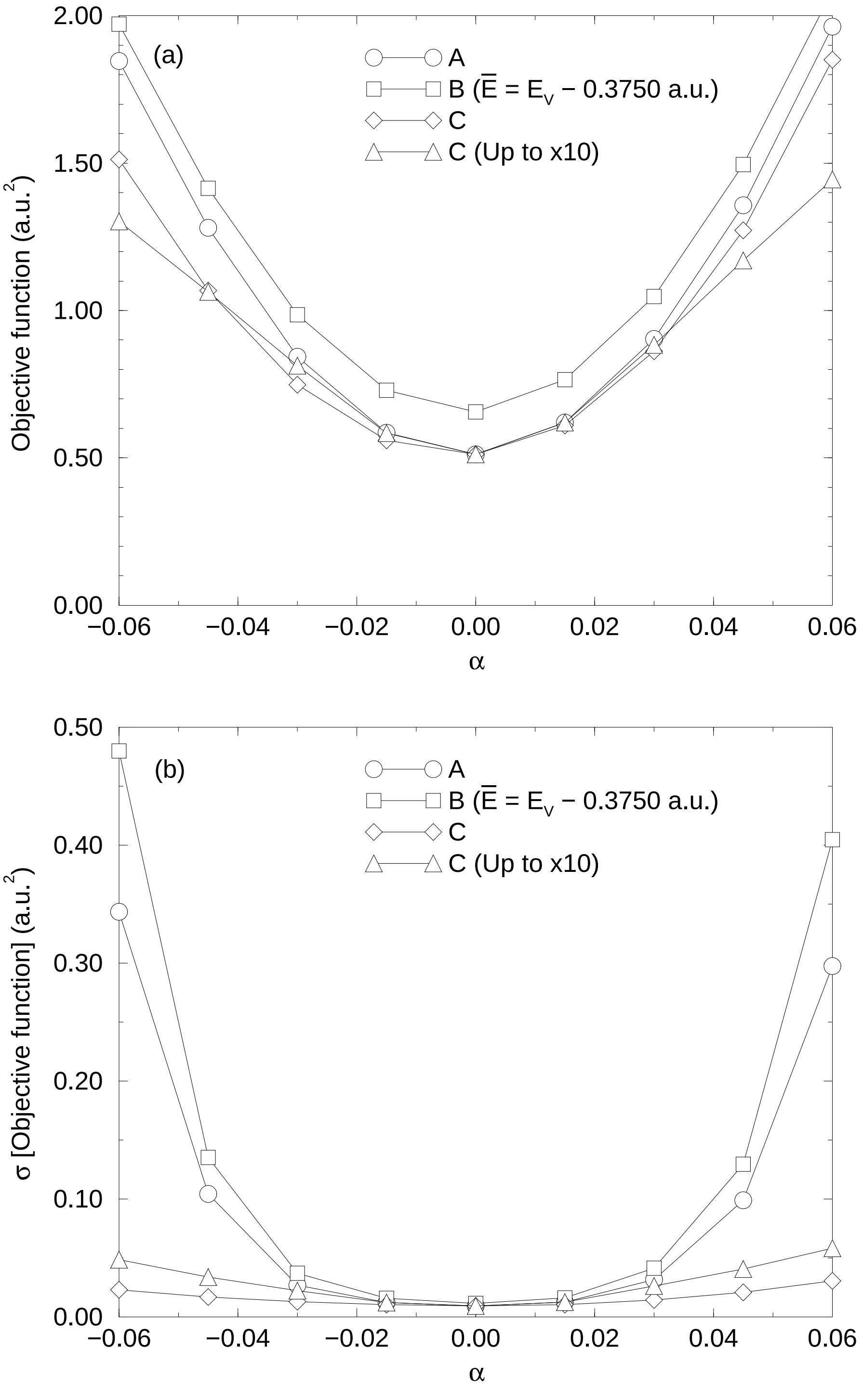

Figure 5.7: Comparison of variance–like objective functions with $\alpha_0 = 0$ as a function of α.

5.10 Conclusions

Several energy and variance minimisation schemes for optimising many–body wavefunctions were examined, where the integrals involved are evaluated statistically. Two reasons why variance minimisation techniques are numerically more stable than energy minimisation techniques were identified: *(i)* In variance minimisation it is allowable to limit the weights or set them equal to unity, which reduces the variance of the objective function while introducing only a "weak bias", which disappears as the process converges. Altering the weights in energy minimisation normally leads to a badly behaved objective function. *(ii)* Variance minimisation, with or without altering the weights, shows greater numerical stability against errors introduced by finite sampling because the positions of the minima of the variance are not shifted by the finite sampling, whereas those of the (properly weighted) energy are.

By investigating a model system, configurations with outlying energies were found to be present and contribute significantly to the variances of the objective functions. The outlying energies result from divergences in the kinetic energy in certain, small, regions of configuration space. Limiting these outlying energies was found to significantly reduce the variance of variance–like objective functions, without significantly altering their value.

The best strategy found for optimising wavefunctions based on these observations is:

1. Minimise the variance of the unweighted local energy (objective function C, equation 5.6). This objective function is robust to finite sampling and has the best statistical properties of the all objective functions examined.

2. Limit outlying values of the local energy according to equation 5.15. Limiting the outlying energies greatly reduces the variance of variance–like objective functions, such as the variance of the unweighted local energy.

3. Regenerate the configurations several times with the updated parameter values until convergence is obtained. Regeneration of the configurations reduces the effects of finite sampling, particularly when the parameters are varied substantially during optimisation.

This strategy may be applied to both ground and excited states of atoms, molecules and solids. The behaviour observed in numerous wavefunction optimizations for large systems, presented later in this thesis, is entirely consistent with the analysis and model system results presented in this chapter. Although for small systems with few parameters other methods may be preferable, the above strategy is expected to be optimal for large systems containing many electrons where efficiency is critical.

This strategy is applied in the later chapters of this thesis, particularly in chapter 8 where wavefunctions are optimised for 13 different carbon clusters. In this application, very high efficiency was important due to the number (13) and size of the clusters (up to 128 valence electrons).

Chapter 6

Finite-size errors in supercell calculations

6.1 Introduction

Most simulations of extended quantum systems are performed using finite simulation cells. This introduces "finite size errors" which are one of the major problems limiting the application of accurate many–body techniques to extended systems. The standard method of reducing the finite size errors is to apply periodic boundary conditions, but important finite size errors often remain. Minimising and quantifying the size of these errors is essential where high accuracy is required.

In this chapter, developments of the theory of finite size effects in quantum many–body simulations subject to periodic boundary conditions are presented. The motivation is to understand and reduce the finite size effects encountered in quantum Monte Carlo simulations, although the techniques described are of wide generality and can be readily applied to other many–body electronic structure methods. The work contained in this chapter is an extension of earlier work by Fraser *et al.* [86] and Williamson*et al.* [87] A more thorough analysis of the finite size effects in presented, including application to excitation energies. The theory is applied to large silicon supercells of up to 250 atoms (1000 valence electrons) in DFT, HF, VMC and DMC. These calculations are computationally very demanding.

By more accurately modelling the infinite system, the techniques developed in this chapter enable far greater accuracy to be obtained at lower computational cost than would otherwise be possible.

6.2 Finite size errors in many–body calculations

Finite size errors in many–body calculations have traditionally been corrected for by extrapolation techniques and/or by using the results of more approximate calculations, such as those based on density functional theory (DFT). The finite size errors are typically fitted to a lower order polynomial. While extrapolation techniques can certainly be successful they are very costly. In addition, the finite size errors in periodic boundary condition DFT calculations are significantly different from those in the standard formulations of quantum many–body techniques such as QMC, and therefore a simple application of a DFT finite size correction may not lead to accurate results.

In this chapter a way to reduce the finite size errors in quantum many–body simulations is presented. A key step in this process is the use of a new periodic interaction instead of the conventionally used Ewald interaction. This interaction more accurately describes the interactions in the infinite system than the Ewald potential, which is designed to model the artificial array of supercells. The new “Model Periodic Coulomb” (MPC) interaction also has the additional advantage that the residual finite size effects are reasonably well described by standard DFT calculations. The accuracy can be further increased by using an extrapolation procedure, but the extrapolation corrections are considerably reduced and can therefore be evaluated using a smaller range of system sizes.

In section 6.3 the Hamiltonian within periodic boundary conditions is described. In section 6.4 various finite size correction and extrapolation procedures that have previously been used are presented. In section 6.6 the “independent particle finite size effects” review, showing how k–space sampling techniques in many–body theory are related to those used in mean–field theories. The MPC interaction for reducing finite size effects in periodic systems is introduced in section 6.7. It is shown that it can be applied to all the Coulomb interactions. The MPC interaction is tested within HF theory (section 6.8.1), VMC (section 6.8.2), and DMC (section 6.8.3).

In section 6.9 the finite size errors present in calculations of excitation energies are discussed. HF and VMC calculations of excitation energies using the MPC interaction are presented in sections 6.9.1 and section 6.9.2. VMC results for the “optical absorption” and “photoemission” gaps are presented, the latter being the first such calculations for a three–dimensional periodic system. Calculations with up to 512 electrons demonstrate that the finite size errors in the “optical absorption” and “photoemission” gaps are similar and that the finite size effects are quite small even for a 64–electron simulation cell. Conclusions are given in section 6.10.

6.3 The Hamiltonian within periodic boundary conditions

The many–body Hamiltonian for a system of electrons at positions $\mathbf{r}_i$ and static ions at positions $\mathbf{x}_\alpha$ is

$$\begin{aligned} \hat{H} &= \sum_i -\frac{1}{2}\nabla_i^2 + \sum_i \sum_\alpha v_\alpha(\mathbf{r}_i, \mathbf{x}_\alpha) \\ &+ \frac{1}{2}\sum_i \sum_{j \neq i} v(\mathbf{r}_i, \mathbf{r}_j) + \frac{1}{2}\sum_\alpha \sum_{\beta \neq \alpha} v_{\alpha\beta}(\mathbf{x}_\alpha, \mathbf{x}_\beta) \quad . \end{aligned} \tag{6.1}$$

An infinite system is normally modeled by a finite simulation cell subject to periodic boundary conditions. The model interaction terms, v_α, v, and $v_{\alpha\beta}$, are chosen such that the potential energy of the model system, which involves only the positions of the particles in the finite simulation cell, mimics the potential energy of the infinite system as closely as possible. Since the potential energy of the infinite system depends on the positions of all the charges in the solid, only a few of which are included in the simulation, the model interaction energy is approximate even in crystalline solids. To enforce the periodic boundary conditions the functions v_α, v, and $v_{\alpha\beta}$ must be invariant under the translation of either argument by a member of the set of translation vectors of the simulation cell lattice, $\{\mathbf{R}\}$. The standard approach is to choose the model Hamiltonian such that the full potential energy of equation 6.1, evaluated by summing the model interactions between all pairs of particles in the simulation cell, equals the potential energy per cell of an infinite array of identical copies of the simulation cell. However, even when each simulation cell in the array of copies is restricted to be overall charge neutral, the sum of inter–particle Coulomb $1/r$ interactions is only conditionally convergent, [76] and to define this model interaction uniquely the boundary conditions at infinity must be specified. The standard procedure is to define the potential by solving Poisson's equation subject to periodic boundary conditions, in which case the model interaction is the Ewald interaction. [51] Further details of the Ewald interaction and the supercell approach were given in chapter 4.

6.4 Finite size correction and extrapolation formulae

The basic idea behind finite size correction formulae is to write the energy for the infinite system as

$$E_\infty = E_N + (E_\infty - E_N) \ , \tag{6.2}$$

where the subscript denotes the system size. A highly accurate many–body calculation for the N–particle system is then performed to obtain an approximation for E_N, and the correction term in brackets is approximated using a much less expensive scheme which can be applied to very large systems:

$$E_\infty \simeq E_N + (E'_\infty - E'_N) \; , \tag{6.3}$$

where the prime indicates that a less expensive scheme is used. In an extrapolation procedure the energy E_N is calculated for a range of system sizes and fitted to a chosen functional form which contains parameters. Correction and extrapolation procedures can be combined to give an expression

$$E_\infty \simeq E_N + (E'_\infty - E'_N) + F(N) \; , \tag{6.4}$$

where the prime again indicates that a less expensive scheme is used and $F(N)$ is an extrapolation function. Clearly, the optimal form of $F(N)$ depends on the method used to calculate the correction term. The practical difference between correction and extrapolation is that correction requires a single calculation of E_N using the accurate (and normally very costly) many–body technique, while extrapolation requires several such calculations for different values of N and a subsequent fit to the chosen functional form $F(N)$. The extrapolation procedure is costly because it involves more calculations and is prone to inaccuracy because one has to perform a fit with only a few data points. In designing a correction/extrapolation procedure one therefore tries to make the extrapolation term as small as possible.

Candidate methods for evaluating the correction term include Kohn–Sham DFT and HF theory. The most convenient methods are the local density approximation (LDA) to DFT, or extensions such as the Generalized Gradient Approximation. These methods are very widely applied in periodic boundary conditions calculations and are computationally inexpensive, while retaining a realistic description of the system. Within an independent particle theory, such as Kohn–Sham DFT, calculations for periodic systems are normally performed by solving within the primitive unit cell.[1] To obtain the correct result for the infinite system it is necessary to integrate over k–space, and the integral is normally approximated by a sum over a finite set of k–points. A determinant formed from the occupied orbitals at a single k–point in the first Brillouin zone (BZ) of the primitive unit cell is a many–body wave function for a simulation cell of the size of the primitive unit cell. Adding a second k–point doubles the size of the determinant and is equivalent to doubling

[1] The reduction of the problem to one within the primitive unit cell is only possible within independent particle theories; in a true many–body theory one has to solve over the entire simulation cell.

the size of the simulation cell, etc.

An important subtlety arises when finite size corrections derived from an independent particle theory such as LDA–DFT are used to correct the results of a true many–body method such as QMC. Suppose the many–body calculation is performed using the Ewald interaction, so that the Coulomb energy is that of an infinite periodic array of copies of the simulation cell. Given that the solid we are trying to model is crystalline, and hence the charge density is truly periodic, the Ewald interaction gives a good description of the classical Coulomb or Hartree energy. However, because the electronic positions are mirrored exactly in every copy of the simulation cell, the *electronic correlations* are also forced to be periodic, and the exchange–correlation (XC) energy corresponds to a system with a *periodic* XC hole. This unphysical approximation is particularly inaccurate when the simulation cell is small. In Kohn–Sham DFT calculations the XC energy is evaluated using a standard functional such as the LDA of Perdew and Zunger [88]. This was obtained by fitting to the results of DMC calculations for jellium, [13] which were extrapolated to the infinite–system–size limit *prior* to performing the fit.

The consequences of this difference may be illustrated by considering many–body and LDA calculations for jellium, in which the charge density is uniform. The LDA gives the XC energy for the infinite system irrespective of the size of the simulation cell, but the XC energy obtained from a many–body simulation using the Hamiltonian of equation 6.1 with the Ewald interaction gives an XC energy which depends on the size and shape of the simulation cell. Consequently, evaluating the correction term in equation 6.4 within the LDA does not give a good approximation to the finite size correction in the many–body simulation and a significant extrapolation term remains.

6.4.1 Previous work

Several methods for computing finite size corrections in many–body calculations occur in the literature.

The issue of finite size corrections to both the kinetic and interaction energies has been addressed by Ceperley and coworkers. [33, 89–91] Their approach is to add separate extrapolation terms for the kinetic and interaction energies. In their work on hydrogen solids, Ceperley and Alder [90] performed DMC calculations for a number of different system sizes and fitted to the formula

$$E_\infty^{\mathrm{DMC}} \simeq E_N^{\mathrm{DMC}} + a(T_\infty - T_N) + \frac{b}{N} \quad , \tag{6.5}$$

where a and b are parameters, and T is the kinetic energy of the non–interacting electron gas. The b/N term accounts for the finite size effects arising from the interaction energy and the difference of the parameter a from unity accounts for the difference between the kinetic energies of the interacting and non–interacting systems. Normally $a > 1$ because the interacting kinetic energy is larger than the non–interacting kinetic energy. Engel *et al.* [92] used the following formula for inhomogeneous systems:

$$E_{\infty}^{\mathrm{QMC}} \simeq E_{N}^{\mathrm{QMC}} + a(E_{\infty}^{\mathrm{LDA}} - E_{N}^{\mathrm{LDA}}) + \frac{b}{N} \ , \tag{6.6}$$

which reduces to equation 6.5 for a homogeneous system.

6.5 General approach

A large part of the finite size error in the interaction energy arises from the use of the Ewald interaction. [86, 87] The general approach to the problem of the finite size errors taken here differs from that of Ceperley *et al.* in an important respect. Here the finite size effects are reduced within the *many–body* calculation by *modifying the interaction terms* in the many–body Hamiltonian.

The Ewald interaction is a periodic function which differs from $1/r$ in such a way that the sum of interactions between pairs of particles *within one cell* gives the exact Coulomb energy per cell of a periodic lattice of identical cells. This ensures that the Ewald Hamiltonian gives the correct Hartree energy, but the deviations from $1/r$ give rise to a spurious contribution to the XC energy, corresponding to the periodic repetition of the XC hole. [86, 87] A solution is to change the many–body Hamiltonian so that the interaction with the XC hole is exactly $1/r$ (see section 6.7) for any size and shape of simulation cell, without altering the form of the Hartree energy. This source of finite size error is therefore eliminated and equation 6.3, with the correction term calculated within a the LDA, gives a much better description of the remaining finite size errors. Greater accuracy can be obtained by adding an extrapolation term, but this term is much smaller than if the Ewald interaction is used.

The utility of changing the Hamiltonian to reduce the finite size errors is readily seen by considering an example. Suppose that a periodic boundary condition QMC technique is used to study an isolated molecule. If the molecule is placed at the center of the periodically repeated simulation cell, the calculated energy is that of an array of identical molecules, including unwanted inter–molecular interactions. The results improve as the simulation cell is made larger, but the convergence is slow, especially for molecules with permanent dipole moments. A better solution is to cut off all Coulomb interactions between charges on different molecules, i.e., to replace the Ewald interactions by truncated Coulomb $1/r$ interactions acting only within the simulation cell. As long as the molecular wave function has decayed to zero before the simulation cell boundary is reached, this procedure should give essentially exact results.

An alternative procedure is to evaluate the correction term in equation 6.3 using HF data. HF theory is an approximate method for solving the many–body Hamiltonian, and if the Ewald formula is used for the electron–electron interaction terms in both the many–body and HF theories, the finite size error in the HF exchange energy will tend to cancel the finite size error in the many–body interaction energy. However, this procedure gives too large a correction, presumably because the HF exchange hole is significantly different from the screened XC hole of the many–body system.

6.6 Independent particle finite size effects

The "independent particle finite size effect" (IPFSE) is defined by the correction term $(E_\infty^{\rm LDA} - E_N^{\rm LDA})$. The finite size effect arising from the particular model interaction used is then the remaining "Coulomb finite size effect" (CFSE). In this section the IPFSE is discussed in detail. Particular attention is paid to k–point selection, or BZ sampling.

The "finite size error" in an LDA calculation for a perfect crystal arises from errors in the BZ integration. In an LDA calculation the computational cost is proportional to the number of k–points used. In a QMC calculation the volume of the simulation cell is proportional to the number of k–points in the primitive BZ. The computational cost increases approximately as the cube of the volume of the simulation cell, meaning that the selection of k–points is even more critical in QMC calculations.

Additional errors arise when using the supercell approximation for systems which break translational symmetry. Consider a crystal containing a point defect. In the supercell approach a finite simulation cell containing a single defect is repeated throughout space to form an infinite three–dimensional array. The supercell must be sufficiently large that the interaction between defects in different cells is negligible. This type of finite size effect is distinct from the BZ integration error because it persists even when the BZ integration is performed exactly. However, the correction to the defect formation energy from this type of finite size error should be described reasonably well by a $(E_\infty^{\rm LDA} - E_N^{\rm LDA})$ correction term, where $E_\infty^{\rm LDA}$ and $E_N^{\rm LDA}$ are, respectively, the defect formation energies for a large simulation cell and a smaller one, each containing a single defect.

A determinantal Bloch wave function suitable for use in a QMC calculation may be formed from a set of single particle orbitals at a single k–point, $\mathbf{k}_{\rm s}$, in the Brillouin zone of the simulation cell reciprocal lattice. The IPFSE can be greatly reduced for insulating systems by a careful choice of $\mathbf{k}_{\rm s}$ using the method of "special k–points" [93, 94] borrowed from band structure theory. [95, 96] Baldereschi [95] defined the "mean–value point", a k–point at which smooth periodic functions of wave vector accurately approximate their averages over the BZ, and clearly the Baldereschi mean–value point is a strong candidate for $\mathbf{k}_\mathrm{s}$. It is convenient to choose $\mathbf{k}_\mathrm{s}$ to be equal to half a translation vector of the simulation cell reciprocal lattice ($\mathbf{k}_\mathrm{s} = \mathbf{G}/2$), which allows the construction of real single–particle orbitals and hence is computationally more efficient. Some freedom is still left in the selection of $\mathbf{k}_\mathrm{s}$ and the symmetrized plane wave test of BZ integration quality introduced by Baldereschi [95] provides an accurate guide to final selection. For example, for an fcc simulation cell the half–reciprocal–lattice vectors correspond to the Γ, X and L points of the BZ of the simulation cell reciprocal lattice. For a crystal with the full cubic symmetry $\mathbf{k}_\mathrm{s} \equiv L$ gives the best BZ integration and $\mathbf{k}_\mathrm{s} \equiv \Gamma$ the worst.

6.7 Coulomb finite size effects

In this section the CFSE is considered and the theory of the MPC interaction is described. These effects do not occur in single–particle theories and are more difficult to compensate for than the IPFSE.

The origin of the CFSE was described in section 6.4 as arising from the XC energy and the dependence of the Ewald interaction, v_E, on the size and shape of the simulation cell. Expanding v_E around zero separation gives [86, 87]

$$v_\mathrm{E}(\mathbf{r}) \quad = \quad \text{constant} + \frac{1}{r} + \frac{2\pi}{3\Omega}\,\mathbf{r}^T \cdot \mathbf{D} \cdot \mathbf{r} + \mathcal{O}\left(\frac{r^4}{\Omega^{5/3}}\right) \quad , \tag{6.7}$$

where Ω is the volume of the simulation cell, and the tensor $\mathbf{D}$ depends on the shape of the simulation cell. (For a cubic cell $\mathbf{D}$=$\mathbf{I}$.) The constant term arises from the condition that the average of v_E over the simulation cell is zero. The term of order r^2 and the higher order deviations from $1/r$ make the Ewald interaction periodic and ensure that the sum of interactions between the particles in the simulation cell gives the potential energy per cell of an infinite periodic lattice. These terms are responsible for the spurious contribution to the XC energy, which is the source of the large finite size effect in many–body calculations using the Ewald interaction. [86, 87] For cubic cells the interaction at short distances is larger than $1/r$ and therefore the XC energy is more negative than it should be, and because the leading order correction is proportional to the inverse of the simulation cell volume the error per electron is inversely proportional to the number of electrons in the cell.

Clearly it is desirable to remove this spurious contribution to the XC energy, but the Hartree energy is correctly evaluated using the Ewald interaction. The key requirements for a model Coulomb interaction giving small CFSEs in simulations with periodic boundary conditions are therefore: (i) it should give the Ewald interaction for the Hartree terms, and (ii) it should be exactly $1/r$ for the interaction with the XC hole. Unfortunately, the only periodic solution of Poisson's equation for a periodic array of charges is the Ewald interaction, which obeys criterion (ii) only in the limit of an infinitely large simulation cell. The "Model Periodic Coulomb" (MPC) interaction satisfies both criteria, but does not satisfy Poisson's equation.

The MPC interaction can be written as [87]

$$\begin{aligned}\hat{H}_{\mathrm{e-e}} &= \sum_{i>j} f(\mathbf{r}_i - \mathbf{r}_j) \\ &+ \sum_i \int_{\mathrm{WS}} \left[v_{\mathrm{E}}(\mathbf{r}_i - \mathbf{r}) - f(\mathbf{r}_i - \mathbf{r})\right] \rho(\mathbf{r})\, \mathrm{d}\mathbf{r} \quad , \end{aligned} \tag{6.8}$$

where ρ is the electronic charge density and

$$f(\mathbf{r}) = \frac{1}{r_{\mathrm{m}}} \quad . \tag{6.9}$$

The definition of the cutoff Coulomb function, f, involves a minimum image convention whereby the inter–electron distance, $\mathbf{r}$, is reduced into the Wigner–Seitz (WS) cell of the simulation cell lattice by removal of simulation cell lattice translation vectors, leaving a vector $\mathbf{r}_{\mathrm{m}}$. This ensures that $\hat{H}_{\mathrm{e-e}}$ has the correct translational and point group symmetry. The first term in equation 6.8 is a direct Coulomb interaction between electrons within the simulation cell and the second term is a sum of potentials due to electrons "outside the simulation cell". Note that the second term is a one–body potential similar to the Hartree potential. It depends on the electronic charge density, ρ, but is not a function of inter–particle separation.

The electron–electron contribution to the total energy is evaluated as the expectation value of $\hat{H}_{\mathrm{e-e}}$ with the many-electron wave function, ϕ, minus a double counting term:

$$E_{\mathrm{e-e}} = \langle \phi | \hat{H}_{\mathrm{e-e}} | \phi \rangle - \frac{1}{2} \int_{\mathrm{WS}} \rho(\mathbf{r}) \rho(\mathbf{r}') \left[v_{\mathrm{E}}(\mathbf{r} - \mathbf{r}') - f(\mathbf{r} - \mathbf{r}')\right] \, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}' \quad . \tag{6.10}$$

Evaluating the expectation value gives

$$\begin{aligned} E_{\mathrm{e-e}} &= \frac{1}{2} \int_{\mathrm{WS}} \rho(\mathbf{r}) \rho(\mathbf{r}') v_{\mathrm{E}}(\mathbf{r} - \mathbf{r}')\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}' \\ &+ \left(\int_{\mathrm{WS}} |\phi|^2 \sum_{i>j} f(\mathbf{r}_i - \mathbf{r}_j)\, \Pi_k\, \mathrm{d}\mathbf{r}_k - \frac{1}{2} \int_{\mathrm{WS}} \rho(\mathbf{r}) \rho(\mathbf{r}') f(\mathbf{r} - \mathbf{r}')\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}' \right) \quad , \end{aligned} \tag{6.11}$$

where the first term on the right hand side is the Hartree energy and the term in brackets is the XC energy. We can see immediately that the Hartree energy is calculated with the Ewald interaction while the XC energy (expressed as the difference between a full Coulomb term and a Hartree term) is calculated with the cutoff interaction f.

The charge density ρ appears in equations 6.8, 6.10, and 6.11. In QMC methods, the charge density is known with greatest statistical accuracy at the end of the calculation. This is not a serious complication for VMC simulations as the interaction energy may be evaluated at the end of the simulation using the accumulated charge density. In DMC this is not possible because the local energy is required at every step. This point is investigated in section 6.8.3.

6.7.1 Systems of electrons and nuclei

In the previous sections only the electron–electron interaction was considered, not the electron–nucleus and nucleus–nucleus interactions. If the MPC interaction is physically reasonable it should be possible to apply it to all the interactions in the problem, not just the electron–electron terms. For systems of electrons and nuclei, under quite general conditions, the expressions for the MPC interaction simplify so that the electron–nucleus and nucleus–nucleus terms reduce to the Ewald form.

Consider a simulation cell with periodic boundary conditions containing N electrons at positions $\mathbf{r}_i$ and M nuclei of charge Z_α at positions $\mathbf{x}_\alpha$. The wave function of the electrons and nuclei is $\Psi(\{\mathbf{r}_i\}, \{\mathbf{x}_\alpha\})$, and the total charge density (electrons and nuclei) is $\rho_{\mathrm{T}}(\mathbf{r})$. The interaction energy calculated with the MPC interaction is

$$\begin{aligned} E_{\mathrm{int}} &= \frac{1}{2}\int_{\mathrm{WS}} \rho_{\mathrm{T}}(\mathbf{r})\rho_{\mathrm{T}}(\mathbf{r}')[v_{\mathrm{E}}(\mathbf{r}-\mathbf{r}') - f(\mathbf{r}-\mathbf{r}')]\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \\ &+ \int_{\mathrm{WS}} |\Psi|^2 \left[\sum_{i>j} f(\mathbf{r}_i - \mathbf{r}_j) - \sum_i \sum_\alpha Z_\alpha f(\mathbf{r}_i - \mathbf{x}_\alpha)\right. \\ &+ \left.\sum_{\alpha>\beta} Z_\alpha Z_\beta f(\mathbf{x}_\alpha - \mathbf{x}_\beta)\right] \Pi_k\,\mathrm{d}\mathbf{r}_k\,\Pi_\gamma\,\mathrm{d}\mathbf{x}_\gamma \quad . \end{aligned} \tag{6.12}$$

We now employ the adiabatic approximation to separate the electronic and nuclear dynamical variables:

$$\Psi(\{\mathbf{r}_i\}, \{\mathbf{x}_\alpha\}) = \phi(\{\mathbf{r}_i\}; \{\mathbf{x}_\alpha\})\Phi(\{\mathbf{x}_\alpha\}) \quad , \tag{6.13}$$

where the $\{\mathbf{x}_\alpha\}$ appear only as parameters in ϕ. To make further progress a form for the nuclear part of the wave function, Φ, must be assumed. The simplest assumption is that Φ can be written as an appropriately symmetrized product of very strongly localized non–overlapping single–nucleus functions. Regardless of whether the product is antisymmetrized, symmetrized, or not symmetrized, equation 6.12 reduces to

$$\begin{aligned} E_{\rm int} &= \frac{1}{2}\int_{\rm WS} \rho(\mathbf{r})\rho(\mathbf{r}')[v_{\rm E}(\mathbf{r}-\mathbf{r}') - f(\mathbf{r}-\mathbf{r}')]\,{\rm d}\mathbf{r}\,{\rm d}\mathbf{r}' \\ &+ \int_{\rm WS} |\phi|^2 \sum_{i>j} f(\mathbf{r}_i - \mathbf{r}_j)\,\Pi_k\,{\rm d}\mathbf{r}_k \\ &- \int_{\rm WS} |\phi|^2 \sum_i \sum_\alpha Z_\alpha\, v_{\rm E}(\mathbf{r}_i - \overline{\mathbf{x}}_\alpha)\,\Pi_k\,{\rm d}\mathbf{r}_k \\ &+ \sum_{\alpha>\beta} Z_\alpha Z_\beta\, v_{\rm E}(\overline{\mathbf{x}}_\alpha - \overline{\mathbf{x}}_\beta) \quad , \end{aligned} \tag{6.14}$$

where the $\overline{\mathbf{x}}_\alpha$ denote the centers of the single–nucleus functions and ρ is the electron density. Note that the electron–nucleus and nucleus–nucleus terms involve only the Ewald interaction and that the first two terms of equation 6.14 correspond precisely to the electron–electron interaction of equation 6.11. This result justifies the use of equation 6.8 for the electron–electron interactions while retaining the Ewald interaction for the electron–nucleus and nucleus–nucleus terms.

6.8 Tests of the MPC interaction

The MPC interaction was tested by performing calculations using HF theory, VMC and DMC. LDA-DFT was used for the IPFSE corrections. A series of calculations on diamond structure silicon using fcc simulation cells of increasing size. Cells whose translation vectors are n times those of the primitive unit cell were used.

6.8.1 Application within HF theory

In an initial set of tests LDA and HF results were compared. These tests are inexpensive and are not subject to statistical errors because Monte Carlo techniques are not involved. Simulation cells with n = 1, 2, 3, 4, and 5, which contain 2, 16, 54, 128, and 250 ions, respectively were considered. The Si^{4+} ions were represented by norm–conserving non–local LDA pseudopotentials and the calculations were performed using a plane–wave basis set with a cutoff energy of 15 Ry.

To measure the IPFSE and the effect of k–point selection, calculations were performed using $\mathbf{k}_s \equiv L$ and $\mathbf{k}_s \equiv \Gamma$. To facilitate comparison the HF energy was evaluated non-consistently with the LDA orbitals, so that the energy differences arise solely from the difference between the LDA XC energy and the HF exchange energy.

The HF energy evaluated with the MPC interaction is

$$
\begin{aligned}
E^{\mathrm{HF}} &= \sum_i -\frac{1}{2}\int \phi_i^*(\mathbf{r})\nabla^2\phi_i(\mathbf{r})\,\mathrm{d}\mathbf{r} \\
&+ \frac{1}{2}\int \rho(\mathbf{r})\rho(\mathbf{r}')v_{\mathrm{E}}(\mathbf{r}-\mathbf{r}')\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \\
&- \frac{1}{2}\sum_{i,j}^{N}\delta_{s_is_j}\int \phi_i^*(\mathbf{r})\phi_i(\mathbf{r}')f(\mathbf{r}-\mathbf{r}')\phi_j^*(\mathbf{r}')\phi_j(\mathbf{r})\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \\
&+ \int V_{\mathrm{ext}}(\mathbf{r})\rho(\mathbf{r})\,\mathrm{d}\mathbf{r} + E_{\mathrm{I-I}} \quad , \qquad (6.15)
\end{aligned}
$$

where $E_{\mathrm{I-I}}$ is the ion–ion energy calculated with the Ewald interaction. Note that the Hartree energy is evaluated with the Ewald interaction while the exchange energy is evaluated with f, the cutoff Coulomb interaction.

In figure 6.1 the deviations of the LDA and HF energies from the fully converged values as a function of system size for $\mathbf{k}_s \equiv \Gamma$ wave functions are shown. The LDA energy converges smoothly with system size but for small system sizes the IPFSE error is large because of the Γ–point sampling. HF energies calculated with the Ewald and MPC interactions, with and without incorporating the IPFSE corrections obtained from the LDA data are shown. The data incorporating the IPFSE (filled symbols) show the residual CFSE errors. The IPFSE is positive while the CFSE is negative, in accord with the analysis of the CFSE given in section 6.7. The IPFSE corrected data show that the CFSE for the MPC interaction is roughly half that for the Ewald interaction.

In figure 6.2 similar data for $\mathbf{k}_s \equiv L$ is shown. The LDA energy converges rapidly and smoothly with system size, and therefore the IPFSE is small, which demonstrates the efficacy of the "special k–points" method. The IPFSE and CFSE errors tend to cancel and for $\mathbf{k}_s \equiv L$ sampling the HF data converge more rapidly without the IPFSE corrections. For $n = 3$, corresponding to a 54 atom simulation cell, the LDA finite size error (IPFSE only) is 0.011 eV per atom, which is very much smaller than the HF (Ewald) finite size error of -0.211 eV per atom or the equivalent HF (MPC) error of -0.071 eV per atom. As in the case of Γ–point sampling, the CFSE for the MPC interaction is roughly half that for the Ewald interaction.

After applying the IPFSE correction obtained from the LDA data the HF results for $\mathbf{k}_s \equiv \Gamma$ and $\mathbf{k}_s \equiv L$ sampling are very similar. The correspondence of the IPFSE corrected data for the L– and Γ–points demonstrates that to a very good approximation the IPFSE and CFSE are independent. Estimating the energy of the infinite system by averaging the energy over a set of $\mathbf{k}_s$ vectors removes the IPFSE but does not remove the CFSE.

The CFSE errors from the filled data points in figures 6.1 and 6.2 were fitted to the form b/N^x, where b and x are parameters and $N = 8n^3$ is the number of electrons in the simulation cell. The fits give values of x in the region of unity for both the Ewald and MPC interactions. This extrapolation function is therefore suitable for both interactions, although the size of the extrapolation term is smaller for the MPC interaction.

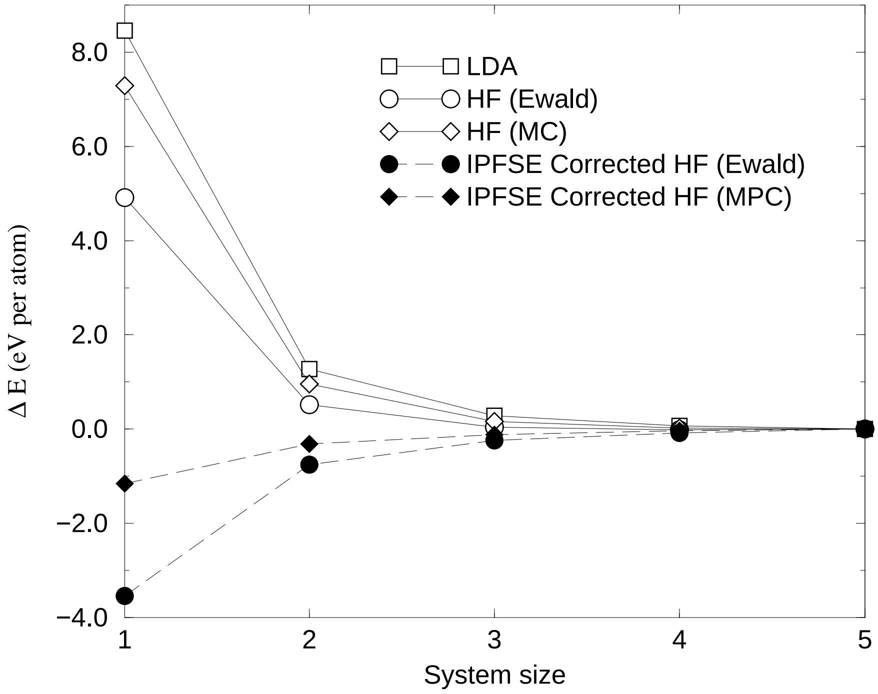

Figure 6.1: Convergence of the LDA and HF ground state energies per atom of silicon as a function of simulation cell size, n, using Γ–point BZ sampling.

6.8.2 Application within VMC

Single determinant wave functions of Slater-Jastrow form, as described in chapter 4, were optimised using both the Ewald and MPC interactions. 32 variable parameters were optimised, and the resultant wave functions obtained approximately 90% of the fixed-node correlation energy. The same pseudopotential as for the LDA and HF calculations was used.

The wave functions optimised using the different interactions were almost identical. Properties other than the energy, such as pair correlation functions, are therefore hardly affected by the choice of interaction. As the MPC interaction gives the correct interaction between the electrons at short distances it may give a better account of, for example, the short distance behaviour of the pair correlation function, but this point was not investigated.

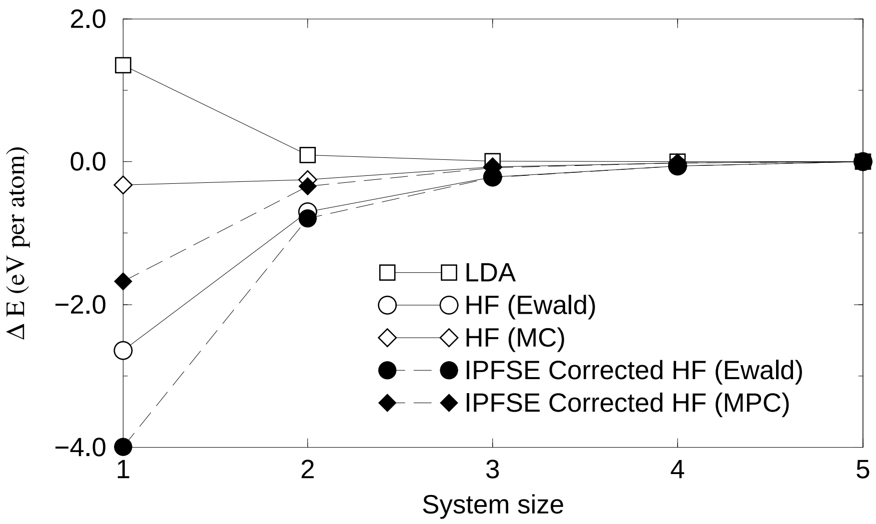

Figure 6.2: The LDA and HF ground state energies per atom of silicon as a function of simulation cell size, n, using L–point BZ sampling.

In figure 6.3 VMC results for energies of the same systems as in the previous section are shown.[2] The total energies were calculated to a statistical accuracy of ± 0.01 eV per atom. The VMC data display a smaller CFSE than the HF data, probably because the HF exchange hole is longer ranged than the screened hole obtained in the correlated calculations. For $n = 2$, the MPC interaction reduces the VMC finite size error by more than 50%, from -0.403 to -0.187 eV per atom.

6.8.3 Application within DMC

The MPC interaction is more complicated to apply in DMC than in VMC because the evaluation of the importance sampled Green's function requires the local energy. The modified Hamiltonian, and hence the local energy, depends on the charge density, and therefore the charge density must be known before the DMC calculation. Fortunately the "exact" DMC charge density is not required because the local energy is relatively insensitive to the charge density used in the Hamiltonian (equation 6.8) because $v_{\rm E}(\mathbf{r}) - f(\mathbf{r})$ is small when $|\mathbf{r}|$ is small.

Two candidate charge densities are the charge density of the determinantal part of the QMC wave function and the charge density of the VMC guiding wave function. Even for a small system ($n = 2$), where the potential errors are largest, it was found to be

[2] These results are similar to those given in Ref. [87], but they have been fully recalculated and corrected.

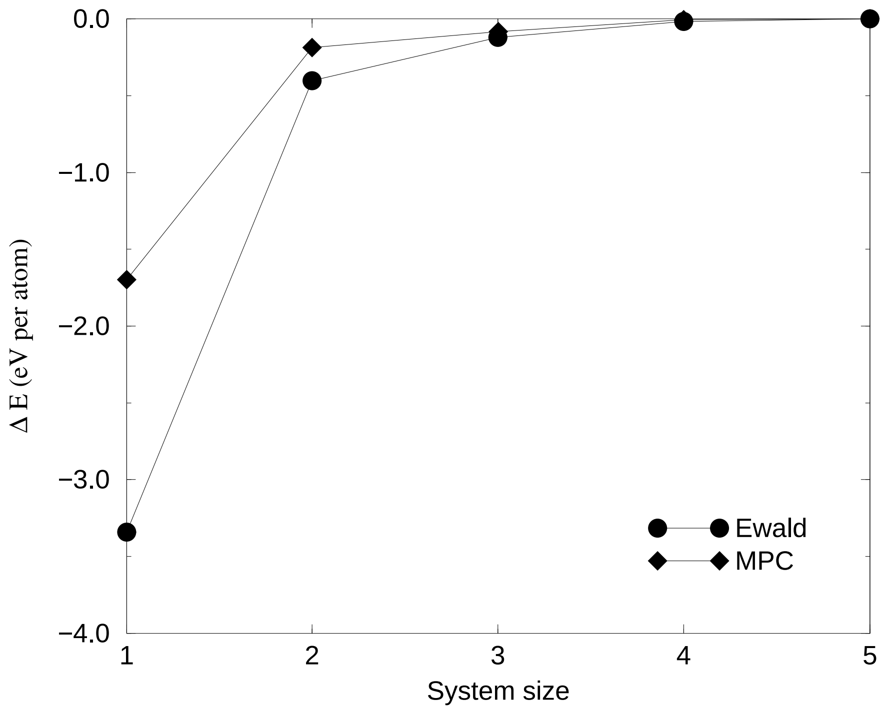

Figure 6.3: The VMC ground state energies per atom of silicon as a function of simulation cell size, n. A correction for the IPFSE is included. The statistical error bars are ± 0.01 eV per atom.

sufficient to use the LDA charge density during the calculation of the Green's function and to re–evaluate the charge density dependent term in the interaction energy using the DMC density obtained at the end of the calculation. The insensitivity of the Green's function to the charge density used in the Hamiltonian was tested by running calculations with LDA, VMC and "self-consistent" DMC densities. The sensitivity rapidly reduces with increasing system size. Using an LDA density gives errors of less than 0.03 eV per atom for $n = 2$, and less than 0.01 eV per atom for larger system sizes. Therefore, the requirement of having a good approximation to the charge density in advance of the DMC calculation does not pose a significant difficulty. A successful DMC calculation requires a good quality VMC trial function and its charge density can be obtained during the process of wave function optimization.

The DMC calculations were performed using the optimised wavefunctions of the previous section. A time step of 0.01 a.u. was used. Li *et al.* [97] found that using a time step of 0.015 a.u. gave a time–step error of less than 0.03 eV per atom in silicon, so that our time step error should be even smaller. Total energies were calculated to a statistical accuracy of ± 0.02 eV per atom.

In figure 6.4 the results of DMC calculations are shown. The results include a correction for the IPFSE. The convergence behaviour is very similar to the VMC data. The MPC energies are always above those for the Ewald interaction and the MPC interaction significantly reduces the CFSE. These results demonstrate that the finite size errors within VMC and DMC calculations are very similar and that the MPC interaction is similarly effective in both methods, although a little more complex to apply in DMC.

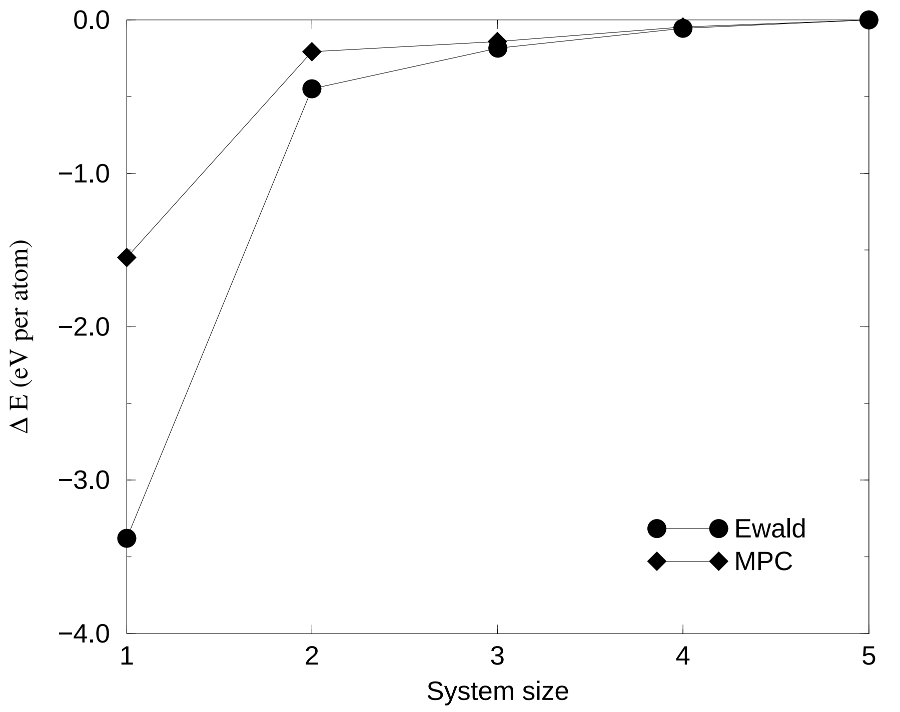

Figure 6.4: The DMC ground state energies per atom of silicon as a function of simulation cell size, n. A correction for the IPFSE is included. The statistical error bars are ± 0.02 eV per atom.

The residual CFSE errors in both the VMC and DMC calculations are reasonably fit by a function to the form b/N^x. The fit gives a value of x close to unity for both the Ewald and MPC data. For these data it is possible to obtain more accurate approximations to E_∞^{QMC} by using such an extrapolation function, although the calculations for the large system sizes are costly, especially within DMC. The extrapolated energies should be more accurate for the MPC interaction because the extrapolation term is significantly smaller. The quality of the fit and hence extrapolation cannot be accurately determined due to the limited number of data points and statistical errors present in the data.

Many interesting applications of VMC and DMC methods will be to problems in which the quantity of physical interest is the difference in energy between two large systems. Examples of such problems are calculations of excitation energies and defect energies in solids. In such cases the energy of interest is approximately independent of the size of the simulation cell, so that for each simulation cell size it is the energy of the *whole simulation cell* which must be converged to the required tolerance, not the energy per atom as we plotted in figures 6.3 and 6.4. In these cases extrapolation will be so costly that it can hardly be contemplated. In some cases the CFSE will largely cancel between the two systems, as occurs in excitation energy calculations (see next section). This cancellation cannot always be relied upon, particularly when the simulation cells contain different numbers of particles or different supercell geometries. The use of the MPC interaction should be particularly beneficial in such cases.

6.9 Excitation energies

The quasiparticle excitation energies are the energies for adding an electron to the system or subtracting one from it. A quasiparticle energy has both real and imaginary parts, the latter giving the quasiparticle lifetime. For the minimum energy electron and hole quasiparticle excitations the imaginary parts of the quasiparticle energies are zero and the quasiparticles have an infinite lifetime. In this case the exact quasiparticle energy gap can be written as

$$E_{\mathrm{g}} = (E_{N+1} - E_N) + (E_{N-1} - E_N) \ , \tag{6.16}$$

where E_{N+1}, E_{N-1}, and E_N are the ground state total energies of the $N+1$, $N-1$ and N electron systems. [3] A general quasiparticle energy gap cannot be written in terms of differences between energies of exact eigenstates of the system, but such an approximation is often accurate for low energy gaps. The quasiparticle energies are measured in photoemission and inverse photoemission experiments. In an optical absorption experiment a different process occurs in which an electron is excited from the valence to the conduction band. This introduces two quasiparticles into the system, the electron and hole, which interact and can form an exciton, in which case the lowest excitation energy is smaller than E_{g} by the exciton binding energy, E_{b}.

The HF method gives approximations to the energies of quasiparticles and the interactions between them. According to Koopmans' theorem, if orbital relaxation is neglected, the quasiparticle energies are equal to the HF eigenvalues. The *extended* Koopmans' theorem [98, 99] has been used in conjunction with VMC methods to calculate quasiparticle energies in silicon (see chapter 7 and reference [100]). In both of these methods the "quasiparticle energies" are real as they are obtained as approximations to the energy differences between exact eigenstates of the system.

Recently there has been significant progress in applying QMC techniques to calculate approximate excitation energies from eigenstates using "direct methods". In these approaches an excitation energy is obtained by performing separate QMC calculations for the ground and excited states. A Slater–Jastrow wave function is used for the ground state, and for the optical gap the excited state is formed by replacing a valence band single particle orbital by a conduction band one. This calculation is referred to as a "promotion" calculation, and such calculations have been reported for a nitrogen solid [42], diamond [101, 102], and silicon [103]. Photoemission/inverse photoemission gaps may be obtained by using QMC to calculate the ground state energies of the $N+1$, $N-1$, and N electron systems. Wave functions for the $N+1$ and $N-1$ electron systems may be formed by adding or subtracting an orbital from the up– or down–spin determinants of the N–electron wave function - an "addition/subtraction" calculation. For calculations with periodic boundary conditions the simulation cell is made charge neutral by adding a compensating uniform background charge density. Calculations of this type have been reported for one– [104] and two–dimensional [92] model systems, while the results results for a three–dimensional system (silicon) are reported in this thesis.

QMC calculations of excitation energies in extended systems are computationally very demanding because they are '$\frac{1}{N}$' effects, i.e., the fractional change in energy is inversely proportional to the number of electrons in the system. This means that high statistical accuracy is required to obtain good results. The largest system for which excitation energies have been calculated prior to this work is 16 atoms (64 electrons). [103] The total finite size error in the ground state energy for that system was estimated to be about 16 eV per simulation cell, while the energy scale of interest for the excitations is of order 0.1 eV. Like almost all methods for calculating excitation energies, QMC calculations of this type only work because of a strong cancellation of errors between the ground and excited states. Finite size errors tend to reduce the energy gap, while the errors in the trial wave functions are usually larger for the excited states than for the ground state and so increase the energy gap. Although good agreement with experimental excitation energies has been found using small simulation cells, [101–103] one is left with the suspicion that if larger simulation cells were used the agreement with experiment might be significantly worse because the finite size effects would be smaller.

In the next sections, the key questions as to the the size and origin of finite size effects in excitation energy calculations and the differences in finite size effects between promotion and addition/subtraction calculations are addressed. Before these QMC techniques can be relied upon for the calculation of excitation energies it is necessary that finite size effects be properly explored, and answers to these questions found.

6.9.1 HF theory of excitation energies

First excitation energies for solids are considered within HF theory. The HF equations with the MPC interaction are obtained by minimizing the HF energy of equation 6.15 with respect to the single–particle orbitals, giving

$$\begin{aligned} & -\frac{1}{2}\nabla^2\phi_i + \int \rho(\mathbf{r}')v_{\mathrm{E}}(\mathbf{r}-\mathbf{r}')\,\mathrm{d}\mathbf{r}'\,\phi_i \\ - & \sum_{i,j}^{N}\delta_{s_i s_j}\int \phi_j^*(\mathbf{r}')\phi_j(\mathbf{r})\phi_i(\mathbf{r}')f(\mathbf{r}-\mathbf{r}')\,\mathrm{d}\mathbf{r}' + V_{\mathrm{ext}}\phi_i = \epsilon_i\phi_i \quad . \end{aligned} \tag{6.17}$$

If orbital relaxation is neglected, the energy required to excite an electron from the j^{th} (occupied) orbital into the i^{th} (unoccupied) orbital is

$$\begin{aligned} \Delta E_{ij} &= (\epsilon_i - \epsilon_j) \\ &- \int \rho_i(\mathbf{r})\rho_j(\mathbf{r}')v_{\mathrm{E}}(\mathbf{r}-\mathbf{r}')\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \\ &+ \delta_{s_i s_j}\int \phi_i^*(\mathbf{r})\phi_j^*(\mathbf{r}')\phi_j(\mathbf{r})\phi_i(\mathbf{r}')f(\mathbf{r}-\mathbf{r}')\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \\ &+ \frac{1}{2}\int \rho_i(\mathbf{r})\rho_i(\mathbf{r}')\left[v_{\mathrm{E}}(\mathbf{r}-\mathbf{r}') - f(\mathbf{r}-\mathbf{r}')\right]\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \\ &+ \frac{1}{2}\int \rho_j(\mathbf{r})\rho_j(\mathbf{r}')\left[v_{\mathrm{E}}(\mathbf{r}-\mathbf{r}') - f(\mathbf{r}-\mathbf{r}')\right]\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}'\ , \end{aligned} \tag{6.18}$$

where $\rho_k = |\phi_k|^2$ is the charge density from the k^{th} orbital. The first term is the eigenvalue difference for the excitation while the second and third terms are the Hartree and exchange interactions between the electron and hole. Within this approximation the electron–hole terms go to zero in the limit of an infinitely large simulation cell. The fourth and fifth terms on the right hand side are absent if one uses the Ewald interaction instead of the MPC interaction, i.e., f is replaced by v_{E}. When the relaxation of the orbitals is neglected these terms also go to zero when the size of the simulation cell goes to infinity because v_{E} tends to $1/r$ over most of the simulation cell.

The addition/subtraction gap is given by

$$\begin{aligned} E_{\mathrm{g}} &= \left(E_{+i}^{\mathrm{HF}} - E_0^{\mathrm{HF}}\right) - \left(E_0^{\mathrm{HF}} - E_{-j}^{\mathrm{HF}}\right) \\ &= (\epsilon_i - \epsilon_j) \end{aligned}$$

$$
\begin{aligned}
&+ \quad \frac{1}{2}\int \rho_i(\mathbf{r})\rho_i(\mathbf{r}')\,[v_{\mathrm{E}}(\mathbf{r}-\mathbf{r}') - f(\mathbf{r}-\mathbf{r}')]\ \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \\
&+ \quad \frac{1}{2}\int \rho_j(\mathbf{r})\rho_j(\mathbf{r}')\,[v_{\mathrm{E}}(\mathbf{r}-\mathbf{r}') - f(\mathbf{r}-\mathbf{r}')]\ \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \quad , \qquad (6.19)
\end{aligned}
$$

where E_0^{HF} is the HF ground state energy of the N–electron system, E_{+i}^{HF} is the energy of the state with an electron added to the i^{th} (previously unoccupied) orbital, along with the uniform background charge, and E_{-j}^{HF} is the energy of the state where an electron is removed from the j^{th} orbital, along with the background charge. The standard Koopmans' theorem has been modified and contains two additional terms, which also occur in the promotion energy, ΔE_{ij}. These additional terms, when evaluated using LDA orbitals, are small even for a small simulation cell ($n = 2$), being in the range ± 0.05 eV, and they decrease rapidly with system size. The use of exact HF orbitals or orbital relaxation is not expected to greatly affect these results.

In figure 6.5 the addition/subtraction energies calculated using equation 6.19 for the $\Gamma_{25'} \rightarrow \Gamma_{15}$ energy gap and the valence band width, calculated with both the Ewald and MPC interactions, are shown along with the LDA values. The results for other low–lying excitation energies show similar behaviour. The promotion energies are not shown figure 6.5 because they differ from the addition/subtraction energies only by the exciton binding energy, which decreases with increasing system size quite rapidly. Figure 6.5 shows that the HF results for the Ewald and MPC interactions are very similar. The band width converges by about $n = 7$, but the band gap is still slowly increasing at $n = 12$, and the Ewald and MPC values are not yet equal, which they must be at convergence. For the largest system size studied ($n = 12$) the MPC gap and valence band width are 7.4 eV and 17.7 eV respectively, which are a little smaller than the HF values of 8.0 eV and 18.9 eV cited in Ref. [105]. The major reasons for these differences are the use of LDA wave functions and LDA–derived pseudopotentials, although as noted above there is clear evidence that in our calculations the HF energy gap has still not fully converged at $n = 12$. The LDA excitation energies converge very rapidly with system size.

In a recent study of silicon clusters, Öğüt *et al.* [106] found large differences between the band gap in the LDA eigenvalues and the band gap calculated by electron addition/subtraction. As shown by Franceschetti *et al.* [107], these differences are due to the charging of the cluster when an electron is added or subtracted, which does not occur in our calculations because a uniform background is added to preserve charge neutrality. The slow convergence of the HF excitation energies apparent in figure 6.5 therefore arises from the exchange energy. Moreover, because the results with the Ewald and MPC interactions are almost the same, the source of the error is not the interaction with the exchange hole, but the shape of the exchange hole. This is because the excitation energy depends on the change in the exchange hole due to the excitation, which is not strongly localised.

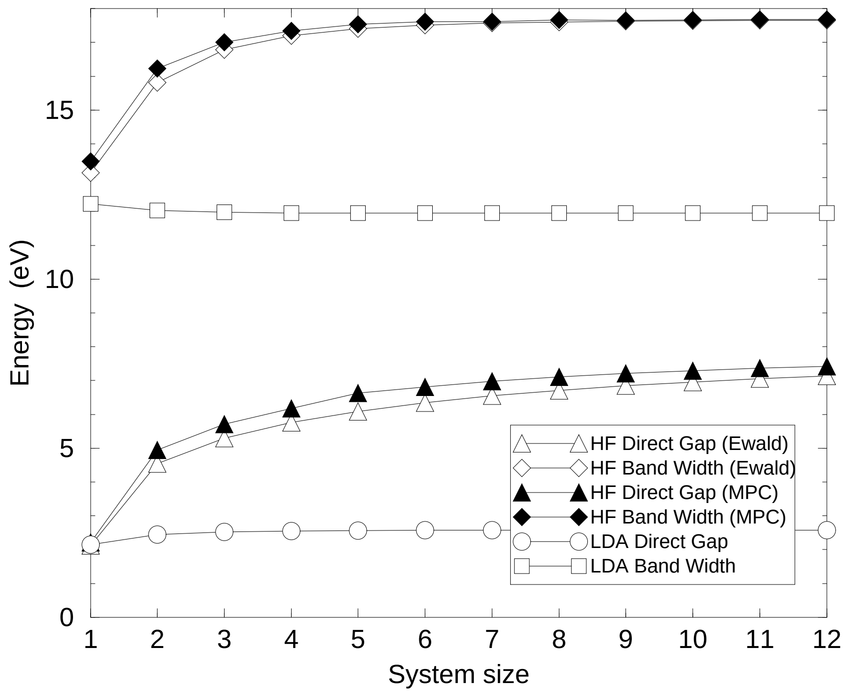

Figure 6.5: Convergence of the $\Gamma_{25'} \rightarrow \Gamma_{15}$ excitation energy and valence band width of silicon. The data correspond to addition/subtraction energies calculated within HF theory as a function of simulation cell size, n, using both the MPC and Ewald interactions. LDA results are also shown.

In a set of calculations on a model two–dimensional system using LDA, GW, VMC and DMC techniques, Engel *et al.* [92] performed a number of VMC calculations with increasing system size and found that the addition/subtraction gap tended to increase with system size. This is the same behaviour as found in the present HF calculations. Engel *et al.* went on to give an explanation of this effect. Their explanation was that in the $N + 1$ $(N - 1)$ electron systems there is an additional electrostatic energy due to the interaction of the extra electron (hole) with the additional electron (hole) in the other simulation cells. Taking into account the additional compensating uniform background charge that was added to keep each cell neutral, this additional energy is negative and therefore the energies E_{N+1} and E_{N-1} are lower than they should be. Engel *et al.* showed that the observed finite size effect is much smaller than the Madelung energy for point charges, and to explain this they argued that the effect would be screened by the response of the other electrons. This argument implies that the finite size effects in addition/subtraction calculations are larger than those in promotion calculations.

A reanalysis of the situation follows. HF calculations are considered for simplicity, where the interaction energy can be divided into Hartree and exchange contributions. The significant underestimation of the HF band gaps of small systems is not due to the Hartree terms, which by construction are the same as for our LDA calculations and give very small finite size effects in the band gaps. The finite size error in the HF gaps therefore arises from the exchange energy. By comparing band gaps calculated with the Ewald and MPC interactions we can see whether the problem lies with the interaction or with the shape of the exchange hole. Because the band gaps calculated with the Ewald and MPC interactions are very similar, the form of the interaction is not the important consideration. Therefore the source of the problem must be the finite size errors present in the shape of the exchange hole. This argument implies that the finite size effects in addition/subtraction calculations are similar to those in promotion calculations. This view is supported by the HF results that have been presented in this section and also by the VMC results of the next section.

In summary, the HF excitation energies calculated with the Ewald and MPC interactions are very similar. Within HF theory the largest finite size error in excitation energies arises from the shape of the exchange hole, which leads to slow convergence with system size. The finite size errors in promotion and addition/subtraction HF calculations are of similar size.

6.9.2 QMC theory of excitation energies

We now apply the theory developed in the previous section to QMC calculations of excitation energies. Although the HF gaps converge rather slowly with system size, it was shown in section 6.8.2 and section 6.8.3 that the finite size effects in VMC and DMC ground state energies are smaller than in HF theory. It is important to see whether this also applies to excitation energies.

VMC is computationally cheaper than DMC and so excitation energies may be computed using VMC over a larger range of system sizes. It is expected that the finite size effects in DMC will follow those in VMC, as the VMC calculations retrieve about 90% of the fixed–node correlation energy. The $\Gamma_{25'} \rightarrow \Gamma_{15}$ excitation energies were computed within VMC in silicon for the system sizes n = 1, 2, 3, 4 using both promotion and addition/subtraction methods. The calculations were performed with $\mathbf{k}_s \equiv \Gamma$ and the other computational details were the same as for the ground state calculations. The Jastrow factors were left unchanged for the excited state. In tests on the $n = 2$ system, separately optimizing the Jastrow factors for both ground and excited states did not significantly change the results. The computational cost of the $n = 4$ calculations is very large; an error bar in the excitation energies of ± 0.3 eV requires an error bar of ± 0.0006 eV per electron. Although the computational effort is large, such a study is necessary to establish the accuracy of QMC excitation energy calculations.

In figure 6.6 excitation energies obtained with the Ewald interaction via the promotion and addition/subtraction methods are shown. Results for the MPC interaction were found to be very similar. The promotion and addition/subtraction results are nearly the same, but the promotion energies are slightly smaller because they include an exciton binding energy, which decreases as the system size increases. The results are consistent with a slow increase in the excitation energy with system size and indicate that reasonable convergence is already obtained at $n = 2$. The increase in excitation energies with system size is the same trend as in HF calculations, although the finite size errors are smaller in the correlated calculations. The finite size errors in the promotion and addition/subtraction methods are not significantly different at this level of statistical accuracy. On general grounds one would expect the finite size effect in the promotion calculation to be slightly larger. This follows because the trend for both promotion and addition/subtraction calculations is that the excitation energy is reduced for small system sizes and this effect is enhanced in the promotion calculation by the exciton binding energy which is larger for small systems. The calculated excitation energy is roughly 4 eV, which is larger than the experimental value of 3.40 eV, [108] and also a little larger than the DMC value for an $n = 2$ simulation cell of 3.7 eV. [103] This study demonstrates that the largest contribution to the error in the VMC band gap for $n \geq 2$ arises from the approximate nature of the trial wave functions, and not from finite size effects.

The exciton binding energy can be calculated as the difference between the promotion and addition/subtraction gaps. The exciton binding energy for the $\Gamma_{25'} \rightarrow \Gamma_{15}$ excitation is small and it is only statistically resolvable for the smallest ($n = 1$) cell, which gives a value of 0.28 ± 0.01 eV. In earlier QMC calculations [42, 101, 102] the exciton binding energy was estimated using the Mott–Wannier formula

$$E_{\mathrm{b}} \sim \frac{1}{2\epsilon r} \quad , \tag{6.20}$$

where ϵ is the relative permittivity and r is the radius of the localisation region. Using $\epsilon = 11.7$ and the appropriate radius for $n = 1$ of $r = 4.0$ a.u. gives an exciton binding energy of 0.29 eV. This is extremely close to the VMC value, but the excellent agreement is probably fortuitous since the $n = 1$ cell is so small that it is appropriate to use a value of ϵ at finite wave vector, which would be smaller. The exciton binding energy may also be evaluated within HF theory as the sum of the second and third terms on the r.h.s. of equation 6.18. This gives 0.75 eV for the $n = 1$ cell using the Ewald interaction, which is considerably larger than the VMC value because the latter calculation includes screening of the electron–electron interaction.

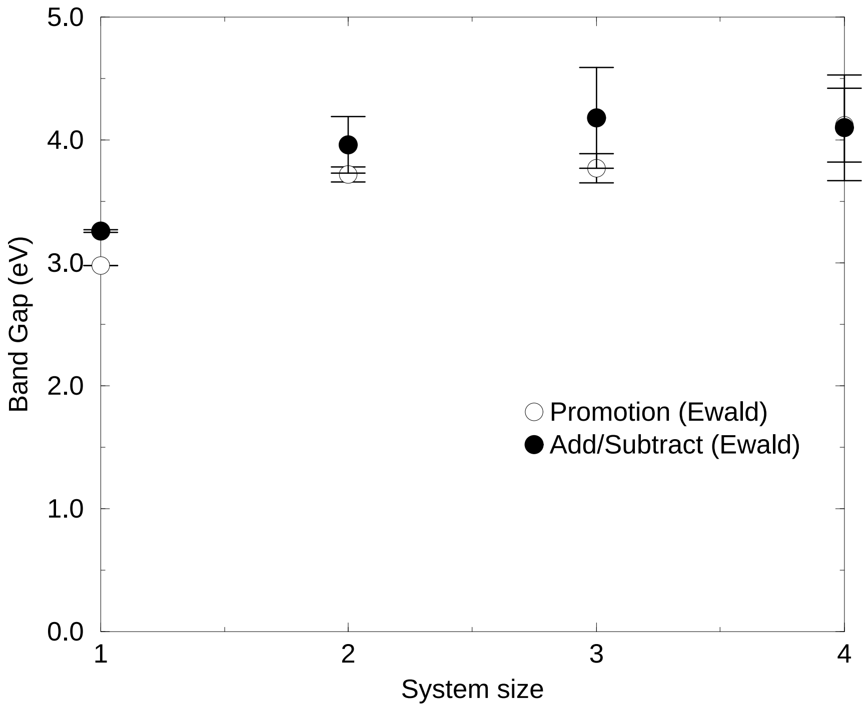

Figure 6.6: The $\Gamma_{25'} \rightarrow \Gamma_{15}$ excitation energy of silicon calculated within VMC theory as a function of simulation cell size, n. Data for the Ewald interaction are shown obtained via both the promotion and addition/subtraction methods.

Note that the promotion and addition/subtraction methods differ significantly in the required computational effort. If the intrinsic variance of the local energy is the same for each of the energies, which is a good approximation for silicon wavefunctions, then the most efficient way to achieve the same error bar in an addition/subtraction gap is to perform $2M$ moves for each of the $N+1$ and $N-1$ systems and $4M$ for the ground state, giving a total cost of $8M$ moves, where an acceptable error bar is obtained in a promotion calculation by performing M Monte Carlo moves for both the ground and excited states. It is therefore four times more expensive to calculate an energy gap to some given accuracy by the addition/subtraction method than by the promotion method.

6.9.3 Modified interaction for excitation energies

In this section a modified electron–electron interaction specifically designed to describe excitation energies within periodic boundary conditions simulations in introduced. Two problems arise when trying to model excitations using finite simulation cells subject to periodic boundary conditions. One is that the excitation is "squeezed" into the simulation cell, and the other is that there are spurious interaction between the periodic replicas of the simulation cell. Here the problem of the spurious interactions is addressed using the ideas of the MPC interaction. The charge density on promotion or addition/subtraction of an electron can be written as the sum of the ground state charge density and a small deviation, i.e., $\tilde{\rho}(\mathbf{r}) = \rho(\mathbf{r}) + \Delta(\mathbf{r})$. The Hartree term can be modified so that this charge density interacts with the *ground state* density in the replicas. This leads to the interaction energy

$$\begin{aligned}
\tilde{E}_{\mathrm{e-e}} &= \int |\tilde{\phi}|^2 \sum_{i>j} f(\mathbf{r}_i - \mathbf{r}_j) \prod_k \mathrm{d}\mathbf{r}_k \\
&+ \int \tilde{\rho}(\mathbf{r})\rho(\mathbf{r}') \left[v_{\mathrm{E}}(\mathbf{r}-\mathbf{r}') - f(\mathbf{r}-\mathbf{r}')\right] \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \\
&- \frac{1}{2}\int \rho(\mathbf{r})\rho(\mathbf{r}') \left[v_{\mathrm{E}}(\mathbf{r}-\mathbf{r}') - f(\mathbf{r}-\mathbf{r}')\right] \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}' \;\;, \qquad (6.21)
\end{aligned}$$

where $|\tilde{\phi}|^2$ generates the charge density $\tilde{\rho}$ and the ground state charge density, ρ, is fixed. A HF analysis of this interaction shows that the HF equations are identical to equation 6.17, so that the orbitals and eigenvalues are unaltered. However, the ground and excited state energy expressions are modified. For the excited states analogues of equations 6.18 and 6.19 are obtained, but without the terms involving $(v_{\mathrm{E}} - f)$, i.e., the standard Koopmans' theorem is retrieved. Although these terms are small for silicon, they will be significant in cases when the change in the charge density of the exciton is strongly localized. This analysis also provides further evidence that the electrostatic interactions between the simulation cell and its replicas is not necessarily an important source of finite size error in excited state energy calculations; for these silicon systems, the finite size errors in the excitations are small.

6.10 Conclusions

Large Coulomb finite size errors arise in many–body calculations when using the Ewald form of the electron–electron interaction. These errors are different from those encountered in single–particle theories such as DFT, and corrections derived only from single–particle theories are unsatisfactory.

By modifying the interaction terms in the Hamiltonian, the finite size errors may be considerably reduced. A modified interaction, the "Modified Periodic Coulomb" (MPC) interaction, in which the inter–particle interaction is exactly equal to $1/r$ at short distances and the long range interactions are replaced by a mean–field–like one–electron potential, is found in actual HF, VMC and DMC calculations to substantially reduce the finite–size errors in total energies.

The Ewald and MPC interactions may be used in tandem as an efficient diagnostic of Coulomb finite size errors. If the Ewald and MPC results agree then the Coulomb finite size error should be small. If the simulation cell is too small then the confinement of the XC hole makes the XC energy more negative. This source of error is intrinsic to using a finite simulation cell. However, even when the XC hole is artificially confined by a small simulation cell the MPC interaction still gives a better estimate of the XC energy than the Ewald interaction.

Excitation energies calculated within fixed–LDA–orbital HF theory show significant finite size effects. However, in correlated calculations the finite size effects are smaller and accurate excitation energies can be obtained using quite small simulation cells. In silicon the finite size errors in VMC electron–promotion ("optical absorption") and electron addition/subtraction ("photoemission") calculations are similar, and the optical promotion method has the greater statistical efficiency. The finite size errors for low lying excitations in silicon are small, and quite accurate results may be obtained from 16 atom cells. It appears that the limiting factor in obtaining accurate results for excitation energies is the accuracy of the excited state wavefunctions and not the size of supercell.

In every test the energy calculated with the MPC interaction is closer than the Ewald energy to the value for a very large system. Use of the MPC interaction therefore permits the use of smaller simulation cells for a specified accuracy than use of the Ewald interaction. The MPC is therefore an attractive and practical alternative to the Ewald interaction, suitable for use in all quantum many–body calculations of total energies in systems with periodic boundary conditions.

Chapter 7

The one–body density matrix and excitation energies of silicon

7.1 Introduction

The one–body density matrix is an object fundamental to quantum mechanics. Its complete knowledge enables the computation of all one–body expectation values without reference to a system's wavefunction. The one–body density matrix [109–111] for a normalized wave function, ψ, is defined as

$$\begin{aligned}\rho(\mathbf{r},\mathbf{r}') &= N\int \psi^*(\mathbf{r},\mathbf{r}_2,\ldots,\mathbf{r}_N) \\ &\times \psi(\mathbf{r}',\mathbf{r}_2,\ldots,\mathbf{r}_N)d\mathbf{r}_2\ldots d\mathbf{r}_N\ .\end{aligned} \tag{7.1}$$

The pair–correlation function of the valence electrons of bulk silicon, which is related to components of the *two–body* density matrix, was recently computed using VMC. [14] Substantial inhomogeneity and deviations from single particle theories were found, suggesting that other distribution functions, such as the one–body density matrix should be examined. No previous studies of the entire one–body density matrix using accurate many–body wavefunctions were found in the literature.

Although the one–body density matrix is of immediate interest, the matrix also provides a potential route to wavefunction optimisation and to the computation of excitation energies. In this chapter, a QMC scheme for sampling the one–body density matrix is developed and applied to the valence electrons of bulk silicon. The natural orbitals are obtained by diagonalising the density matrix and tested for improvements in the DMC fixed–node energy. The relation between the density matrix, the momentum density and

Compton profiles is briefly discussed, and the zero momentum component of the momentum density computed. Approximate excitation energies are also obtained, via an extension of Koopmans' theorem, [112] again with the density matrix as a key quantity.

7.1.1 Orbital optimisation

It is highly desirable to improve the quality of the trial wave functions used in QMC calculations. Improvements to trial wave functions can be classified into three types: (i) improvement of the Jastrow factor, (ii) using a linear combination of determinants, and (iii) improvements in the orbitals forming the determinants. The use of natural orbitals, which diagonalise the one–body density matrix are a possible improvement of type (iii).

The question of which single-particle orbitals lead to the best approximation to the exact many-body wave function is still open. Furthermore, this choice fixes the nodal surface of the trial wave function and therefore determines the accuracy of the fixed-node approximation. LDA and HF orbitals have been used successfully in a number of atomic, [37, 46] molecular [113, 114] and solid [115] QMC calculations, but so far it has not proved possible to perform a direct optimization of the single-particle orbitals of an extended system. A study of first row atoms and molecules [53, 116] showed that lower energies can be obtained in both VMC and DMC using a trial wave function containing several determinants obtained from a multi-configuration self-consistent field (MCSCF) calculation. However, a similar study for small silicon clusters [45] found that trial wave functions containing a single determinant of natural orbitals computed within an MCSCF scheme gave better DMC results than some multi-determinant wave functions. This result strongly suggests that the natural orbitals result in improved nodal surfaces, and motivates the calculation of the natural orbitals for bulk silicon.

An expansion of a wave function in Slater determinants of natural orbitals requires a smaller number of terms for a given accuracy than expansions using other orbitals. [2, 4] Calculation of the natural orbitals is, however, costly, and less expensive schemes such as natural pair orbitals [117, 118] have been proposed to improve convergence in quantum chemical calculations. It is not clear that orbitals arising in schemes designed to accelerate convergence of configuration interaction (CI) calculations should give smaller fixed-node errors in DMC calculations than LDA or HF orbitals. Although a CI wavefunction can well approximate an exact eigenstate, whether the low order terms in the expansion improve the nodes in general is not known. However, as mentioned above, there is some evidence to suggest that natural orbitals have this property. Natural orbitals have not frequently been computed within fermion QMC, although VMC and DMC calculations of natural orbitals have been reported for the ground states of the Li, C, and Ne atoms. [67] No calculations of natural orbitals for realistic extended fermion systems using many-body wavefunctions have appeared in the literature prior to this study, although for homogeneous systems the translational symmetry requires the natural orbitals to be plane waves.

Systematic studies of multi-determinant wave functions in QMC are lacking for solids. It seems reasonable to assume that multi-determinant wave functions will have improved nodes, and therefore give a better representation of the exact wave function, but there is little direct evidence to support this. Multi-configurational approaches include correlation effects, but do so relatively inefficiently - large numbers of terms (configurations) are usually required to obtain a significant proportion of the correlation energy. This form of wave function is unattractive for QMC as a representation of the wave function which is both accurate and can be evaluated rapidly is required. By obtaining the one-body density matrix and hence the natural orbitals from a VMC calculation using a correlated trial wave function, the need to determine them using a multi-determinantal calculation is bypassed.

7.1.2 Excited states

QMC techniques may be applied to excited states, and consequently there is considerable interest in calculating excitation energies and optical properties of materials. Excitation energies may be obtained by analyzing DMC decay curves, [119, 120] but this method has not proven very useful due to the large statistical noise. Furthermore, as the quality of the ground state trial wave function improves, less information about excited states is obtained. A combination of ground and excited state wave functions must then be used to obtain upper bounds for the excitation energies.

Direct methods for calculating excitation energies have met with more success. Mitáš and Martin have calculated an excitation energy in a molecular nitrogen solid by performing DMC calculations for the ground and excited states. [42] Mitáš has also reported similar calculations for two excitation energies in diamond. [101] Recently [103] 27 excitation energies of silicon were calculated using DMC, obtaining very good agreement with experiment for the low lying excitation energies, while the energies of the higher lying excitations were somewhat too large. Although successful, these techniques are computationally very expensive. The energy of each state of interest is calculated separately and excitation energies computed as energy differences.

A different approach to the direct methods is taken here, using the "Extended Koopmans' Theorem" (EKT). [98,99] The theorem derives from quantum chemistry and involves the one-body density matrix. The EKT has not previously been applied to extended systems, or applied to explicitly many–body wavefunctions.

The EKT may be approximated by a simplified diagonal approximation to the full theory. This approximation was used previously to estimate quasihole energies in silicon [121] and NiO. [122] The EKT and its simplified form require only a *single* ground state calculation in contrast to direct techniques, allowing more efficient studies of excitations in large systems. If sufficiently accurate the EKT would clearly be a significant advance over direct techniques.

7.2 Silicon

In this section, the silicon test system, the many-body wavefunction and its parameterisation are described. Bulk silicon in the diamond structure was chosen as a test system for several reasons. The material is commonly used as a test system for electronic structure methods and consequently a substantial literature of comparative data exists. Crucially it was considered that calculations of sufficient statistical accuracy to resolve the density matrix would be computationally affordable, partly because the same system had previously been used in studies of the pair–correlation function [14] using QMC techniques.

7.2.1 The supercell

For this study an fcc supercell containing 54 Si^{4+} ions at the experimental lattice constant and 216 electrons was selected. The electron-ion potential, v_α, was modeled by a norm-conserving non-local pseudopotential [57] obtained from atomic calculations performed within the local density approximation (LDA) to density functional theory.

Energies were computed in VMC using both the Ewald and MPC interactions, using a wave function which was optimized using the MPC interaction (see chapter 6). By using both interactions, some of the finite size effects could be isolated and quantified, providing a useful diagnostic as to the likely size of any finite size errors, without resorting to costly and error prone extrapolation procedures.

7.2.2 The trial wavefunction

A Slater-Jastrow wavefunction of the type described in chapter 4 was used, consisting of a single product of spin-up and spin-down determinants multiplied by a Jastrow factor. The Jastrow factor consisted of a one-body χ function and two-body correlation factor. A total of 16 parameters were optimized by minimizing the variance of the energy (see chapter 5), obtaining approximately 85% of the fixed–node correlation energy.

The spin-up and spin-down Slater determinants were formed from single-particle orbitals obtained from an LDA calculation employing the same pseudopotential as in the

QMC calculations. The LDA orbitals were calculated at the Γ-point of the simulation cell Brillouin zone using a plane wave basis set with an energy cutoff of 15 Ry. Although the Γ-point scheme does not give optimal Brillouin zone sampling, it allows comparison with a wider number of established results. The Γ-point of the simulation cell Brillouin zone unfolds to four inequivalent **k**-points in the primitive Brillouin zone. These are: (0,0,0) (the Γ-point), $(0,0,\frac{2}{3})\frac{2\pi}{a}$ (a point along the Δ axis, hereafter referred to as the Δ-point), $(0,\frac{2}{3},\frac{2}{3})\frac{2\pi}{a}$ (a point along the Σ axis, hereafter referred to as the Σ-point), and $(\frac{1}{3},\frac{1}{3},\frac{1}{3})\frac{2\pi}{a}$ (a point along the Λ axis, hereafter referred to as the Λ-point).

7.2.3 VMC and DMC calculations

The VMC and fixed–node DMC calculations were carried out following the methods and procedures described in chapter 4.

In the DMC calculations a time step of 0.015 a.u. was used, which has been shown to give a small time-step error in silicon. [97] This timestep (and wavefunction) gave an acceptance/rejection ratio greater than 99.9% in the DMC calculations.

7.3 The density matrix and natural orbitals

The one-body density matrix, defined in equation 7.1, is conventionally expanded in a basis of orbitals, ϕ_i, where

$$\rho(\mathbf{r},\mathbf{r}') = \sum_{i,j} \rho_{ij}\phi_i(\mathbf{r})\phi_j^*(\mathbf{r}') \,. \tag{7.2}$$

The diagonal elements, ρ_{ii}, are commonly referred to as the orbital occupation numbers. For wave functions consisting of a single determinant, such as HF or LDA wave functions, the density matrix is idempotent ($\rho = \rho^2$) and takes the form of a sum over the occupied orbitals, i.e.,

$$\rho(\mathbf{r},\mathbf{r}') = 2\sum_{i=1}^{N/2} \phi_i(\mathbf{r})\phi_i^*(\mathbf{r}') \,, \tag{7.3}$$

so that the occupation numbers are 2 (including spin degeneracy) for occupied orbitals and 0 for unoccupied orbitals.

The matrix elements of the interacting density matrix, ρ_{ij}, are written as expectation values over the distribution $|\psi|^2$:

$$\rho_{ij} = N \int \phi_i^*(\mathbf{r}_1)\phi_j(\mathbf{r}') \frac{\psi(\mathbf{r}', \mathbf{r}_2, \ldots, \mathbf{r}_N)}{\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)} \times |\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)|^2 \, d\mathbf{r}' d\mathbf{r}_1 \ldots d\mathbf{r}_N \, . \tag{7.4}$$

The permutation symmetry allows a rewriting of equation 7.4 in an efficient way for Monte Carlo evaluation. Denoting the average over the distribution, $|\psi|^2$, as $\langle \ldots \rangle_{|\psi|^2}$, the Monte Carlo expectation value is rewritten as

$$\rho_{ij} = \left\langle \sum_{n=1}^{N} \int \phi_i^*(\mathbf{r}_n)\phi_j(\mathbf{r}') \frac{\psi(\ldots, \mathbf{r}', \ldots)}{\psi(\ldots, \mathbf{r}_n, \ldots)} d\mathbf{r}' \right\rangle_{|\psi|^2} , \tag{7.5}$$

so that N values are accumulated at each step along the VMC walk.

The integral over $d\mathbf{r}'$ is performed by summing over a grid of uniform spacing whose origin is chosen randomly for each electron configuration. The same grid in $\mathbf{r}'$ is used for each term in equation 7.5, which further reduces the computational cost. A series of grid sizes were tested for the $\mathbf{r}'$ integral, using identical configurations for each grid size to obtain a correlated sampling estimate [23] of the difference between the integrals. A grid containing 125 points in the simulation cell was tested and found to sample the integral with sufficient numerical accuracy.

Provided that the density of points in the $\mathbf{r}'$ integral is kept constant, the statistical error in the individual elements of the density matrix for a given number of statistically independent configurations is approximately independent of system size.

A basis set consisting of the lowest energy LDA orbitals at the 27 $\mathbf{k}$-points in the primitive Brillouin zone was used. LDA orbitals were expected to form an efficient basis for the density matrix as the quasiparticle orbitals of silicon are very close to the LDA orbitals. The effect of varying the number of orbitals in the basis was tested, and approximately 40 orbitals per $\mathbf{k}$-point were sufficient, although to retain the symmetry all rows of a degenerate representation (a degeneracy) were included, so that the actual number used was either 39 or 40 orbitals, depending on the $\mathbf{k}$-point.

The normalization used in equation 7.1 requires that

$$N = \mathrm{Tr}\rho \, , \tag{7.6}$$

which provides a practical test for the completeness of the basis set. The total occupation of the matrix ($\mathrm{Tr}\rho$) was 215.9(2), which is within the statistical error of the number of electrons in the system, indicating sufficient completeness of the basis at the level for the statistical accuracy obtained.

7.3.1 Symmetry and efficiency considerations

Although for a given number of Monte Carlo moves, the best statistics were found to be obtained by accumulating all non-zero matrix elements and applying the symmetry afterwards, it was found to be computationally very expensive to accumulate all of them. A computationally more efficient procedure was found to involve accumulating only the independent non-zero matrix elements, and applying symmetry relationships to obtain the remaining elements. The basis set of LDA orbitals are basis functions of the unitary irreducible representations of the symmetry group O_h^7. Using the "orthogonality condition for matrix representations" [123] it was inferred that elements involving products of orbitals from inequivalent **k**-points and of differing representations are zero. An explicitly fully symmetrised basis set was constructed numerically from a partially symmetrised basis set using the projection operator technique [123]. This procedure ensured that every occurrence of a given representation was identical, so that products between functions belonging to different rows were orthogonal. This procedure reduced the total number of independent and non-zero matrix elements, ρ_{ij}, from 42094 to 582 elements. This reduction resulted in substantial gains in computational efficiency, enabling substantially more accurate statistics to be obtained using the available computer time.

The 582 matrix elements were sampled using approximately 6.6×10^5 statistically independent configurations. The correlation lengths along the VMC walks of both the local energy and density matrix were found to be essentially the same.

7.3.2 Density matrix results

Elements of the accumulated density matrix, ρ_{ii}, are shown in figure 7.1. The accumulated density matrix, ρ_{ij}, was very nearly diagonal with little coupling between LDA orbitals. The occupation numbers decrease almost linearly with increasing LDA energy for both the occupied and unoccupied bands.

Denoting double occupancy (spin-up and down) of orbitals by the value 2.0, the maximum difference between an interacting occupation number, ρ_{ii}, and the LDA occupation number was 0.0625(5). This difference occurred at the $\Gamma_{25'}$ state at the top of the valence band. The magnitude of the largest off-diagonal matrix element was 0.014(1), which is of similar order to the occupation number of the lowest unoccupied orbitals. The fractional errors in occupation numbers for orbitals of low occupation were large in comparison to those of high occupation. Approximately 97.6% of the total occupation of the density matrix is contained within the four occupied LDA bands at each **k**-point, and 99.0% is obtained within the first ten bands.

In figure 7.2 the density matrix, $\rho(\mathbf{r}, \mathbf{r}')$, in the (110) plane. The coordinate $\mathbf{r}$ is fixed

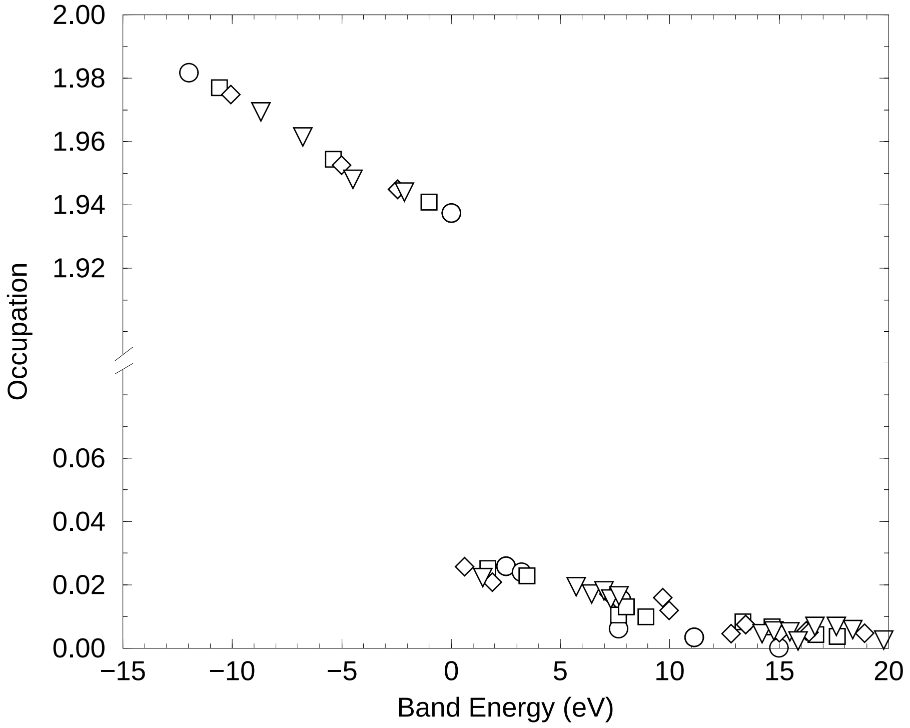

Figure 7.1: Occupation numbers (diagonal elements of the density matrix in the basis of LDA orbitals) plotted against the LDA energy for each **k**-point. These are the Γ-point($\circ$), Δ ($\diamond$), Σ ($\bigtriangledown$), and Λ ($\square$), points on the Δ, Σ, and Λ axes respectively. The statistical error bars are approximately equal to the sizes of the symbols for the conduction band states and are about 5 times smaller than the symbols for the valence band states.

at the center of a covalent bond, and $\mathbf{r}'$ ranges over the (110) plane passing through the atomic positions. The small differences between the VMC and LDA matrices are shown separately. The density matrix consists of an asymmetric central peak, reduced in width along the bonding direction. A longer ranged structure is present in areas of high valence charge density, smaller by approximately one order of magnitude than the peak.

The VMC value for the peak in the density matrix on the bond center at $\mathbf{r}=\mathbf{r}'$ is 1.7% smaller than the LDA value. The VMC density matrix has a larger magnitude around the neighboring silicon ions than the LDA density matrix, which consequently has a slightly smaller range. The density matrix was examined in the interstitial regions, where more structure was found. Again, the LDA and VMC results were very similar, with small differences between the two cases arising principally from the differing charge densities.

To investigate the possible finite size effects the LDA density matrices were computed for $3\times3\times3$ and $4\times4\times4$ **k**-point meshes, corresponding to simulation cells containing 54 and 128 atoms respectively. The central peak was largely unchanged and the longer ranged

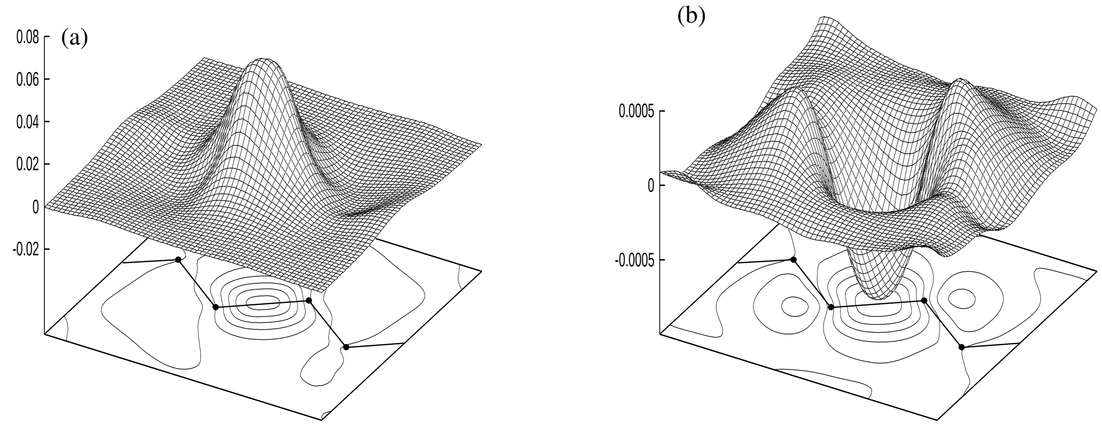

Figure 7.2: (a) The VMC one-body density matrix, $\rho_{\text{VMC}}(\mathbf{r},\mathbf{r}')$ and (b) $\rho_{\text{VMC}}(\mathbf{r},\mathbf{r}') - \rho_{\text{LDA}}(\mathbf{r},\mathbf{r}')$, in the (110) plane passing through the atoms with $\mathbf{r}$ fixed at the bond center. ρ is normalized such that $\rho(\mathbf{r},\mathbf{r}) = n(\mathbf{r})$, the charge density at the bond center. The silicon atoms and bonds are shown schematically.

structure was in qualitative agreement. The central maximum in the density matrix, in figure 7.2, is at the point $\mathbf{r}=\mathbf{r}'$ and its magnitude is directly proportional to the valence charge density at that point, which differed by 4.9% between the two simulation cell sizes. These differences are small and indicate that no significant deviations between the LDA and VMC results are likely to be resolved by using a larger cell.

In exact Kohn-Sham density functional theory the total energy can be written entirely in terms of the one-body density matrix of the Kohn-Sham orbitals, whereas in a fully interacting system both the one-body density matrix and the pair-correlation function are required. Results for the pair-correlation function from accurate correlated wave functions and LDA calculations are extremely different. [14] Exact Kohn-Sham density functional theory reproduces the exact charge density and therefore exactly reproduces the diagonal $\mathbf{r} = \mathbf{r}'$ part of the density matrix. The off-diagonal part of the exact Kohn-Sham and interacting density matrices are not required to be the same. In silicon the LDA is expected to give a good approximation to the exact Kohn-Sham density matrix, as the LDA performs very well. For this system the results show that the *entire* density matrix is very similar in VMC and LDA. It is reasonable to speculate that the exact Kohn-Sham and exact one–body density matrices are probably very similar in other covalent *sp* systems.

7.3.3 Natural orbital results

The natural orbitals were obtained by diagonalising the density matrix in the basis of the LDA orbitals. An assessment of the statistical errors in the eigenvalues and eigenvectors was made by subjecting the matrix to random perturbations of order the statistical error. The eigenvalues varied by up to ± 0.0004 on application of the small perturbations. All the calculated eigenvalues, λ_i, of the density matrix lie in the range $0 \leq \lambda_i \leq 2$, as is required. [109–111] Identical results were obtained when elements within statistical error of zero were explicitly zeroed. The overlap of the space occupied by the LDA orbitals and the natural orbitals is measured by the absolute value of the determinant of the matrix of overlaps between these two sets of vectors. This gave a value of 0.9948, indicating that the spaces spanned are almost the same.

The eigenvalues of the density matrix for the "occupied" natural orbitals were very slightly larger than the corresponding matrix elements ρ_{ii} (by about 0.001). Consequently, the eigenvalues of the "unoccupied" natural orbitals were very slightly decreased, so that $\mathrm{Tr}\rho$ is invariant. Therefore a plot of the eigenvalues of the density matrix would be indistinguishable from figure 7.1, which shows the diagonal elements of the density matrix in the basis of LDA orbitals.

7.3.4 Relation to Compton profiles

Compton scattering is a form of inelastic x-ray scattering characterised by large energy and momentum transfer, and can be used to probe the electronic structure of materials by measuring the electron momentum distribution and its anisotropies. [124] Improvements in synchrotron x-ray sources have led to significant developments of this technique and resolutions of down to $\sim$ 0.02 a.u. (at best) for momenta are currently possible. In this section, the zero momentum component of the momentum density for bulk silicon is determined from the density matrix and compared with previous results for the momentum density and Compton profile.

Accurate predictions of Compton profiles are desired to help confirm the more detailed information about electronic structure that simulations can provide; from accurate electronic structure calculations it should be possible to compute the experimentally observed Compton profile. At the time of writing, most theoretical calculations of Compton profiles have been computed using a single-particle framework such as LDA-DFT within the "impulse approximation". [125] Within this approximation, x-rays are considered to transfer

$N_{\rm VMC}(0) - N_{\rm LDA}(0)$	Value (electrons/primitive cell)/(a.u.)^3
This work	-0.058(2)
Králik *et al.*	-0.053(10)

Table 7.1: Calculated $\mathbf{k} = 0$ components of the momentum density of bulk silicon, given as the difference between the VMC and LDA-DFT values. For this work, a 54 atom supercell was used, compared with the 16 atom supercell of Králik *et al.* [126]. (The units are the same as figure 1 of [126].)

momentum and energy to *free* electrons. [125] The Compton profile, $J(\mathbf{q})$, is then given in terms of an integral of the momentum density, $N(\mathbf{k})$,

$$J(\mathbf{q}) = \int N(\mathbf{k}')\delta(\hat{\mathbf{q}}.\mathbf{k}' - q)d\mathbf{k}' \quad , \tag{7.7}$$

where the $\mathbf{k}'$ integral is over the Brillouin zone. The momentum density, for a many-body system, is given by an expression similar to the expression for the density matrix (equation 7.1),

$$\begin{aligned} N(\mathbf{k}) &= \int \psi^*(\mathbf{r}_1, \ldots, \mathbf{r}_N) \exp(i\mathbf{k}.\mathbf{s}) \\ &\times \psi(\mathbf{r}_1 + \mathbf{s}, \mathbf{r}_2, \ldots, \mathbf{r}_N) d\mathbf{s} d\mathbf{r}_1 \ldots d\mathbf{r}_N \ . \end{aligned} \tag{7.8}$$

Recently, Králik *et al.* [126] and Filippi and Ceperley [127] computed the Compton profiles of bulk silicon and lithium respectively, using QMC techniques. Large deviations between experiment and the QMC results were apparent at small $\mathbf{q}$, although the results did improve on those of LDA-DFT calculations and were in agreement the Lam-Platzman correction [128] to the LDA-DFT results. This correction is in principle exact within DFT, but must be approximated using an approximate functional such as the LDA. The QMC calculations for bulk silicon [126] used a small simulation cell of 16 atoms. Hence, it is possible that finite-size effects and not the impulse approximation could be responsible for the residual discrepancy between the QMC results and experiment.

The value of $N(0)$ is given by equation 7.8 in terms of the natural orbitals of the density matrix and their occupancies. $N(0)$ was evaluated using the VMC density matrix for the 54 atom system used in this chapter. The results are shown in table 7.1 in terms of the difference between the VMC and LDA-DFT values.

Although a larger supercell is used in this work, the value of $N(0)$ obtained is in good agreement with Králik *et al.* and has a smaller statistical error. This result strongly suggests that the finite size effects in the calculations are small, and that a similar Compton profile would be obtained if a full calculation was performed for the larger cell. In a later paper on the Compton profile of silicon, [129] Delaney *et al.* report that the effect of using pseudopotentials on the Compton profile is also small.

Filippi *et al.* [127] performed DMC calculations as well as VMC calculations for lithium. They found only very small differences in the momentum distribution between the two calculations indicating the VMC wavefunctions are of sufficient quality for this work. They also found deviations between experiment and the QMC results at small $\mathbf{q}$.

These results strongly support the conclusion that the residual difference between experiment and theory is not due to the treatment of electron-electron correlation in these materials. The residual differences in both silicon and lithium are therefore most likely due to a breakdown of the impulse approximation [125], although there may also be contributions from finite temperature effects that are not accounted in the experimental analysis.

7.4 Tests of orbitals

7.4.1 LDA and natural orbitals compared

Fixed node DMC calculations were performed using LDA and natural orbitals to form the determinants, in the expectation that a determinant of natural orbitals would have lower nodal error than a determinant of LDA orbitals. Re-optimization of the Jastrow and χ functions to improve sampling efficiency in the DMC calculation was found to be unnecessary in the two cases. VMC calculations were also performed to test for the possibility that the variational energy had been improved.

The energies obtained were -107.59 eV (LDA), $-107.69(1)$ eV (VMC with LDA orbitals), $-107.71(1)$ eV (VMC with natural orbitals), $-108.10(1)$ eV (DMC with LDA orbitals), and $-108.09(1)$ eV (DMC with natural orbitals). The VMC wave function appears to show a very slight improvement with natural orbitals compared with LDA orbitals. However, to within statistical accuracy, the DMC energies obtained with LDA and natural orbitals are the same. This indicates that the nodal surfaces given by the LDA and natural orbitals are of the same quality.

7.4.2 LDA and HF orbitals compared

In light of the above results it is interesting to compare the quality of the nodal surfaces obtained with LDA and HF orbitals, which are both commonly used in the determinantal parts of trial wave functions for QMC calculations.

This was investigated by performing DMC calculations in silicon with an fcc simulation cell containing 16 atoms. The smaller simulation cell enabled a large number of independent configurations to be obtained rapidly. Wave functions expanded in a basis of atom-centered Gaussians were obtained from the HF and DFT code CRYSTAL95. [130] Special care was taken to ensure that the LDA and HF calculations were done in equivalent ways to try and eliminate any bias in the comparison. A basis set of four uncontracted *sp* functions and one *d* polarization function per pseudo-atom was optimized separately for each calculation. The quality of the basis set is high - to obtain the same energy within a plane wave calculation would require a basis set cutoff of 12.5 Ry. The same non-local LDA pseudopotential was used as in the earlier calculations. The Ewald interaction was used in the many-body Hamiltonian to avoid any charge-density dependence on the interaction. Approximately 6.7×10^5 walker moves obtained DMC total energies and statistical accuracies of -107.488(3) eV per atom and -107.464(3) eV per atom for the LDA and HF guiding wave functions respectively.

The walker energies were approximately normally distributed. Using a conventional students t-test, [19] the 95% confidence interval on the difference in energies obtained was $0.002 - 0.046$ eV per atom, showing that for this system it is very likely that the DMC energy from a determinant of LDA orbitals is lower than that from a determinant of HF orbitals. Therefore, for this system, a determinant of LDA orbitals has a marginally better nodal structure than a determinant of HF orbitals.

7.5 The Extended Koopmans' theorem

The calculation of excited state energies in solids using QMC methods is a fairly new area of research. The method described here obtains a large number of excitation energies from a single QMC calculation involving averages over the ground state wave function.

The Extended Koopmans' Theorem (EKT) was derived apparently independently by Smith, Day and Garrod [98] and by Morrell, Parr and Levy. [99] The EKT is closely related to the earlier work of Feynman [131] on calculating excitation energies in the superfluid state of He^4, although the quantum chemists appear to have developed the theory independently. The EKT has been shown to give very good excitation energies for simple molecular systems, [132] and has been applied to atomic and diatomic systems. [133,134] It appears particularly well suited to QMC calculations in which explicitly correlated many-body wave functions are used. The derivation is reviewed following Ref. [98], and a variation suitable for QMC is formulated.

7.5.1 Valence band energies

In this method the band energies are calculated as ionisation energies. The starting point is an approximation to the normalized ground state wave function, ψ^N. The wave function for the N-1 electron system is *approximated* by the Ansatz of eliminating an orbital from ψ^N:

$$\psi^{N-1}(\mathbf{r}_2, \ldots, \mathbf{r}_N) = \int u_v^*(\mathbf{r}_1)\psi^N(\mathbf{r}_1, \ldots, \mathbf{r}_N)d\mathbf{r}_1 \, . \tag{7.9}$$

The valence orbital to be eliminated, u_v, will be determined variationally. This Ansatz is reminiscent of a quasiparticle wave function for the state, although the formulation is for N-1 particle eigenstates of the Hamiltonian. Expressing equation 7.9 in second quantization yields

$$|\psi^{N-1} >= \hat{\mathcal{O}}_v|\psi^N > , \tag{7.10}$$

where $\hat{\mathcal{O}}_v$ is the destruction operator for the state u_v.

The ionisation energy is given by the difference in the expectation values of the Hamiltonian calculated with the N and N-1 electron wave functions: [99, 133]

$$\epsilon_v = \langle\psi^N|\hat{H}|\psi^N\rangle - \frac{\langle\hat{\mathcal{O}}_v\psi^N|\hat{H}|\hat{\mathcal{O}}_v\psi^N\rangle}{\langle\hat{\mathcal{O}}_v\psi^N|\hat{\mathcal{O}}_v\psi^N\rangle} \, . \tag{7.11}$$

If ψ^N is an eigenfunction of $\hat{H}$, equation 7.11 may be written as

$$\epsilon_v = -\frac{\langle\psi^N|\hat{\mathcal{O}}_v^\dagger[\hat{H}, \hat{\mathcal{O}}_v]|\psi^N\rangle}{\langle\psi^N|\hat{\mathcal{O}}_v^\dagger\hat{\mathcal{O}}_v|\psi^N\rangle} \, . \tag{7.12}$$

The denominator in equation 7.12 is the one-body density matrix. Solution of this equation is facilitated by expanding in a set of orbitals, $\{\phi_i\}$, so that $u_v(\mathbf{r}) = \sum c_{iv}\phi_i(\mathbf{r})$, and $\hat{\mathcal{O}}_v = \sum c_{iv}\hat{a}_i$, where $\hat{a}_i$ is the destruction operator for ϕ_i. The condition for a stationary value of ϵ_v generates a secular equation

$$(\mathbf{V}^v - \epsilon_v\mathbf{S}^v)\mathbf{c}_v = \mathbf{0} \, . \tag{7.13}$$

The matrix $\mathbf{S}^v$ is the one-body density matrix, and the elements of $\mathbf{V}^v$ are

$$V_{ij}^v = \langle\psi^N|\hat{a}_j^\dagger[\hat{H}, \hat{a}_i]|\psi^N\rangle \ \ , \tag{7.14}$$

where

$$V_{ij}^v \ = \ N\int \phi_i(\mathbf{r}_1)\phi_j^*(\mathbf{r}')\psi^*(\mathbf{r}_1, \ldots, \mathbf{r}_N)$$

$$\times \hat{H}_1 \psi(\mathbf{r}', \mathbf{r}_2, \ldots, \mathbf{r}_N) d\mathbf{r}' d\mathbf{r}_1 \ldots d\mathbf{r}_N \,. \tag{7.15}$$

$\hat{H}_1$ consists of the terms in the N-electron Hamiltonian involving coordinate $\mathbf{r}_1$, so that

$$\hat{H}_1 = \hat{h}_1 + \sum_{j \neq 1}^{N} v(\mathbf{r}_1, \mathbf{r}_j) \,, \tag{7.16}$$

where $\hat{h}$ consists of the one-body kinetic energy operator and ionic potential, including both local and non-local pseudopotential components, and v is the electron-electron interaction potential.

If a HF wave function is used for ψ^N and the density matrix is expanded in a basis set of HF orbitals, then V^v_{ij} reduces to a matrix with the HF N-particle eigenvalues on the occupied part of the diagonal, and zeros everywhere else. The resulting excitation energies are those given by the well-known Koopmans' theorem. [112] The contents of the EKT method are now reasonably clear. The method consists of a quasiparticle-like Ansatz for the wave function of the N-1 particle system, which is used to calculate the ionisation energies of the system. Electron correlations are included, but no allowance is made for relaxation of the other orbitals in the presence of the excitation. Although this relaxation can be important in small systems, it is expected to be much less important for excitations in extended systems such as the silicon crystal studied here.

7.5.2 Conduction band energies

An analogous theory exists for the conduction band energies. The wave function for the N+1 electron system is approximated by the Ansatz of adding an orbital to ψ^N:

$$\psi^{N+1}(\mathbf{r}_0, \ldots, \mathbf{r}_N) = \hat{\mathcal{A}} u_c(\mathbf{r}_0) \psi^N(\mathbf{r}_1, \ldots, \mathbf{r}_N) \,, \tag{7.17}$$

where $\hat{\mathcal{A}}$ is the antisymmetriser and the orbital u_c is to be determined variationally. In second quantization we have

$$|\psi^{N+1} >= \hat{\mathcal{O}}^\dagger_c |\psi^N > \;\; . \tag{7.18}$$

The excitation energies ϵ_c are defined by

$$\epsilon_c = \frac{\langle \psi^N | \hat{\mathcal{O}}_c [\hat{H}, \hat{\mathcal{O}}^\dagger_c] | \psi^N \rangle}{\langle \psi^N | \hat{\mathcal{O}}_c \hat{\mathcal{O}}^\dagger_c | \psi^N \rangle} \,. \tag{7.19}$$

Expanding in a set of orbitals gives $u_c(\mathbf{r}) = \sum c_{ic} \phi_i(\mathbf{r})$ and $\hat{\mathcal{O}}^\dagger_c = \sum c_{ic} \hat{a}^\dagger_i$. The coefficients c_{ic} are the solutions of the secular equation

$$(\mathbf{V}^c - \epsilon_c \mathbf{S}^c)\mathbf{c}_c = \mathbf{0}\,, \tag{7.20}$$

where the matrix elements are $V^c_{ij} = \langle \psi^N | \hat{a}_i [\hat{H}, \hat{a}^\dagger_j] | \psi^N \rangle$ and $S^c_{ij} = \langle \psi^N | \hat{a}_i \hat{a}^\dagger_j | \psi^N \rangle = \delta_{ij} - \rho_{ij}$.

It is instructive to introduce a new potential with matrix elements $V_{ij} = V^v_{ij} + V^c_{ij}$:

$$\begin{aligned} V_{ij} \quad = \int \quad & \phi_i(\mathbf{r}_0)\hat{h}_0 \phi^*_j(\mathbf{r}_0) d\mathbf{r}_0 \\ +N\int \quad & \phi_i(\mathbf{r}_0)\phi^*_j(\mathbf{r}_0) v(\mathbf{r}_0, \mathbf{r}_1) \\ & \times |\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)|^2 d\mathbf{r}_0 d\mathbf{r}_1 \ldots d\mathbf{r}_N \\ -N\int \quad & \phi_i(\mathbf{r}_0)\phi^*_j(\mathbf{r}_1) v(\mathbf{r}_0, \mathbf{r}_1) \psi(\mathbf{r}_0, \mathbf{r}_2, \ldots, \mathbf{r}_N) \\ & \times \psi^*(\mathbf{r}_1, \ldots, \mathbf{r}_N) d\mathbf{r}_0 d\mathbf{r}_1 \ldots d\mathbf{r}_N\,. \end{aligned} \tag{7.21}$$

If an HF wave function a basis of HF orbitals is used, V_{ij} reduces to a matrix with the HF N-particle eigenvalues on the diagonal, and zeros everywhere else. In this case one can readily identify the second and third terms in equation 7.21 as, respectively, the Hartree and exchange terms. (Similarly V^c_{ij} reduces to the HF energy eigenvalues on the unoccupied part of the diagonal.)

7.5.3 VMC formulation of the EKT

The matrix elements of V^v_{ij} and V_{ij} were accumulated, subsequently forming the matrix $V^c_{ij} = V_{ij} - V^v_{ij}$. The matrix elements, V^v_{ij}, are given by

$$\begin{aligned} V^v_{ij} = N \int \phi_i(\mathbf{r}_1)\phi^*_j(\mathbf{r}') & \frac{\hat{H}_1 \psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)}{\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)} \\ \times \frac{\psi(\mathbf{r}', \mathbf{r}_2, \ldots, \mathbf{r}_N)}{\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)} & |\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)|^2 d\mathbf{r}' d\mathbf{r}_1 \ldots d\mathbf{r}_N\,. \end{aligned} \tag{7.22}$$

As before, permutation symmetry allows this to be rewritten as

$$\begin{aligned} V^v_{ij} = \Bigg\langle \sum_{n=1}^{N} \int \phi^*_i(\mathbf{r}_n)\phi_j(\mathbf{r}') & \frac{\hat{H}_n \psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)}{\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)} \\ & \times \frac{\psi(\ldots, \mathbf{r}', \ldots)}{\psi(\ldots, \mathbf{r}_n, \ldots)} d\mathbf{r}' \Bigg\rangle_{|\psi|^2}\,, \end{aligned} \tag{7.23}$$

so that N values are accumulated at each step. The terms contributing to $\hat{H}_n\psi/\psi$ are already available in a VMC calculation, allowing V_{ij}^{v} to be accumulated with virtually no additional cost beyond that required for the density matrix. An analogous VMC formulation for calculating V_{ij} exists. The single particle terms, $\hat{h}_0$, appearing in the first term in equation 7.21 were evaluated directly without use of Monte Carlo integration. Using Monte Carlo integration for all the terms resulted in a small increase in the variance of the matrix elements. It is therefore preferable, but less elegant, to calculate the $\hat{h}_0$ terms directly. The matrix elements V_{ij}^{v} and V_{ij} were accumulated at the same time as the elements ρ_{ij}. The full crystal symmetry was again used, which reduces the number of matrix elements which must be accumulated and ensures the correct symmetry in the presence of statistical noise.

7.5.4 Results

Full EKT

The correlation lengths along the VMC walks for V_{ij}^{v} and V_{ij} were found to be similar to those of the local energy and density matrix, and the distribution of statistical errors was similar to that for the density matrix. The elements of V_{ij}^{v} and V_{ij}^{c} with the largest statistical errors were the diagonal elements. The statistical error bars on these elements are estimated to be ± 0.2 eV.

The matrix equations 7.13 and 7.20 were diagonalised using a generalized eigenvalue solver. The ratios of the statistical error bars to the mean values were significantly larger for the diagonal elements of V_{ij}^{v} and V_{ij}^{c} than for ρ_{ij}. Just as for the density matrix, small perturbations were applied to estimate the statistical errors in the eigenvalues and eigenvectors. The numerical stability of the diagonalisation was improved by explicitly zeroing elements of V_{ij}^{v} and V_{ij}^{c} within statistical error of zero. The accuracy of the eigenvalues was further verified by gradually increasing the number of bands in the diagonalisation procedure. The valence and low lying conduction band energies were stable to within ± 0.4 eV.

The resulting band energies are shown in figure 7.3 and given in table 7.2, with the energy at the top of the valence band set to zero. Of the **k**-points computed here, the available experimental data (Exp) is limited to the Γ-point. Because of this we also give empirical pseudopotential (Emp) data [135] in table 7.2 and in figure 7.3. The empirical pseudopotential data should provide a good interpolation between this data. Results for the Ewald and MPC interactions are in good agreement, indicating that the finite size errors are not significant at the level of statistical accuracy obtained here. The energies are in good qualitative agreement with the empirical data at all **k**-points except for the upper Σ_1 valence band state, and the $\Delta_{2'}$ conduction band state. The source of the error is principally the value of the diagonal matrix elements for these states, V_{ii}^{v} and V_{ii}^{c} (see

Band	VMC				DMC[e]	HF[f]	LDA[g]	GW[h]	Emp[i]	Exp[j]
	EKT[a]	EKT[b]	DEKT[c]	DEKT[d]						
$\Gamma_{2'}$	4.4	4.5	5.1	5.2	4.6	9.0	3.22	3.89	4.1	4.23,4.1
Γ_{15}	3.8	3.9	4.4	4.4	3.7	8.0	2.51	3.36	3.4	3.40,3.05
$\Gamma_{25'}$	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.0	0.00
Γ_1	-12.9	-13.1	-13.3	-13.4	-13.6	-18.9	-11.98	-11.95	-12.36	-12.5±0.6
Λ_3	4.0	4.7	5.4	5.7		9.0	3.46		4.5	
Λ_1	3.0	3.1	3.6	3.6		6.8	1.68		2.5	
Λ_3	-0.6	-1.2	-1.3	-1.5		-1.7	-1.01		-0.9	
Λ_1	-5.0	-5.0	-5.4	-5.4		-6.3	-5.38		-5.2	
Λ_1	-11.6	-11.9	-12.3	-12.5		-16.4	-10.60		-10.6	
$\Delta_{2'}$	4.7	4.8	5.7	6.0		7.2	1.88		2.4	
Δ_1	2.1	2.3	2.7	3.0		5.5	0.61		1.3	
Δ_5	-1.7	-1.8	-2.4	-2.5		-3.8	-2.45		-2.5	
$\Delta_{2'}$	-4.9	-4.8	-5.5	-5.7		-7.8	-5.01		-5.0	
Δ_1	-11.6	-12.0	-11.7	-12.2		-16.0	-10.08		-10.2	
Σ_1	6.4	5.4	6.7	6.1			5.73		6.0	
Σ_3	3.4	3.1	4.1	4.0			1.45		2.0	
Σ_2	-1.1	-1.6	-1.2	-2.0			-2.14		-2.1	
Σ_1	-1.8	-2.5	-1.9	-2.5			-4.50		-4.6	
Σ_3	-6.1	-6.3	-6.4	-6.8			-6.79		-6.7	
Σ_1	-8.9	-9.0	-9.5	-9.8			-8.70		-8.7	

Table 7.2: Band energies of silicon in eV. a,b- VMC-EKT energies using the MPC and Ewald interactions, respectively. The statistical error bars are ±0.4 eV. c,d- Diagonal approximation to the VMC-EKT energies using the MPC and Ewald interactions, respectively. The statistical error bars are ±0.2 eV. e- Direct DMC calculations, with statistical error bars of ±0.2 eV, from Ref. [103]. f- Ref. [105]. g- This work. h- Ref. [136]. i- Ref. [135] j- From the compilation given in Ref. [136].

next section).

The EKT band energies at the Γ-point are in good agreement with the available experimental data, and are also in good agreement with the DMC data from direct calculations of the excitation energies. [103] The EKT valence band width of 12.9(6) eV is smaller than the value of 13.6(3) eV obtained from the DMC calculations and is in good agreement with the experimental value of 12.5(6) eV. In comparison the HF data shows the well known overestimation of band gaps and band widths which is due to the neglect of correlation energy, the LDA gives excellent valence band energies while the conduction bands are too low in energy by 0.7-1.0 eV, and the GW data is in very good agreement with experiment.

The EKT is a formulation for the eigenstates of the N-1 and N+1 electron systems, while the direct method is aimed at calculating the eigenstates of the N electron system. The EKT and direct results should therefore differ by the exciton binding energy, but this energy is small for silicon and cannot be resolved at the level of statistical accuracy obtained. Depending on the application one would like to be able to choose whether to include excitonic effects, so that it is advantageous to have the different QMC techniques available. Clearly further refinement of excited state QMC methods is required, but the results from the EKT and direct approaches are promising for the study of more strongly correlated systems, for which the LDA and GW methods are less reliable.

7.5.5 Approximations to the EKT

If the off-diagonal elements of V_{ij}^v, V_{ij}^c, and ρ_{ij} are neglected then the valence and conduction band energies can be approximated by

$$\epsilon_{iv}^{\mathrm{DEKT}} = \frac{V_{ii}^v}{\rho_{ii}} \; , \; \epsilon_{ic}^{\mathrm{DEKT}} = \frac{V_{ii}^c}{1-\rho_{ii}} \; , \tag{7.24}$$

where the superscript DEKT denotes the diagonal approximation to the EKT. This approximation has been used within VMC and without reference to the EKT by Fahy *et al.* [121] to calculate quasihole energies in silicon and also by Tanaka [122] to calculate the quasihole energies in NiO. This approximation has the computational advantage that far fewer matrix elements are required, and also that the problems of statistical noise are reduced, because the values of only two matrix elements enter the calculation of each band energy. However the results are basis set dependent, and could differ significantly between, for example, a HF and LDA basis.

If an HF wave function and basis set of HF orbitals is used then the matrices V_{ij}^v, V_{ij}^c, and ρ_{ij} are all diagonal, and consequently the full EKT and the DEKT are equivalent. In general, for correlated wave functions, the DEKT gives neither an upper nor lower bound to the energy obtained from the full EKT. Comparison with the data from the DEKT is made in table 7.2. The DEKT values are close to the full EKT values, including the two cases mentioned above where the agreement with experiment is poor. The DEKT works quite well for the valence bands and slightly less well for the conduction bands, because the off-diagonal matrix elements of V_{ij}^c coupling the unoccupied states are more significant. In similar VMC calculations for silicon using the DEKT and a smaller simulation cell of 64 electrons, Fahy *et al.* [121] calculated a valence band width of 14.5(4) eV, which is larger than the value of 13.3(2) eV and the experimental value of 12.5(6) eV. Fahy *et al.* [121] also obtained occupation numbers of 1.96(4) and 1.92(3) for the Γ_1 and $\Gamma_{25'}$ valence band states respectively, compared with the values of 1.9817(2) and 1.9375(5) obtained in the larger cell of this study.

The scaling with system size of the diagonal approximation is very advantageous compared with direct evaluation of excitation energies. In direct methods, the fractional energy change due to the promotion of an electron is considered. This energy change is inversely proportional to the number of electrons in the system, i.e., a '$\frac{1}{N}$' effect. The precision of the calculation must therefore be sufficient to resolve the energy change amid the statistical noise. This requirement is very challenging for small band gap materials, as the system must also be sufficiently large to approximate the infinite solid. In the DEKT, the band energies are computed directly as averages over the square of the ground state wave function, rather than as the difference between averages over the squares of the ground and excited state wave functions. This greatly improves the sampling statistics. We have found that the errors in V_{ii}^v, V_{ii}^c and ρ_{ii} scale as the inverse of the square root of the number of independent electron configurations, independently of the system size. For sufficiently large systems, it therefore requires fewer statistically independent samples in the DEKT to obtain excitation energies to a given statistical accuracy than with direct methods. The scaling behavior of the full EKT is more difficult to analyze because we are required to diagonalise noisy matrix equations, where the number of off-diagonal matrix elements increases as the square of the system size, and where there are certainly statistical correlations between matrix elements.

7.6 Conclusions

The one-body density matrix for the valence electrons of bulk silicon was calculated using QMC techniques.

In real space, the VMC and LDA density matrices are very similar, the greatest differences being due to the differing charge densities in each case. The natural orbitals, obtained by diagonalising the density matrix, very closely resemble the LDA orbitals. The occupation numbers of the natural orbitals differ significantly from the non-interacting values, reducing linearly with increasing energy for the LDA occupied bands. The occupations are about 3% lower than the non-interacting value near the top of the valence band, and above the Fermi level, the occupation numbers fall slowly to zero.

A DMC calculation for the ground state energy of silicon using a trial wave function containing a determinant of natural orbitals gives an energy which is almost identical to that obtained using a determinant of LDA orbitals. This shows that the quality of the nodal surfaces is almost identical in each case. A DMC calculation comparing the quality of the nodal surfaces obtained with LDA and HF orbitals, found that for the system studied the nodal surface of a determinant of LDA orbitals is marginally better.

The relationship between Compton profiles, momentum densities and density matrices was discussed. The zero momentum component of the momentum density was computed from the density matrix. Its value was found to be in good agreement with other calculations, supporting the conclusion that the impulse approximation used in the analysis of Compton profiles is the most likely source of the current discrepancy between calculations and experiment.

A Monte Carlo formulation of the Extended Koopmans' theorem was applied to bulk silicon, and the resulting band energies are in reasonable agreement with the available experimental data. The success of the VMC-EKT relies on a cancellation of errors between the ground and excited state energies. The wave functions for the excited states contain the variational freedom of the orbitals u_v and u_c. This variational freedom in the orbitals reduces the energy of the N-1 and N+1 electron states and improves the agreement with experiment. In the diagonal approximation to the EKT (DEKT), the u_v and u_c orbitals are fixed and there is no variational freedom in the excited state wave functions. We have found that the DEKT works quite well for silicon using LDA orbitals for u_v and u_c. The diagonal approximation is exact within HF theory, so that it is expected to be a good approximation for weakly correlated systems.

Greater accuracy could be obtained with more accurate trial functions, or using DMC in the calculation of V_{ij}^v, V_{ij}^c, and ρ_{ij}. In comparison with direct methods of calculating excitation energies, [42, 101, 103] the EKT has the advantage that only a single calculation involving the ground state is required to obtain many excitation energies. Direct methods methods require that accurate energies are computed for each state of interest and excitation energies are then computed via energy differences, which is very inefficient.

The EKT involves assumptions about the nature of the excited state wave functions, but nevertheless is a practical method for calculating excitation energies including correlation effects for weakly correlated systems. The diagonal approximation to the EKT has a very advantageous scaling with system size compared with direct QMC calculations. This scaling thereby allows the study of excitation energies in larger systems with a greater efficiency than is possible with direct techniques.

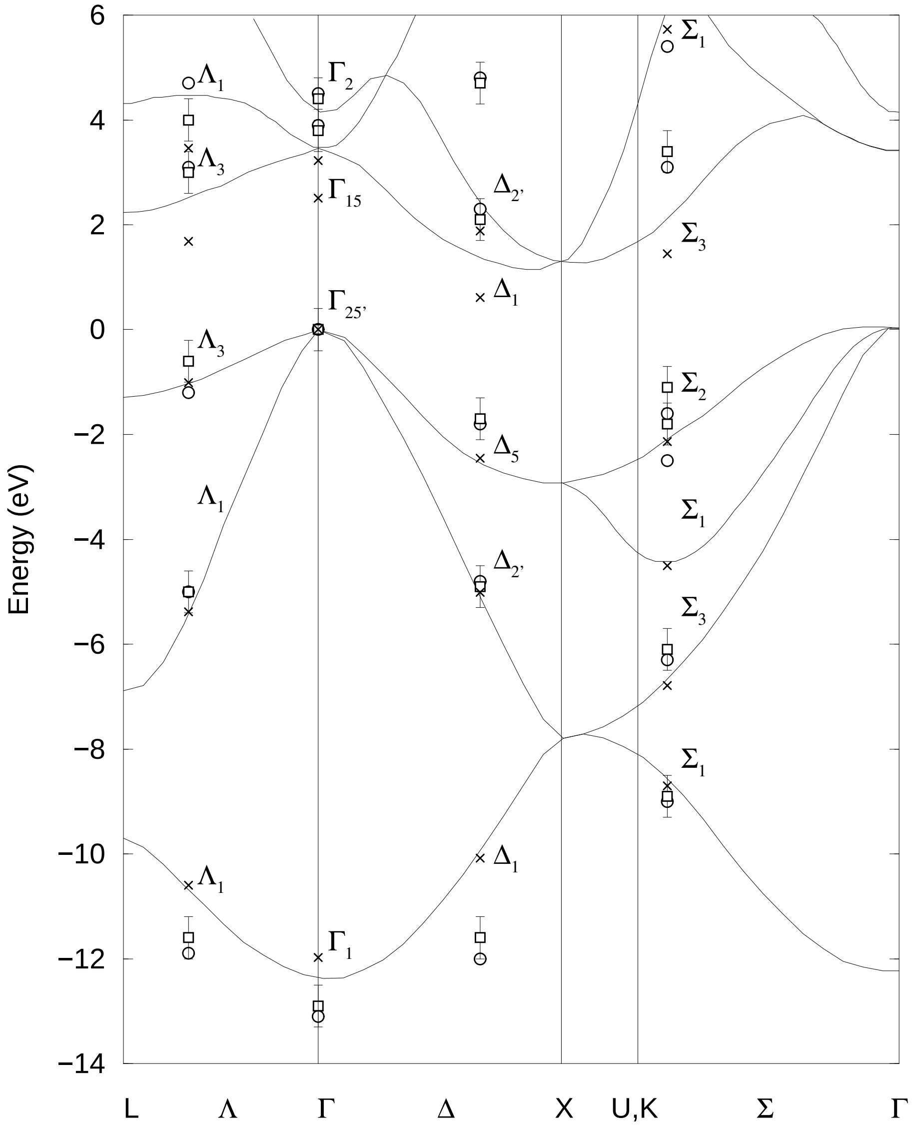

Figure 7.3: The band structure of silicon, as computed within the VMC Extended Koopmans' Theorem using the MPC interaction (□) and the Ewald interaction (○), and within the LDA (×). The results for the MPC interaction are shown with statistical error bars. The statistical error bars for the Ewald results have been omitted for clarity, but they are the same size as for the MPC interaction. As a guide to the eye, the empirical pseudopotential data of [135] is shown (solid lines), which is in good agreement with the available experimental data.

Chapter 8

The energetic stability of small carbon clusters

8.1 Introduction

In this chapter, the energetic stability of isomers of C_{24}, C_{26}, C_{28} and C_{32} carbon clusters is studied using density functional and quantum Monte Carlo methods, with the aim of identifying the smallest energetically stable fullerene. Calculating the energy of small carbon clusters is a strong test of electronic structure methods due to their widely differing geometries and the energetic importance of electron correlation. Recent experimental results comparing the relative abundance of small carbon clusters, including fullerenes, demonstrate a sensitive dependence on experimental conditions and many experimental results are in disagreement with previous electronic structure calculations. By using diffusion Monte Carlo, electron correlation is treated consistently, and accurate relative energies are obtained for systems of up to 128 valence electrons. The results also bear on the existence of fullerene cluster based solids, which at the time of writing have been the subject of considerable theoretical speculation but have not been synthesised in quantity.

8.2 Background

Carbon clusters have been the subject of research for decades. [137] However, the discovery of the fullerene structure of C_{60}, [138] (figure 8.1) led to a renewal and subsequent explosion of interest in the properties of the clusters. An understanding of the physical and chemical properties of these clusters is important for a large variety of systems; they have a role in the chemistry of comets, carbon stars and molecular clouds (References 2-9 in [139]), and their properties are important in chemical vapour deposition processes such as the growth

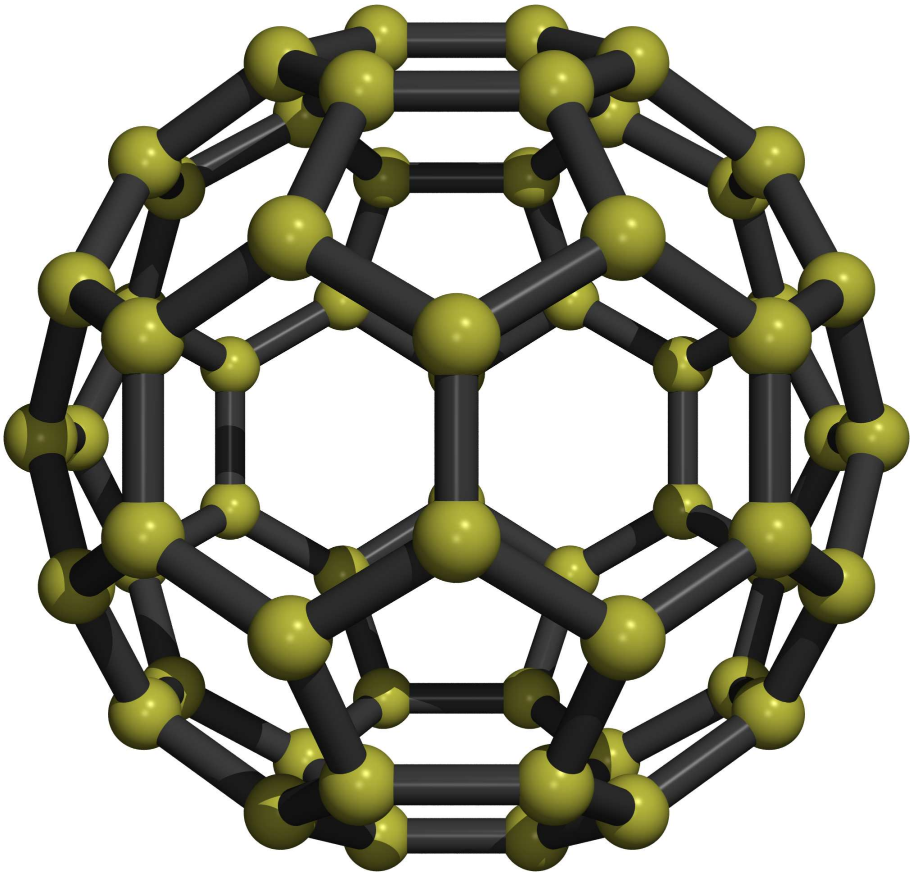

Figure 8.1: The C_{60} "buckyball" fullerene.

of diamond, silicon carbide and carbon nitride thin films.

The discovery of the fullerenes, defined as a closed cage structures containing only pentagonal and hexagonal faces, [140] has subsequently generated a substantial body of research dedicated to their structure and molecular properties. Fullerene clusters may now be produced in macroscopic quantities, although structure and mass selection still remains something of an art. A rich variety of physical and chemical properties has been demonstrated, but despite many experimental and theoretical advances the detailed energetics of these systems are not yet fully understood. The question "which is the smallest stable fullerene?" remains both interesting and contentious due to the sensitivity of cluster formation to experimental conditions and the challenges posed to accurate theoretical methods by system size.

A 20 atom cluster is the smallest geometrically able to form a fullerene. [140] However, the smallest stable fullerenes most commonly identified in experiments are the C_{30} and C_{32} clusters. [141–143] The sizes and types of cluster are usually identified experimentally by

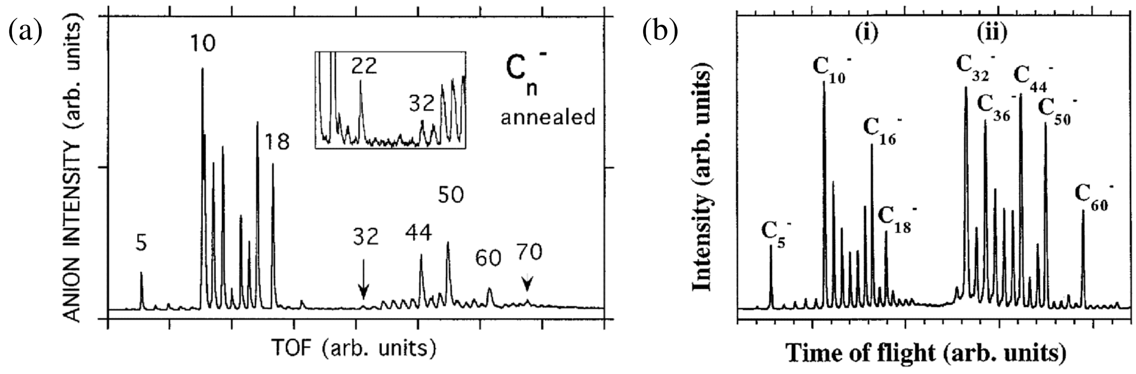

Figure 8.2: Time of flight mass spectra of carbon clusters from two experiments. (a) from H. Handschuh *et al.* [141] and (b) from H. Kietzmann *et al.*, [143] where a higher abundance of C_{32} than C_{60} was observed. The difference in abundance between the two experiments is mostly due to an optimised annealing of the clusters in (b).

mass spectrometry and time of flight measurements. Under typical experimental conditions, rings are found to dominate up to approximately 28 carbon atoms (see for example, Ref. [142]), while for clusters of more than 28 atoms, fullerenes are mostly observed, although rings continue to be present. Figure 8.2 illustrates the influence of experimental conditions on relative cluster abundance.

8.3 Previous work

Carbon clusters have been extensively studied using electronic structure methods such as tight-binding, Hartree-Fock, density functional, quantum chemical and quantum Monte Carlo methods. [43, 139, 144–147] The need for calculations with a sophisticated treatment of electron correlation has been clearly illustrated by several previous studies.

A DFT study of C_{20} isomers showed that the local density approximation (LDA) [88] and the BLYP gradient corrected functional [148, 149] gave different energy orderings for the sheet, bowl and fullerene structures, with neither ordering agreeing with the diffusion quantum Monte Carlo (DMC) ordering. [43] Recent calculations for the monocyclic ring and fullerene of C_{24} have shown substantial differences between the predictions of the LDA, gradient corrected and hybrid density functionals. [150] Second-order Moller-Plesset (MP2) and coupled cluster calculations at the CCSD(T) level were also performed, and the authors concluded that the fullerene lies lower in energy than the ring. Raghavachari *et al.* [151] and Martin *et al.* [152] performed DFT calculations on seven isomers of C_{24}, but obtained conflicting energetic orderings.

Grossman *et al.* [43] have performed pseudopotential DMC calculations for C_{20} clusters, finding that the fullerene isomer of C_{20} is not energetically stable. They also demonstrated in tests on a small series of hydrocarbons, CH_4 to C_6H_6, that pseudopotential DMC give close to "chemical accuracy" of 1–2 kcal mol^{-1}. In an earlier study, Greef *et al.* [113] found similar accuracies for a series of silanes, SiH to Si_3H_8. DMC appears an ideal method for both the study of these molecular systems and for benchmarking other electronic structure methods.

Carbon clusters are challenging to model accurately due to the wide range of geometries and the occurrence of single, double, and triple bonds. These differences result in a non-cancellation of errors in relative energies, exaggerating errors or approximations made in calculations. QMC is therefore ideal for systematically investigating the energetic ordering of different isomers due to the consistent treatment of electron correlation. The approach selected here was to use DFT calculations as a basis for selecting low energy structures and determining geometries, followed by DMC calculations for a more accurate treatment of electron correlation and better ground state total energy.

8.4 Methodology

The range of isomers studied, C_{24} to C_{32}, was selected to cover the range of masses from where the fullerene is clearly predicted to be unstable to where a fullerene is clearly observed. This enables the prediction of the smallest energetically stable fullerene.

To determine the lowest energy isomers a hierarchy of methods of increasing accuracy and computational cost is commonly used. The first stage is to select candidate structural isomers via empirical methods such as bond counting and geometric rules such as "minimize the number of adjacent pentagons". [153] Quantum mechanical calculations such as tight-binding and density functional theory (DFT) based methods are then used to refine the selection. In order to finally establish the energetic ordering of different isomers, very accurate calculations must be performed. Although it would be desirable to perform coupled cluster calculations as well as QMC calculations, these are severely limited by the size of basis set that is computationally affordable for systems of the size of interest.

In this study, isomer selection was based on several previous density functional and quantum chemical studies. Geometries were then optimised within DFT, using the B3LYP hybrid functional. [154]

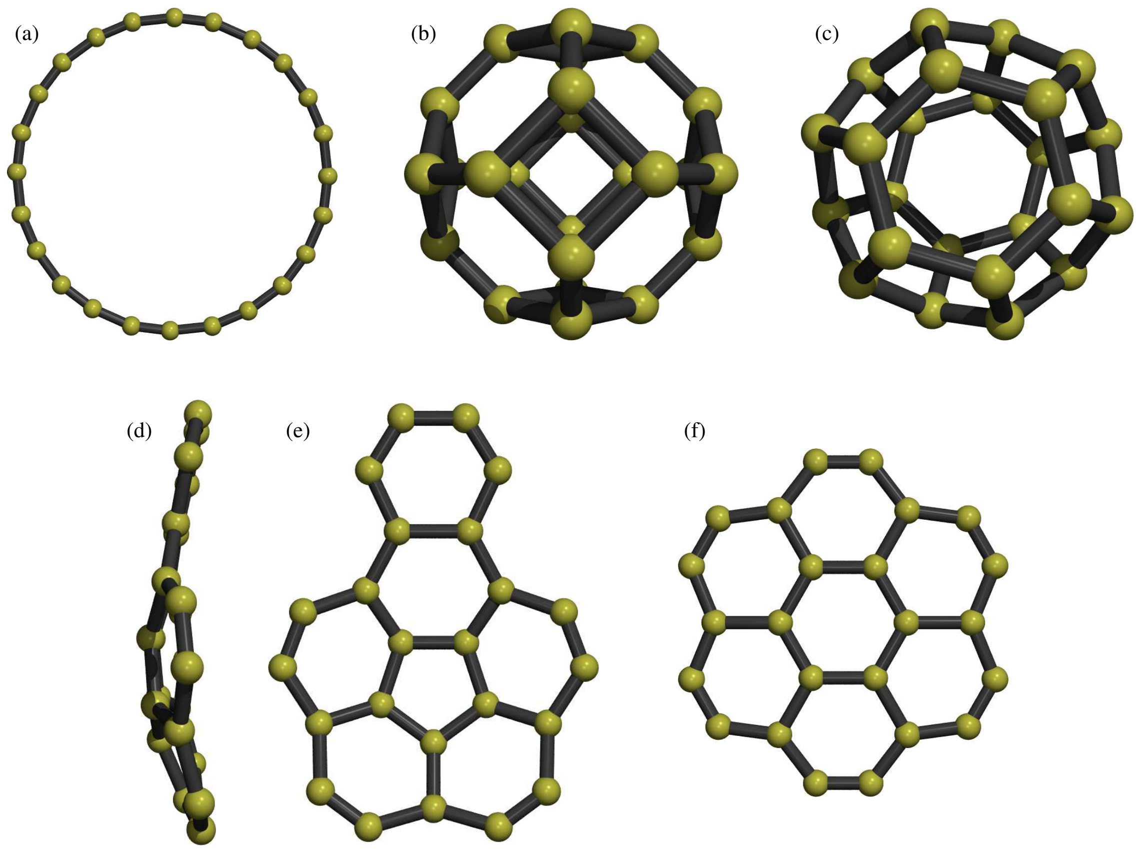

Figure 8.3: Structural isomers of C_{24}: (a) ring, (b) O_h symmetry cage, (c) D_6 symmetry fullerene, (d,e) (side, front) one-pentagon bowl, (f) graphitic sheet.

8.5 Candidate geometries

For clusters containing between 20 and 32 atoms, three types of structural isomer are energetically competitive: fullerenes, planar or near-planar sheets and bowls, and monocyclic rings. Five isomers of C_{24}, three isomers of C_{26}, three isomers of C_{28}, and two isomers of C_{32} were studied. These are illustrated in figures 8.3, 8.4, 8.5,8.6 respectively.

The following isomers of C_{24}, depicted in figure 8.3, were considered: a monocyclic ring, a flat graphitic sheet, a bowl-shaped structure with one pentagon, and a caged structure with a mixture of square, pentagonal and hexagonal faces, and a fullerene. Other candidate structures, such as bicyclic rings and a 3-pentagon bowl were excluded on the grounds that DFT calculations using several different density functionals have shown them to be significantly higher in energy. [151, 152]

Three isomers of C_{26} were studied. A previous density functional study [155] considered 10 isomers, but the effects of basis set incompleteness are apparent and as only two den-

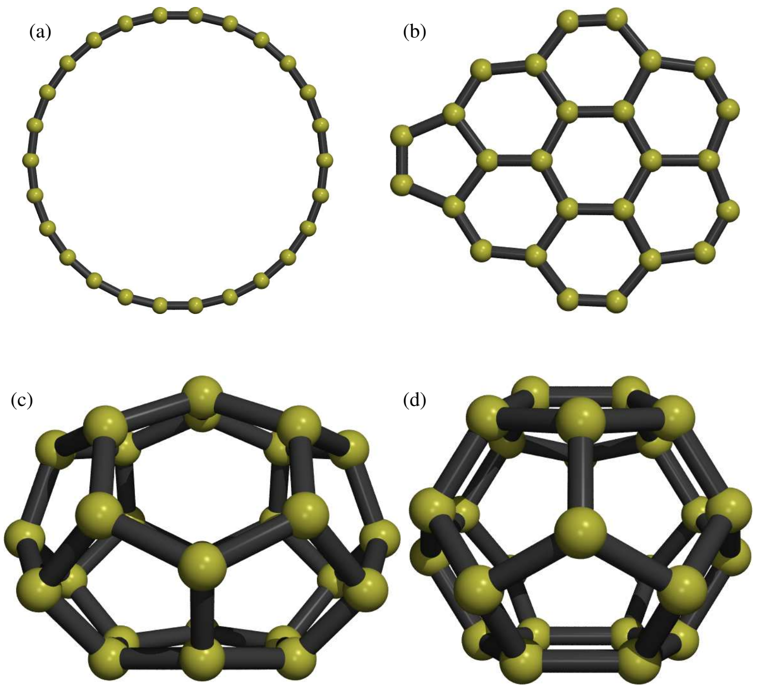

Figure 8.4: Structural isomers of C_{26}: (a) ring, (b) one-pentagon sheet, (c,d) (side, front) C_{2v} symmetry cage.

sity functionals were used, the low energy ordering of the structures is difficult to ascribe. Structures in the study included 3 graphitic sheet–like structures, a monocyclic "cumulenic"[1] ring, a cage and a fullerene. The fullerene, ring and one of the sheets were selected for DMC calculations due to their apparent low energy. The fullerene is significantly distorted and has only C_{2v} symmetry. The sheet resembles the purely graphitic C_{24} sheet but for the addition of a pentagon.

Three structures of C_{28} were investigated: a monocyclic ring, a graphitic sheet and a fullerene of T_d symmetry (figure 8.5). Other bowl and sheet–like structures [156] were excluded on energetic grounds.

Two C_{32} structures were examined: a monocyclic ring and a fullerene (figure 8.6). HF calculations suggest that the fullerene form of C_{32} is energetically favourable, [146] demonstrating the reduced role of electron correlation in determining energetic orderings with increased cluster size. These structures were included to provide additional information about trends in binding energies.

[1]Have only identical bond lengths.

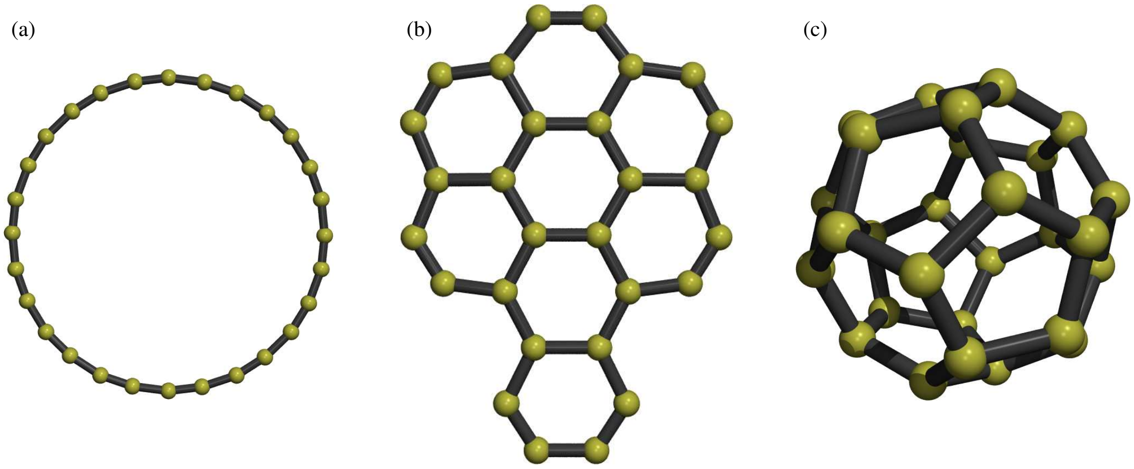

Figure 8.5: Structural isomers of C_{28}: (a) ring, (b) sheet, (c) T_d symmetry fullerene.

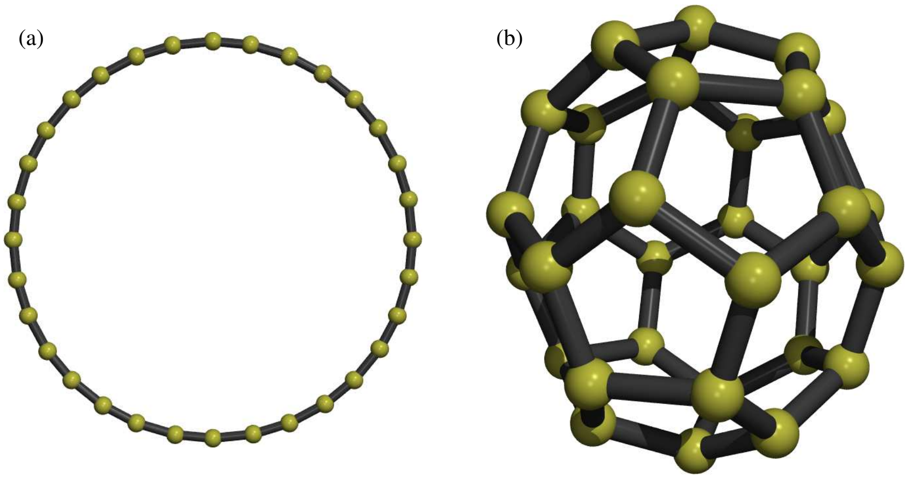

Figure 8.6: Structural isomers of C_{32}: (a) ring, (b) fullerene.

8.6 DFT calculations

The structures were relaxed by performing highly converged density functional calculations. The geometries were obtained from all-electron calculations [157] using the B3LYP hybrid functional [154] and Dunning's cc-pVDZ basis set [158], which has been found an accurate and computationally affordable combination. [152, 156, 159] The B3LYP functional is a so-called *hybrid* functional because it includes a fraction of exact Hartree-Fock exchange.

All-electron, and not pseudopotential calculations were performed in order to facilitate comparison of the geometries obtained with those of previous studies. DFT calculations using the LDA, two gradient corrected functionals (PBE [160] and BLYP) and a hybrid

	Geometry	Method	cc-pVDZ	aug-cc-pVDZ
Ring	B3LYP	B3LYP	-913.896423049	-913.923269541
		BLYP	-913.682193134	-913.716096665
	BLYP	B3LYP	-913.891887356	
		BLYP	-913.687457613	
D_6 Cage	B3LYP	B3LYP	-913.905987459	-913.931247631
		BLYP	-913.628848548	-913.660348572
	BLYP	B3LYP	-913.901588045	-913.926145351
		BLYP	-913.633002984	-913.663869638

Table 8.1: Basis set and geometry dependence of C_{24} isomers from all-electron density functional calculations. Energies are in a.u. per molecule. Sufficiently converged results were not obtained for the BLYP geometry, aug-cc-pVDZ basis.

functional (B3LYP) were then performed at the B3LYP optimised geometries for comparison with the DMC results.

The assertion that cc-pVDZ basis sets are sufficient [152, 156, 159] was tested by performing a series of calculations with Dunning's more extensive basis sets. The cc-pVDZ basis is an all-electron basis designed for correlated calculations consisting of two large s-type contracted sets of Gaussians for the 1s electrons, an additional uncontracted s-type Gaussian, a single p-type contraction, an additional p-type Gaussian and a single d-type Gaussian. Dunning [158] also defined an augmented basis set, aug-cc-pVDZ which augments the cc-pVDZ basis by an additional s,p and d Gaussians.

The ring and D_6 fullerene isomers of C_{24} (see figure 8.3) which have significantly different coordination. The results of these calculations are given in table 8.1. Although the cc-pVDZ basis is "incomplete", the relative energies obtained are very similar to those obtained at the aug-cc-pVDZ level indicating the cc-pVDZ basis is probably sufficient for the structures tested.

The sensitivity of the total energies to changes in the geometries was assessed by comparing the energies of the fully relaxed ring and D_6 fullerene isomers of C_{24} using the BLYP and B3LYP functionals. These functionals give significantly different energetic orderings for the isomers. The differences between the BLYP and B3LYP geometries were small - less than 0.03 angstroms in bond lengths and less than 0.4 degrees in bond angles. These differences in structure give rise to changes in the relative energies of the ring and fullerene of less than 0.27 eV for each of the density functionals investigated. The relative energies are therefore rather insensitive to the density functional used to obtain the geometries, but they are more sensitive to the density functional used to calculate the relative energies. Changes in the relative energies of less than 0.27 eV are small compared with the overall spread of energies, but some changes in the orderings of the isomers closest in energy could occur.

8.7 DMC calculations

DMC calculations were performed using the methods and algorithms of chapters 3 and 4. Core electrons were modelled by an accurate norm-conserving LDA pseudopotential, [58]. The guiding wave functions were of the Slater-Jastrow type, consisting of the product of a Slater determinant of single-particle orbitals obtained from the CRYSTAL [130] or Gaussian94 [157] codes multiplied by a "homogeneous" Jastrow factor (see section 4.3.2). For the single particle orbitals uncontracted valence Gaussian basis sets of four s, four p and one d function were used. The basis sets were optimised for each structural class of isomer (ring, cage, bowl and sheet), yielding basis sets of high efficiency.

Up to 80 parameters in the Jastrow factor were optimized for every structure using the method of variance minimization described in chapter 5. In excess of 75% of the DMC correlation energy was obtained in all cases. No attempt to obtain "benchmark" accuracy VMC results was made: the aim was to obtain wavefunctions of sufficient accuracy for use in DMC calculations.

For the fixed–node DMC calculations, a timestep of 0.003 a.u. and an ensemble of 640 walkers were used. The DMC calculations averaged ~ 30000 moves in length.

A special consideration was made for the C_{26} ring. The C_{26} ring is "cumulenic" whereas the other rings in the study are "polyacetylenic"[2] and have two unique bond lengths. MCHF and CI calculations show that the C_{4N+2}, $N = 1, 2, \ldots$ cumulenic rings have significant multi-determinantal character, unlike the C_{4N} rings. [161] Due to the likely improvement in the trial wavefunction, a CISD calculation was performed and the configurations with largest weight used to form a multi-determinant trial wavefunction. 43 determinants, including the HF reference determinant were selected. Although only a small reduction in the variance of the trial wavefunction resulted (by ~ 0.1 a.u.^{2} to ~ 5.4 a.u.), a small reduction in DMC energy was found due to the improved nodal structure of the wavefunction. Results comparing single and multi-determinant wavefunctions for this isomer are presented in the next section.

[2] Have alternating bond lengths.

8.8 Results and analysis

8.8.1 C_{24}

The results shown in figure 8.7 confirm that the treatment of correlation has a profound effect on the relative energies. All of the density functionals give different orderings of the energies, and none gives the same ordering as DMC. The graphitic sheet is placed lowest in energy by DMC, in agreement with each of the density functionals except BLYP, which places the ring lowest in energy. The low energy of the C_{24} graphitic sheet is expected because the structure accommodates a large number (7) of hexagonal rings without significant strain. This structure is expected to be the smallest stable graphitic fragment, as for smaller clusters ring structures should be lower in energy. Neither DMC or the density functionals find the C_{24} fullerene to be energetically stable.

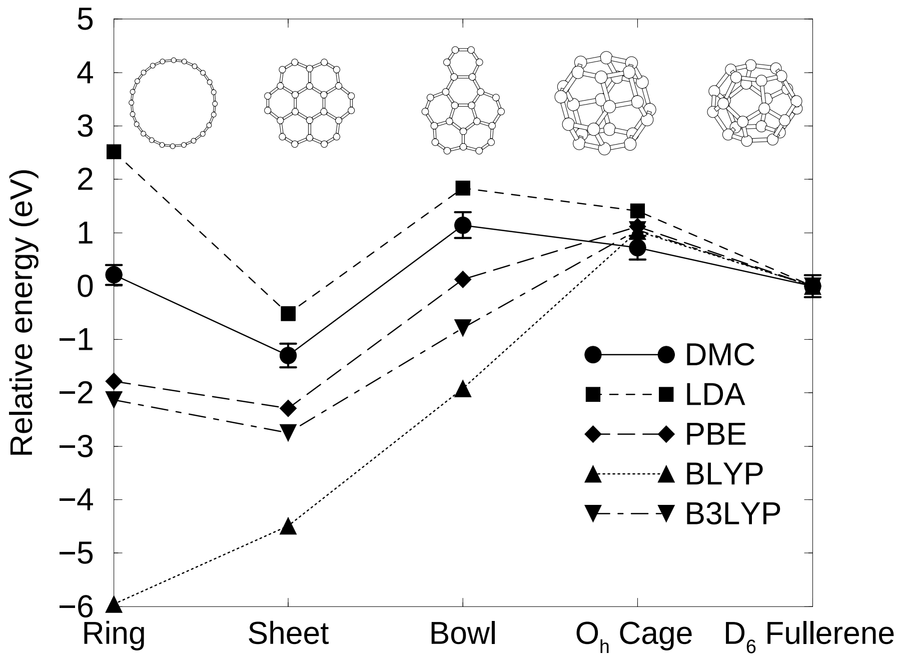

Figure 8.7: Energies of the C_{24} isomers relative to the D_6 fullerene.

8.8.2 C_{26}

The DMC and DFT results shown in figure 8.8 for C_{26} display many of the features established by the C_{24} results. A similar degree of scatter between the different methods is evident. Despite the large strains apparent in the fullerene structure, the fullerene is placed lowest in energy by the DMC and the LDA calculations. The BLYP functional favours the low coordination ring placing it low in energy, and the gradient corrected functionals behave as in C_{24}.

The multi-determinant wavefunction gives a slightly lower DMC energy than the single determinant wavefunction, confirming that CI wavefunctions have better nodal surfaces than HF wavefunctions. However, the ring and sheet-like isomers remain very close in energy, approximately 2.5 eV above the fullerene.

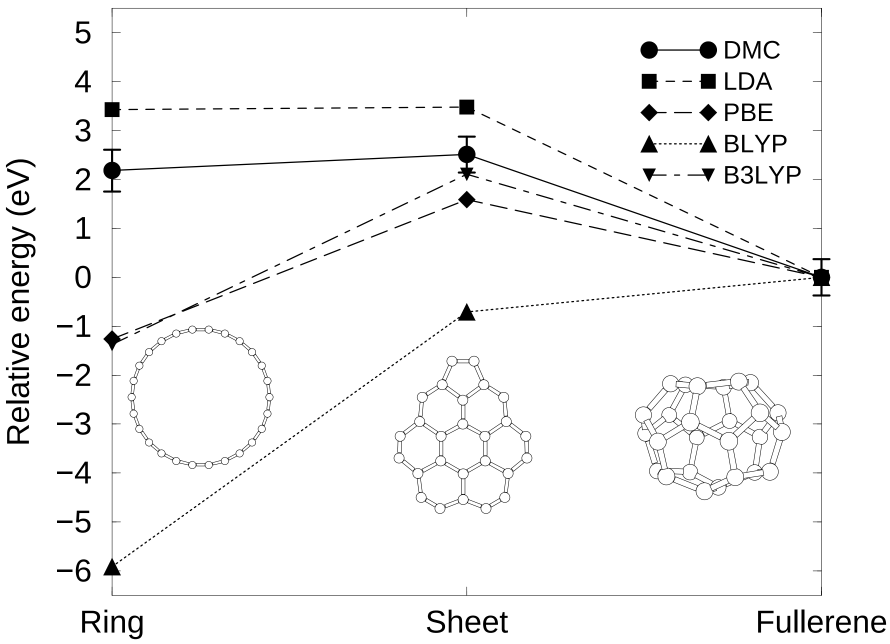

Figure 8.8: Energies of the C_{26} isomers relative to the fullerene.

8.8.3 C_{28}

The DMC results for C_{28}, figure 8.9, again highlight a wide variation in the performance of different DFT functionals. The LDA and B3LYP functionals predict the same ordering as DMC, but the PBE and BLYP functionals predict different orderings.

Spin-polarized DFT calculations show the ground state of the T_d symmetry fullerene to be a spin-polarized $^5A_2$ state. DMC predicts that this spin-polarized fullerene is the lowest energy isomer of C_{28}, and this is supported by each of the density functionals except BLYP.

The spin-polarized fullerene has four unpaired electrons and is therefore reactive. This property has been exploited in atom trapping experiments in which C_{28} fullerenes containing a single atom have been prepared by laser vaporization of a graphite-MO_2 (M =

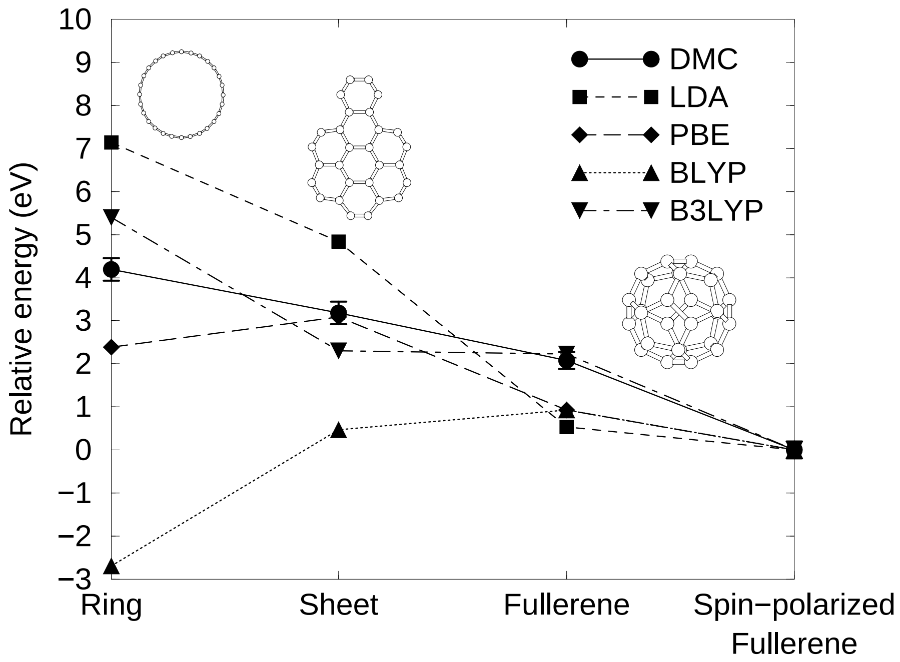

Figure 8.9: Energies of the C_{28} isomers relative to the spin-polarized fullerene.

Ti, Zr, Hf or U) composite rod. [162] The prediction that the fullerene is the most stable isomer of C_{28} indicates that isolated fullerenes might be produced, thereby facilitating the investigation of possible C_{28} fullerene solids, which have been discussed but not yet produced. [162, 163] (A C_{36} fullerene solid has recently been reported. [164])

The DMC data strongly indicates that, of those considered, the T_d fullerene is the most stable C_{28} isomer. The T_d fullerene has the lowest DMC energy in both spin-polarized and non spin-polarized states, and is substantially more stable than the sheet and ring. DMC calculations finds the spin-polarized state to be 2.08(20) eV more stable than the unpolarized state, which is 1.10(26)eV more stable than the sheet. Small changes in the geometries are therefore unlikely to change the energetic ordering.

8.8.4 C_{32}

DMC calculations for a monocyclic ring and fullerene of C_{32} showed that the fullerene is 8.4(4) eV per molecule lower in energy. This is consistent with the observation of a large abundance of C_{32} fullerenes in recent experiments. [143]

8.8.5 Ring and fullerene binding energies

In figure 8.10 the DMC binding energies of the ring and fullerene structures for C_{24}, C_{26}, C_{28} and C_{32} are shown. The binding energy of the fullerenes rises much more rapidly with cluster size than for the rings because of the large amount of strain energy in the smaller fullerenes. The DMC binding energy per atom of the C_{32} fullerene is $\approx$ 1 eV per atom smaller than the experimental binding energy of C_{60}.

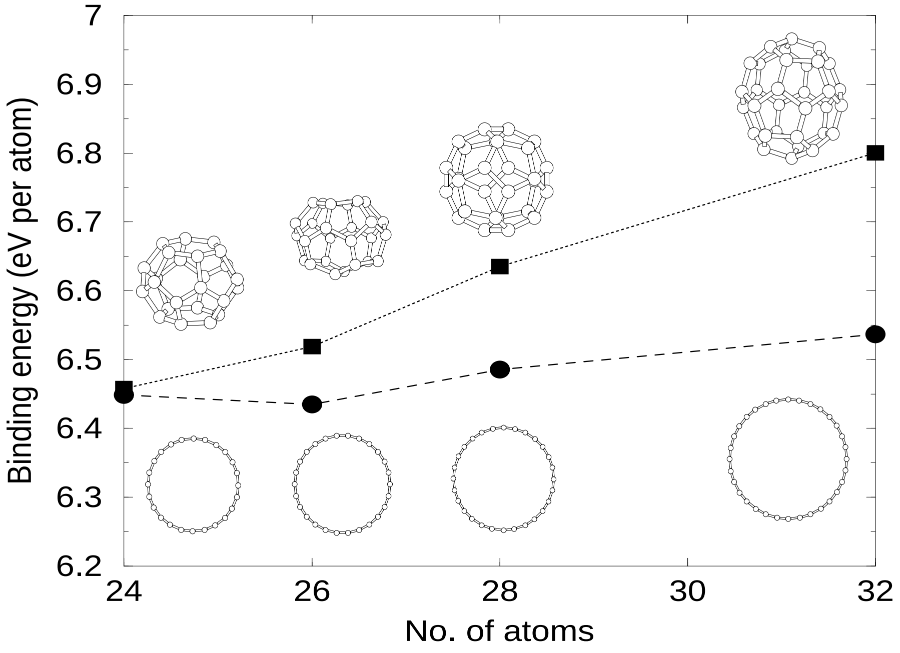

Figure 8.10: DMC binding energies of the ring (○) and fullerene (□) isomers in eV per atom. The lines are drawn for guidance only.

8.9 Discussion

The QMC and DFT results demonstrate the energetic importance of electron correlation. The very wide variation in energy amongst DFT methods gives a confusing and unclear picture of cluster stability. The QMC results, which we believe to be very accurate, indicate that the C_{26} and C_{28} fullerenes should be stable. This is close to the currently quoted range of experimentally determined fullerene stability, around C_{30} and higher.

Comparing the QMC results with those of the various DFT functionals, several important trends in the relative performance of the different functionals are highlighted. The overall quality of a functional for C_{24}, C_{26}, and C_{28} clusters is best judged by the agreement with the DMC data for the overall *shapes* of the relative energy data of figures 8.7

and 8.9. The best agreement is given by the PBE and B3LYP functionals, with the LDA being slightly inferior and the BLYP functional being worst. The tendency of the BLYP functional to favour structures with lower average coordination number and the tendency of the LDA to favour structures with higher average coordination numbers is consistent with the results on C_{20} reported by Grossman *et al.* [43]

The LDA, perhaps surprisingly, performs at least as well as the gradient corrected functionals. The LDA appears to favour systems with high average densities, therefore placing the rings anomalously high in energy. The gradient corrected functionals, which normally would be expected to improve the ordering do *not* consistently reduce the differences between the DFT and DMC results; for C_{24} the rms deviation for the PBE functional is larger than for the LDA and the overall spread in energies is increased from 3.04 to 3.41 eV. The BLYP functional performs poorly, consistently placing the structures of low average coordination substantially lower in energy relative to DMC and other DFT values. The B3LYP functional, which includes an exact exchange component, performs most consistently, and also gives the same increase in stability of C_{28} as DMC when spin-polarization is included. The PBE and B3LYP density functionals give the best description of the relative energies of the isomers, while the BLYP functional gives the poorest.

The DMC binding energies show a clear increase in stability with cluster size for both ring and fullerene isomers. The cumulenic C_{26} ring is found to be more weakly bound *per atom* than the adjacent polyacetylenic C_{24} and C_{28} rings, in agreement with early semi-empirical calculations of C_n rings. [165] However, the difference in binding energies of the fullerene and rings (either cumulenic or polyacetylenic) increases nearly linearly with cluster size.

The final test of the DMC predictions must lie with experiment. It is clear that the actual abundances of different clusters depend sensitively on the precise experimental conditions. The stability of clusters against fragmentation, growth and other chemical reactions is clearly a very complicated subject. One issue is that the clusters are formed at temperatures of order 10^3 K and therefore the vibrational contributions to the free energy may be significant. A relatively simple picture emerges from computations of vibrational properties [151, 152, 156]). Fullerenes are relatively rigid and have smaller vibrational free energies than rings, which have many low-lying vibrational modes. Consequently, the ring isomers are favoured relative to fullerenes at high temperatures. If thermodynamic stability alone were to determine which cluster sizes were observed then only the larger fullerenes would ever be observed, but in a recent experiment the abundance of the C_{32} fullerene was found to be *greater* than C_{60}. [143] There is more evidence that thermodynamic stability to rearrangements of clusters of a particular size are important in determining which isomers are observed. For example, in the experimental study of Ref. [143], fullerenes were mostly observed for clusters containing more than about 30 carbon atoms, while for smaller clusters mostly rings were formed. This behaviour closely matches the critical size for fullerene stability predicted by our DMC calculations.

There have been several proposals that *cluster solids* could be synthesized by surface deposition of fullerenes. [162, 163] Our DMC results support these proposals: the C_{28} fullerene is predicted to be stable, yet it is likely to readily polymerise due to its geometry and T_d spin-polarised ground-state. However, unless sufficiently high abundances of the fullerene are produced, its high reactivity could also *hinder* production of such solids, as the fullerene is likely to react with other clusters of different size or with any impurities present. The C_{26} fullerene, although predicted to be stable, appears less likely to polymerise due to its strained structure and less reactive ground state.

8.10 Conclusions

In conclusion, the prediction of the relative energies of carbon clusters is a severe test of calculational accuracy because of the widely different geometries and the presence of single, double, and triple bonds. Current density functionals give very similar structures but widely differing relative energies. DMC calculations show that for C_{24} the fullerene is not the lowest energy isomer. The lowest energy C_{24} structure is a sheet structure, which is expected to be the smallest stable graphitic fragment. The smallest energetically stable fullerenes are predicted to be the highly strained C_{26} cluster of C_{2v} symmetry and the spin polarized $^5A_2$ state of the C_{28} cluster of T_d symmetry.

This prediction lends support to proposals that a C_{28} solid could be synthesised by surface deposition of C_{28} fullerenes. [162, 163] Depending on its abundance, this route may be either hampered or facilitated by the chemical reactivity of the C_{28} fullerene.

Chapter 9

Conclusions

9.1 Summary

In this thesis, Quantum Monte Carlo (QMC) techniques have been developed and applied to real electronic systems with the aim of obtaining greater accuracy and insight than possible with other current electronic structure methods.

In chapters 3 and 4 the details and subtleties of implementing a viable and accurate QMC computational scheme were discussed. The importance of developing QMC techniques was stressed. Two developments were presented:

Firstly, the poor behaviour of many wavefunction optimisation schemes was investigated, and the increased stability of objective functions based on variance minimisation over energy minimisation was demonstrated numerically. The poor behaviour of the variance sometimes observed during optimisation was shown to be due to the presence of configurations of extreme "outlying" energies which result from commonly used trial wavefunctions. Limiting these outlying energies was shown to greatly improve the statistical properties of common objective functions.

Secondly, the compensation of finite size effects in periodic supercell calculations was demonstrated. The finite size effects were shown to result from a combination of two effects, and methods were presented which minimise both. A modified Coulomb interaction was introduced which minimise the finite size effects. Extensive calculations of ground and excited states demonstrate the efficacy of this approach.

Substantial parallel codes implementing the methods and techniques described have been written or rewritten.

Two large–scale applications of QMC methods are presented:

Firstly, the one-body density matrix of the valence electrons in bulk silicon is studied. A QMC method for the calculation of the density matrix from accurate many–body wavefunctions is presented. The matrix obtained by QMC methods is found to greatly resemble density function results. Natural orbitals, which diagonalise the density matrix, are compared with Hartree-Fock (HF) and local density approximation-density functional theory (LDA-DFT) orbitals. Only small differences in nodal quality are found between the different orbitals. A QMC formulation of the extended Koopmans' theorem is developed, making use of the density matrix. The band structure of silicon is obtained from a single variational Monte Carlo calculation, and the scaling of and approximations to the theory are considered.

Secondly, the stability of small carbon clusters is studied by diffusion Monte Carlo. The zero temperature relative energies of many energetically competitive structures are determined and the smallest stable fullerenes identified as the C_{2v} symmetry C_{26} and T_d symmetry C_{28} fullerenes. Comparisons with density functional results show that most current, popular, density functionals are of low predictive power for the systems studied. These results also demonstrate the applicability of diffusion Monte Carlo using the atom-centred trial wavefunctions described in chapter 4.

These successful results demonstrate the viability of a statistical approach to the many–body problem. Using current resources and wavefunctions of similar quality, i.e. without further technical or resource developments, accurate energy differences between systems of several hundred electrons can be determined. The current obtainable energy resolution, ~ 0.1 eV for several hundred electrons, is sufficient to determine the formation and relative energies of many solid-state and chemical systems and is smaller than the error of other applicable first-principles methods.

QMC calculations of silicon interstitials [166] have already been performed and calculations of the neutral vacancy in diamond are currently underway. [167] Possible future applications could include studies of interstitials and defects in other group IV materials and $III - V$ compounds. A challenging molecular application would be to accurately compute the energy barriers and heats of formation in a reaction with excited state intermediates.

9.2 Future developments

Further technical advances are necessary if QMC methods are to develop significantly beyond the determination of single–point energies.

The development of reliable methods for calculating forces is essential. As shown in chapter 8, there can be considerable energetic differences amongst DFT functionals, which leads to small differences in optimum geometries. Lingering doubts will always remain over geometries determined at a lower level of correlation treatment, even in small systems where quantum chemical techniques may be applied. Accurate methods for calculating forces would represent a considerable refinement in QMC methodology. Improvements in relative energies between structures of significantly different geometry and bonding would result, and it is in these systems where other methods are likely to suffer from the greatest errors. Methods based on VMC will require consistently accurate wavefunctions for different geometries to obtain reliable forces, which is a considerable technical challenge. For DMC based methods, the accuracy of the fixed–node approximation is likely to be critical.

Accurate and efficient methods of calculating excited state energies and properties is a large potential area of research. Although direct methods of calculating excited state energies exist, they are currently computationally very expensive, and satisfactory methods based on correlated sampling might be better.

All of the above developments and applications would benefit from improved wavefunctions and optimisation methods. Specifically, improvements in wavefunction quality are likely to substantially improve the achievable energy resolution, which is determined by wavefunction quality and computational resources. The optimisation of the orbitals in the Slater determinants appears most likely to deliver these improvements. Efficient and reliable methods of orbital optimisation for large systems will need to be developed. Transition metal systems are certain to require improvements over current methodology. Improved pseudopotentials will also need to be developed for these systems, [61] as the LDA and pseudopotentials derived from it are unreliable.

Quantum Monte Carlo methods are currently the only practical methods for large-scale benchmark accuracy electronic structure calculations. Their non-perturbative nature and real-space formulation is ideal for benchmark studies of electron correlation. The techniques are particularly suited to supercomputer architectures and their application will continue to grow as increased computer resources become available to the community. Although the methods will always be more expensive than DFT and HF-based calculations, the scaling of the methods places them substantially ahead of other correlated wavefunction methods for large systems. Therefore it is important that Quantum Monte Carlo techniques are considered in context, as one of a range of electronic structure methods of increasing accuracy, from tight-binding through density functional theory to Quantum Monte Carlo.

Bibliography

[1] Neil W. Ashcroft and N. David Mermin. *Solid State Physics* (Saunders College, Philadelphia, 1976), International edn.

[2] Neil S. Szabo and Attila Ostlund. *Modern Quantum Chemistry* (McGraw-Hill, New York, 1989).

[3] P. Fulde. *Electron Correlations in Molecules and Solids* (Springer, 1995).

[4] R. McWeeny. *Methods of Molecular Quantum Mechanics* (Academic Press, London, 1992).

[5] Gene Golub and Charles Van Loan. *Matrix Compuations* (The Johns Hopkins University Press, 1989).

[6] R. Barrett. *Templates for the Solution of Linear Systems:Building Blocks for Iterative Methods* (Society of Industrial and Applied Mathematics, 1994).

[7] J. Čí̧zek. *Adv. Chem. Phys.* **14**, 35 (1969).

[8] R. M. Dreizler and E. K. U. Gross. *Density Functional Theory* (Springer, 1990).

[9] Robert G. Parr and Yang Weitao. *Density-functional Theory of Atoms and Molecules* (Oxford University Press, 1994).

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

[11] P. Hohenberg and W. Kohn. *Phys. Rev. B* **76**, 6062 (1964).

[12] M. Levy. *Proc. Nat. Acad. Sci.* **76**, 6062 (1979).

[13] D. M. Ceperley and B. J. Alder. Ground State of the Electron Gas by a Stochastic Method. *Phys. Rev. Lett.* **45** (7), 566 (August 1980).

[14] R. Q. Hood, M. Y. Chou, A. J. Williamson, G. Rajagopal and R. J. Needs. Quantum Monte Carlo investigation of exchange and correlation in silicon. *Phys. Rev. Lett.* **78** (17), 3350 (April 1997).

[15] R. Q. Hood, M. Y. Chou, A. J. Williamson, G. Rajagopal and R. J. Needs. Exchange and correlation in silicon. *Phys. Rev. B* **57** (15), 8972 (April 1998).

[16] M. E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics* (Oxford University Press, 1999).

[17] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids* (Oxford University Press, 1987).

[18] W. H. Press, S. A. Teukolsky and W. T. Vetterling. *Numerical Recipes in Fortran 77: The Art of Scientific Computing* (Cambridge University Press, 1992).

[19] John A. Rice. *Mathematical Statistics and Data Analysis* (Duxbury Press, 1995), 2nd edn.

[20] N. Metropolis, A. W. Rosenbluth, N. M. Rosenbluth, A. H. Teller and E. Teller. *J. Chem. Phys.* **21**, 1087 (1953).

[21] W. L. McMillan. *Phys. Rev.* **138**, A442 (1965).

[22] R. Jastrow. *Phys. Rev.* **98**, 1479 (1955).

[23] B. L. Hammond, W. A. Lester, Jr. and P. J. Reynolds. *Monte Carlo Methods in Ab Initio Quantum Chemistry* (World Scientific, 1994).

[24] L. Mitáš. Quantum Monte Carlo. *Current Opinion in Solid State and Materials Science* **2**, 696 (1997).

[25] T. Kato. *Comm. Pure Appl. Math.* **10**, 151 (1957).

[26] S. Fahy, X. W. Wang and S. G. Louie. Variational Quantum Monte-Carlo nonlocal pseudopotential approach to solids - formulation and application to diamond, graphite, and silicon. *Phys. Rev. B* **42** (6), 3503 (August 1990).

[27] M. H. Kalos, D. Levesque and L. Verlet. *Phys. Rev. A* **9**, 2178 (1974).

[28] P. J. Reynolds, D. M. Ceperley, B. J. Alder and W. A. Lester, Jr. Fixed-node quantum Monte-Carlo for molecules. *J. Chem. Phys.* **77** (11), 5593 (December 1982).

[29] M. P. Nightingale C. J. Umrigar and K. J. Runge. A diffusion Monte Carlo algorithm with very small time-step errors. *J. Chem. Phys.* **99** (4), 2865 (August 1993).

[30] James B. Anderson. Quantum chemistry by random walk. *J. Chem. Phys.* **65** (10), 4121 (November 1976).

[31] D. M. Ceperley and B. J. Alder. Quantum Monte Carlo for molecules: Green's function and nodal release. *J. Chem. Phys.* **81** (12), 5833 (December 1984).

[32] Yongkyung Kwon, D. M. Ceperley and R. M. Martin. Effects of three-body and backflow correlations in the two-dimensional electron gas. *Phys. Rev. B* **48** (16), 12037 (October 1993).

[33] Yongkyung Kwon, D. M. Ceperley and R. M. Martin. Effects of backflow correlation in the three-dimensional electron gas: Quantum Monte Carlo study. *Phys. Rev. B* **58** (11), 6800 (September 1998).

[34] R. P. Feynman and M. Cohen. *Phys. Rev.* **102**, 1189 (1956).

[35] D. M. Ceperley. The statistical error of greens-function Monte-Carlo. *J. Stat. Phys* **43** (5-6), 815 (1986).

[36] M. Krauss and W. J. Stevens. Effective potentials in molecular quantum-chemistry. *Ann. Rev. Phys. Chem.* **35**, 357 (1984).

[37] C. J. Umrigar, K. G. Wilson and J. W. Wilkins. Optimized Trial Wave-functions for Quantum Monte-Carlo Calculations. *Phys. Rev. Lett.* **60** (17), 1719 (1988).

[38] S. Fahy, X. W. Wang and S. G. Louie. Variational Quantum Monte-Carlo nonlocal pseudopotential approach to solids - cohesive and structural-properties of diamond. *Phys. Rev. Lett.* **61** (14), 1631 (October 1988).

[39] David B. Cook. *Handbook of Computational Quantum Chemistry* (Oxford University Press, 1998).

[40] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias and J. D. Joannopoulos. Iterative minimization techniques for *ab initio* total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.* **64** (4), 1045 (October 1992).

[41] S. A. Alexander and R. L. Coldwell. Atomic wave function forms. *Int. J. Quantum Chem.* **63** (5), 1001 (1997).

[42] L. Mitáš and R. M. Martin. Quantum Monte-Carlo of Nitrogen - Atom, dimer, atomic, and molecular-solids. *Phys. Rev. Lett.* **72** (15), 2438 (1994).

[43] J. C. Grossman and L. Mitáš. Structure and stability of molecular carbon - Importance of electron correlation. *Phys. Rev. Lett.* **75** (21), 3870 (November 1995). Erratum: Phys. Rev. Lett. **76**, 1006 (1995).

[44] J. C. Grossman and L. Mitáš. High accuracy molecular heats of formation and reaction barriers: Essential role of electron correlation. *Phys. Rev. Lett.* **79** (22), 4353 (December 1997).

[45] J. C. Grossman and L. Mitáš. Quantum Monte-Carlo determination of electronic and structural-properties of Si_n clusters ($n \leq 20$). *Phys. Rev. Lett.* **74** (8), 1323 (February 1995).

[46] A. J. Williamson, S. D. Kenny, G. Rajagopal, A. J. James, R. J. Needs, L. M. Fraser, W. M. C. Foulkes and P. Maccallum. Optimized wave functions for Quantum Monte Carlo studies of atoms and solids. *Phys. Rev. B* **53** (15), 9640 (April 1996).

[47] A. J. Williamson. *Quantum Monte Carlo Calculations of Electronic Excitations*. Ph.D. thesis, University of Cambridge (1996).

[48] K. C. Huang. *Ab Initio Determinations of Energetics and Forces in Molecules*. Master's thesis, University of Cambridge (1999).

[49] P. R. C. Kent, K. C. Huang and R. J. Needs. Unpublished.

[50] D. Ceperley, G. V. Chester and M. Kalos. Monte Carlo simulation of a many-fermion study. *Phys. Rev. B* **16** (7), 3081 (October 1977).

[51] P. P. Ewald. *Ann. Phys. (Leipzig)* **64**, 253 (1921).

[52] G. Rajagopal and R. J. Needs. An optimised Ewald method for long-ranged potentials. *J. Comput. Phys.* **115** (2), 399 (December 1994).

[53] C. Filippi and C. J. Umrigar. Multiconfiguration wave functions for quantum Monte Carlo calculations of first-row diatomic molecules. *J. Chem. Phys.* **105** (1), 213 (July 1996).

[54] Walter J. Stevens, Harold Basch and Morris Krauss. Compact effective potentials and efficient shared-exponent basis sets for the first- and second-row atoms. *J. Chem. Phys.* **81** (12), 6026 (December 1984).

[55] D. R. Hamann, M. Schlüter and C. Chiang. Norm-Conserving Pseudopotentials. *Phys. Rev. Lett.* **43** (20), 1494 (November 1979).

[56] G. B. Bachelet, D. R. Hamann and M. Schlüter. Pseudopotentials that work: From H to Pu. *Phys. Rev. B* **26** (8), 4199 (October 1982).

[57] G. P. Kerker. Non-singular atomic pseudopotentials for solid state applications. *J. Phys. C* **13** (9), 189 (March 1980).

[58] N. Troullier and José Luís Martins. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B* **43** (3), 1993 (January 1991).

[59] David Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B* **41** (11), 7892 (April 1990).

[60] E. L. Shirley and R. M. Martin. *GW* quasi-particle calculations in atoms. *Phys. Rev. B* **47** (23), 15404 (June 1993).

[61] Y. Lee, P. R. C. Kent, M. D. Towler and R. J. Needs. Pseudopotentials for Transition Metals. Unpublished.

[62] A. D. McLaren. *J. Math. Comp.* **17**, 361 (1963).

[63] L. Mitáš, E. L. Shirley and D. M. Ceperley. Nonlocal pseudopotentials and diffusion Monte Carlo. *J. Chem. Phys.* **95** (5), 3467 (September 1991).

[64] Andrea Bosin, Vincenzo Fiorentini, Andrea Lastri and Giovanni B. Bachelet. Local norm-conserving pseudo-Hamiltonians. *Phys. Rev. A* **52** (1), 236 (July 1995).

[65] B. L. Hammond, P. J. Reynolds and W. A. Lester, Jr. Valence quantum Monte Carlo with ab initio effective core potentials. *J. Chem. Phys.* **87** (2), 1130 (July 1987).

[66] M. M. Hurley and P. A. Christiansen. Relativistic effective potentials in quantum Monte-Carlo calculations. *J. Chem. Phys.* **86** (2), 1069 (January 1987).

[67] P. H. Acioli and D. M. Ceperley. Generation of pseudopotentials from correlated wave-functions. *J. Chem. Phys.* **100** (11), 8169 (June 1994).

[68] E. L. Shirley and R. M. Martin. Many-body core-valence partitioning. *Phys. Rev. B* **47** (23), 15413 (June 1993).

[69] P. A. Christiansen. Effective potentials and multiconfiguration wave-functions in quantum Monte-Carlo calculations. *J. Chem. Phys.* **88** (8), 4867 (April 1988).

[70] M. Dolg, U. Wedig, H. Stoll and H. Preuss. Energy-adjusted abintio pseudopotentials from the 1st-row transition-elements. *J. Chem. Phys.* **86** (2), 866 (January 1987).

[71] E. L. Shirley, Xuejun Zhu and S. G. Louie. Core polarization in solids: Formulation and application to semiconductors. *Phys. Rev. B* **56** (11), 6648 (September 1997).

[72] Jack Sugar and Charles Corliss. Atomic Energy Levels of the Iron-Period Elements: Potassium through Nickel. *Journal of Physical and Chemical Reference Data* **14** (2), 147 (1985).

[73] N. K. Wannier. *Rev. Mod. Phys.* **64**, 1045 (1992).

[74] J. M. Ziman. *Principles of the Theory of Solids* (Cambridge University Press, Cambridge, 1972), 2nd edn.

[75] M. C. Rapaport. *The Art of Molecular Dynamics Simulation* (Cambridge University Press, 1995).

[76] S. W. de Leeuw, J. W. Perram and E. R. Smith. *Proc. Roy. Soc. Lond A* **27**, 373 (1980).

[77] J. W. Perram, H. G. Peterson and S. W. de Leeuw. An algorithm for the simulation of condensed matter which grows as the 3/2 power of the number of particles. *Mol. Phys.* **65** (4), 875 (1988).

[78] Christopher A. White and M. J. Head-Gordon. Derivation and efficient implementation of the fast multipole method. *J. Chem. Phys.* **101** (8), 6593 (October 1994).

[79] L. Greengard and V. Rokhlin. On the evaluation of electrostatic interactions in molecular modeling. *Chem. Scripta* **29A**, 139 (1989).

[80] H. Conroy. *J. Chem. Phys.* **41**, 1331 (1964).

[81] K. E. Schmidt and J. W. Moskowitz. Correlated Monte Carlo wave functions for the atoms He through Ne. *J. Chem. Phys.* **93** (6), 4172 (September 1990).

[82] R. N. Barnett, Z. Sun and W. A. Lester, Jr. Fixed-sample optimization in quantum Monte Carlo using a probability density function. *Chem. Phys. Lett.* **273** (5-6), 321 (July 1997).

[83] P. R. C. Kent, R. Q. Hood, A. J. Williamson, R. J. Needs W. M. C. Foulkes and G. Rajagopal. Finite-size errors in quantum many-body simulations of extended systems. *Phys. Rev. B* **59** (3), 1917 (January 1999).

[84] S. Fahy. Unpublished.

[85] M. L. Cohen and T. K. Bergstresser. *Phys. Rev.* **141**, 789 (1966).

[86] L. M. Fraser, W. M. C. Foulkes, G. Rajagopal, R. J. Needs, S. D. Kenny and A. J. Williamson. Finite-size effects and Coulomb interactions in Quantum Monte-Carlo calculations for homogeneous systems with periodic boundary conditions. *Phys. Rev. B* **53** (4), 1814 (January 1996).

[87] A. J. Williamson, G. Rajagopal, R. J. Needs, L. M. Fraser, W. M. C. Foulkes, Y. Wang and M.Y.Chou. Elimination of Coulomb finite-size effects in quantum many-body simulations. *Phys. Rev. B* **55** (8), 4851 (February 1997).

[88] J. P. Perdew and A. Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23** (10), 5048 (May 1981).

[89] D. Ceperley. Ground state of the fermion one-component plasma: A Monte Carlo study in two and three dimensions. *Phys. Rev. B* **18** (7), 3126 (October 1978).

[90] D. M. Ceperley and B. J. Alder. Ground state of solid hydrogen at high pressures. *Phys. Rev. B* **36** (4), 2092 (August 1987).

[91] B. Tanatar and D. M. Ceperley. Ground state of the two-dimensional electron gas. *Phys. Rev. B* **39** (8), 5005 (March 1989).

[92] G. E. Engel, Y. Kwon and R. M. Martin. Quasi-particle bands in a 2-dimensional crystal found by *GW* and quantum Monte-Carlo calculations. *Phys. Rev. B* **51** (19), 13538 (May 1995).

[93] G. Rajagopal, R. J. Needs, S. D. Kenny, W. M. C. Foulkes and A. J. James. Quantum Monte-Carlo Calculations for solids using special k points methods. *Phys. Rev. Lett.* **73** (14), 1959 (October 1994).

[94] G. Rajagopal, R. J. Needs, A. James, S. D. Kenny and W. M. C. Foulkes. Variational and diffusion Quantum Monte Carlo calculations at nonzero wave vectors: Theory and application to diamond-structure germanium. *Phys. Rev. B* **51** (16), 10591 (April 1995).

[95] A. Baldereschi. *Phys. Rev. B* **7**, 5212 (1973).

[96] H. J. Monkhorst and J. D. Pack. Special points for Brillouin-zon integrations. *Phys. Rev. B* **13** (12), 5188 (June 1976).

[97] X. P. Li, D. M. Ceperley and R. M. Martin. ohesive energy of silicon by the Green's-function Monte Carlo method. *Phys. Rev. B* **44** (19), 10929 (November 1991).

[98] O. W. Day, D. W. Smith and C. Garrod. *Int. J. Quantum Chem. Symp.* **8**, 501 (1974).

[99] M. M. Morrell, R. G. Parr and M. Levy. Calculation of ionization potentials from density matrices and natural functions, and the long-range behavior of natural orbitals and electron density. *J. Chem. Phys.* **62** (2), 549 (January 1975).

[100] P. R. C. Kent, R. Q. Hood, M. D. Towler, R. J. Needs W. M. C. Foulkes and G. Rajagopal. Quantum Monte Carlo calculations of the one-body density matrix and excitation energies of silicon. *Phys. Rev. B* **57** (24), 15293 (June 1998).

[101] L. Mitáš. Electronic structure by quantum Monte Carlo: Atoms, molecules and solids. *Comput. Phys. Commun.* **96** (2-3), 107 (1996).

[102] L. Mitáš. *Electronic Properties of Solids Using Cluster Methods*, p. 151 (Plenum, New York, 1995).

[103] A. J. Williamson, R. Q. Hood, R. J. Needs and G. Rajagopal. Diffusion quantum Monte Carlo calculations of the excited states of silicon. *Phys. Rev. B* **57** (19), 12140 (May 1998).

[104] W. Knorr and R. W. Godby. Investigating exact density-functional theory of a model semiconductor. *Phys. Rev. Lett.* **68** (5), 639 (February 1992).

[105] W. Borrmann and P. Fulde. Exchange and correlation effects on the quasiparticle band structure of semiconductors. *Phys. Rev. B* **35** (18), 9569 (June 1987).

[106] S. Öğüt, J. R. Chelikowsky and S. G. Louie. Quantum Confinement and Optical Gaps in Si Nanocrystals. *Phys. Rev. Lett.* **79** (9), 1770 (September 1997).

[107] A. Franceschetti, L. W. Wang and A. Zunger. Comment on "Quantum Confinement and Optical Gaps in Si Nanocrystals". *Phys. Rev. Lett.* **83** (6), 1269 (September 1999).

[108] K.-H. Hellwege, O. Madelung and Landolt-Börnstein, eds. *Numerical Data and Functional Relationships in Science and Technology* (Springer, Berlin, 1982).

[109] P. Löwdin. *Phys. Rev.* **97**, 1474 (1955).

[110] P. Löwdin. *Phys. Rev.* **97**, 1490 (1955).

[111] P. Löwdin. *Phys. Rev.* **97**, 1509 (1955).

[112] T. Koopmans. *Physica* **1**, 104 (1924).

[113] C. W. Greef and W. A. Lester, Jr. Quantum Monte Carlo binding energies for silicon hydrides. *J. Chem. Phys.* **106** (15), 6412 (April 1997).

[114] A. Lüchow and J. Anderson. First-row hydrides: Dissociation and ground state energies using quantum Monte Carlo. *J. Chem. Phys.* **105** (17), 7573 (November 1996).

[115] H. Eckstein, W. Schattke, M. Reigrotzki and R. Redmer. Variational quantum Monte Carlo ground state of GaAs. *Phys. Rev. B* **54** (8), 5512 (August 1996).

[116] Chien-Jung Huang, C. J. Umrigar and M. P. Nightingale. Accuracy of electronic wave functions in quantum Monte Carlo: The effect of high-order correlations. *J. Chem. Phys.* **107** (8), 3007 (1997).

[117] P. O. Löwdin. *J. Chem. Phys.* **61**, 55 (1957).

[118] W. Meyer. PNO-CI studies of electron correlation effects. I. Configuration expansion by means of nonorthogonal orbitals, and application to the ground state and ionized states of methane. *J. Chem. Phys.* **58** (3), 1017 (1973).

[119] M. Caffarel and D. M. Ceperley. A Bayesian analysis of Green's function Monte Carlo correlation functions. *J. Chem. Phys.* **97** (11), 8415 (December 1992).

[120] D. M. Ceperley and B. Bernu. The calculation of excited state properties with quantum Monte Carlo. *J. Chem. Phys.* **89** (10), 6316 (November 1988).

[121] S. Fahy, X. W. Wang and S. G. Louie. Pair-correlation function and single-particle occupation numbers in diamond and silicon. *Phys. Rev. Lett.* **65** (12), 1478 (September 1990).

[122] S. Tanaka. Variational Quantum Monte-Carlo approach to the electronic- structure of NiO. *J. Phys. Soc. Japan* **64** (11), 4270 (1995).

[123] J.F. Cornwell. *Group Theory in Physics Vol. 1* (Academic Press (London), 1984).

[124] B. Williams. *Compton Scattering* (McGraw-Hill, New York, 1976).

[125] E. Eisenberger and P. M. Platzman. Compton scattering of X rays from bound electrons. *Phys. Rev. A* **2** (2), 415 (August 1970).

[126] Balázs Králik, Paul Delaney and Steven G. Louie. Correlation Effects in the Compton Profile of Silicon. *Phys. Rev. Lett.* **80** (19), 4253 (May 1998).

[127] Claudia Filippi and David M. Ceperley. Quantum Monte Carlo calculation of Compton profiles of solid lithium. *Phys. Rev. B* **59** (12), 7907 (March 1999).

[128] L. Lam and P. M. Platzman. Momentum density and Compton profile of the inhomogeneous interacting electronic system. I. Formalism. *Phys. Rev. B* **9** (12), 5122 (June 1974).

[129] Paul Delaney, Balázs Králik and Steven G. Louie. Compton profiles of Si: Pseudopotential calculation and reconstruction effects. *Phys. Rev. B* **58** (8), 4320 (August 1998).

[130] C. Pisani, V. R. Saunders, C. Roetti, M. Causà, N. M. Harrison, R. Orlando and E. Aprà. *CRYSTAL 95 User Manual* (1996).

[131] R. P. Feynman. *Phys. Rev.* **94**, 262 (1954).

[132] R. C. Morrison. The extended Koopmans' theorem and its exactness. *J. Chem. Phys.* **96** (5), 3718 (March 1992).

[133] O. W. Day and D. W. Smith. Extension of Koopmans' theorem. I. Derivation. *J. Chem. Phys.* **62** (1), 113 (January 1975).

[134] L. Adamowicz, J. C. Ellenbogen and E. A. McCullough. Extended-Koopmans-theorem approach to abinitio calculations upon the ground-state and 1st excited-state of the LiH anion. *Int. J. Quantum Chem.* **30** (5), 617 (November 1986).

[135] J. R. Chelikowsky and M. L. Cohen. Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors. *Phys. Rev. B* **14** (2), 55 (July 1976).

[136] M. Rohlfing, P. Krüger and J. Pollmann. Quasiparticle band-structure calculations for C, Si, Ge, GaAs, and SiC using Gaussian-orbital basis sets. *Phys. Rev. B* **48** (24), 17791 (December 1993).

[137] W. Weltner, Jr. and R. J. Van Zee. Carbon Molecules, Ions and Clusters. *Chem. Rev.* **89**, 1713 (1989).

[138] H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl and R. E. Smalley. C_{60} - Buckminsterfullerene. *Nature (London)* **318** (6042), 162 (1985).

[139] Alan Van Orden and Richard J. Saykally. Small Carbon Clusters: Spectroscopy, Structure, and Energetics. *Chem. Rev.* **98** (6), 2313 (1998).

[140] P. W. Fowler and D. E. Manolopoulos. *An Atlas of Fullerenes* (Clarendon Press, Oxford, 1995).

[141] H. Handschuh, G. Gantefor, B. Kessler, P. S. Bechthold and W. Eberhardt. Stable Configurations of Carbon Clusters - Chains, Rings, and Fullerenes. *Phys. Rev. Lett.* **74** (7), 1095 (February 1995).

[142] G.von Helden *et al.* Do small fullerenes exist only on the computer - experimental results on C_{20}(+/-) and C_{24}(+/-). *Chem. Phys. Lett.* **204** (1-2), 15 (March 1993).

[143] H. Kietzmann, R. Rochow, G. Gantefor, W. Eberhardt, K. Vietze, G. Seifert and P. W. Fowler. Electronics structure of small fullerenes: Evidence of high stability of C_{32}. *Phys. Rev. Lett.* **81** (24), 5378 (December 1998).

[144] Jerzy Cioslowski. *Electronic Structure Calculations on Fullerenes and Their Derivatives* (Oxford University Press, 1995).

[145] Gustavo E. Scuseria. Ab initio calculations of fullerenes. *Science* **271** (5251), 942 (February 1996).

[146] Martin Feyereisen, Maciej Gutowski and Jack Simons. Relative stabilities of fullerene, cumulene, and polyacetylene structures for C_n: $n = 18 - 60$. *J. Chem. Phys.* **96** (4), 2926 (February 1992).

[147] Y. Shlakhter, Svetlana Sokolova, Arne Lüchow and ames B. Anderson. Energetics of carbon clusters C_8 and C_{10} from all-electron Quantum Monte Carlo calculations. *J. Chem. Phys.* **110** (22), 10725 (June 1999).

[148] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic-behavior. *Phys. Rev. A* **38** (6), 3098 (1988).

[149] C. Lee, W. Yang and R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37** (2), 785 (January 1988).

[150] Frank Jensen and Henrik Koch. C_{24}: Ring or fullerene? *J. Chem. Phys.* **108** (8), 3213 (February 1988).

[151] Krishnan Raghavachari, B. L. Zhang, J. A. Pople, B. G. Johnson and P. M. W. Gill. Isomers of C_{24} - density-functional studies including gradient corrections. *Chem. Phys. Lett.* **220** (6), 385 (April 1994).

[152] Jan M. L. Martin, J. ElYazal and J. P. Francois. On the structure and vibrational frequencies of C_{24}. *Chem. Phys. Lett.* **255** (1-3), 7 (June 1996).

[153] H. W. Kroto. The stability of the fullerenes C_{24}, C_{28}, C_{32}, C_{36}, C_{50}, C_{60}, and C_{70}. *Nature (London)* **329** (6139), 529 (October 1987).

[154] A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98** (7), 5648 (April 1993).

[155] R. O. Jones. Density functional study of carbon clusters C_{2n}, $(2 \leq n \leq 16)$. I. Structure and bonding in the neutral clusters. *J. Chem. Phys.* **110** (11), 5189 (March 1999).

[156] Jan M. L. Martin. C_{28}: The smallest stable fullerene? *Chem. Phys. Lett.* **255** (1-3), 1 (June 1996).

[157] M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez and J. A. Pople. *Gaussian 94*. Gaussian Inc., Pittsburgh, PA (1995).

[158] T. H. Dunning, Jr. Gaussian-basis sets for use in correlated molecular calculations .I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90** (2), 1007 (1989).

[159] Jan M. L. Martin. Basis set convergence and performance of density functional theory including exact exchange contributions for geometries and harmonic frequencies. *Mol. Phys.* **86** (6), 1437 (1995).

[160] J. P. Perdew, Kieron Burke and Matthias Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77** (18), 3865 (October 1996).

[161] L. Mitáš. Unpublished.

[162] Ting Guo, M. D. Diener, Y. Chai, M. J. Alford, R. E. Haufler, S. M. McClure, T. Ohno, J. H. Weaver, G. E. Scuseria and R. E. Smalley. Uranium stabilization of C_{28} - a tetravalent fullerene. *Science* **257** (5077), 1661 (September 1992).

[163] Jeongnim Kim, Giulia Galli, John W. Wilkins and Andrew Canning. Disordered and ordered C_{28} solids. *J. Chem. Phys.* **108** (6), 2631 (February 1998).

[164] C. Piscotti, T. Yarger and A. Zettl. C_{36}, a new carbon solid. *Nature (London)* **393**, 771 (June 1998).

[165] K. S. Pitzer and E. Clementi. *J. Am. Chem. Soc.* **81**, 4477 (1959).

[166] W. K. Leung, R. J. Needs, G. Rajagopal, S. Itoh and S. Ihara. Calculations of Silicon Self–Interstitial Defects. *Phys. Rev. Lett.* (1999). Accepted for publication.

[167] R. Q. Hood, P. R. C. Kent, R. J. Needs, G. Rajagopal and P. R. Briddon. Calculations of the neutral vacancy in diamond. Unpublished.