Continuum variational and diffusion quantum Monte Carlo calculations

R J Needs, M D Towler, N D Drummond and P López Ríos

Theory of Condensed Matter Group, Cavendish Laboratory, Cambridge CB3 0HE, UK

Abstract.   This topical review describes the methodology of continuum variational and diffusion quantum Monte Carlo calculations. These stochastic methods are based on many-body wave functions and are capable of achieving very high accuracy. The algorithms are intrinsically parallel and well-suited to petascale computers, and the computational cost scales as a polynomial of the number of particles. A guide to the systems and topics which have been investigated using these methods is given. The bulk of the article is devoted to an overview of the basic quantum Monte Carlo methods, the forms and optimisation of wave functions, performing calculations within periodic boundary conditions, using pseudopotentials, excited-state calculations, sources of calculational inaccuracy, and calculating energy differences and forces.

1. Introduction

The variational Monte Carlo (VMC) and diffusion Monte Carlo (DMC) methods are stochastic approaches for evaluating quantum mechanical expectation values with many-body Hamiltonians and wave functions [1]. VMC and DMC methods are used for both continuum and lattice systems, but here we describe their application only to continuum systems. The main attraction of these methods is that the computational cost scales as some reasonable power (normally from the second to fourth power) of the number of particles [2]. This scaling makes it possible to deal with hundreds or even thousands of particles, allowing applications to condensed matter.

Continuum quantum Monte Carlo (QMC) methods, such as VMC and DMC, occupy a special place in the hierarchy of computational approaches for modelling materials. QMC computations are expensive, which limits their applicability at present, but they are the most accurate methods known for computing the energies of large assemblies of interacting quantum particles. There are many problems for which the high accuracy achievable with QMC is necessary to give a faithful description of the underlying science. Most of our work is concerned with correlated electron systems, but these methods can be applied to any combination of fermion and boson particles with any inter-particle potentials and external fields *etc.* Being based on many-body wave functions, these are zero-temperature methods, and for finite temperatures one must use other approaches such as those based on density matrices.

Both the VMC and DMC methods are variational, so that the calculated energy is above the true ground state energy. The computational costs of VMC and DMC calculations scale similarly with the number of particles studied, but the prefactor is larger for the more accurate DMC method. QMC algorithms are intrinsically parallel and are ideal candidates for taking advantage of the petascale computers (10^{15} flops) which are becoming available now and the exascale computers (10^{18} flops) which will be available one day.

DMC has been applied to a wide variety of continuum systems. A partial list of topics investigated within DMC and some references to milestone papers are given below.

- Three-dimensional electron gas [3, 4, 5].
- Two-dimensional electron gas [6, 7, 8].
- The equation of state and other properties of liquid ^{3}He [9, 10].
- Structure of nuclei [11].
- Pairing in ultra-cold atomic gases [12, 13, 14].
- Reconstruction of a crystalline surface [15] and molecules on surfaces [16, 17].
- Quantum dots [18].
- Band structures of insulators [19, 20, 21].
- Transition metal oxide chemistry [22, 23, 24].

- Optical band gaps of nanocrystals [25, 26].
- Defects in semiconductors [27, 28, 29].
- Solid state structural phase transitions [30].
- Equations of state of solids [31, 32, 33, 34].
- Binding of molecules and their excitation energies [35, 36, 37, 38, 39].
- Studies of exchange-correlation [40, 41, 42, 43].

The same basic QMC algorithm can be used for each of the applications mentioned above with only minor modifications. The complexity and sophistication of the computer codes arises not from the algorithm itself, which is in fact quite simple, but from the diversity of the Hamiltonians and many-body wave functions which are involved. A number of computer codes are currently available for performing continuum QMC calculations of the type described here [44]. We have developed the CASINO code [45], which can deal with systems of different dimensionalities, various interactions including the Coulomb potential, external fields, mixtures of particles of different types and different types of many-body wave function.

The VMC and DMC methods are described in section 2, and the types of many-body wave function we use are described in 3. The optimisation of parameters in wave functions using stochastic methods, which are unique to the field, is described in section 4. QMC calculations within periodic boundary conditions are described in section 5, the use of pseudopotentials in QMC calculations is discussed in section 6, and excited-state DMC calculations are briefly described in section 7. Sources of bias in the DMC method and practical methods for handling errors in QMC results are described in section 8. In section 9 we describe how to evaluate other expectation values apart from the energy. Section 10 deals with the calculation of energy differences and energy derivatives in the VMC and DMC methods, and we make our final remarks in section 11.

2. Quantum Monte Carlo methods

The VMC method is conceptually very simple. The energy is calculated as the expectation value of the Hamiltonian with an approximate many-body trial wave function. In the more sophisticated DMC method the estimate of the ground state energy is improved by performing a process described by the evolution of the wave function in imaginary time. Throughout this article we will consider only systems with spin-independent Hamiltonians and collinear spins. We will also restrict the discussion to systems with time-reversal symmetry, for which the wave function may be chosen to be real. It is, however, straightforward to generalise the VMC algorithm to work with complex wave functions, and only a little more complicated to generalise the DMC algorithm to work with them [46].

2.1. The VMC method

The variational theorem of quantum mechanics states that, for a proper trial wave function Ψ_T, the variational energy,

$$E_V = \frac{\int \Psi_T(\mathbf{R}) \hat{H} \Psi_T(\mathbf{R}) \, d\mathbf{R}}{\int \Psi_T^2(\mathbf{R}) \, d\mathbf{R}} \, , \tag{1}$$

is an upper bound on the exact ground state energy E_0, *i.e.*, $E_V \geq E_0$. In equation (1), $\hat{H}$ is the many-body Hamiltonian and $\mathbf{R}$ denotes a $3N$-dimensional vector of particle coordinates. As discussed in section 3.1, the spin variables in equation (1) are implicitly summed over.

To facilitate the stochastic evaluation, E_V is written as

$$E_V = \int p(\mathbf{R}) E_L(\mathbf{R}) \, d\mathbf{R} \, , \tag{2}$$

where the probability distribution p is

$$p(\mathbf{R}) = \frac{\Psi_T^2(\mathbf{R})}{\int \Psi_T^2(\mathbf{R}') \, d\mathbf{R}'} \, , \tag{3}$$

and the local energy,

$$E_L(\mathbf{R}) = \Psi_T^{-1} \hat{H} \Psi_T \, , \tag{4}$$

is straightforward to evaluate at any $\mathbf{R}$.

In VMC the Metropolis algorithm [47] is used to sample the probability distribution $p(\mathbf{R})$. Let the electron configuration at a particular step be $\mathbf{R}'$. A new configuration $\mathbf{R}$ is drawn from the probability density $T(\mathbf{R} \leftarrow \mathbf{R}')$, and the move is accepted with probability

$$A(\mathbf{R} \leftarrow \mathbf{R}') = \min \left\{ 1, \frac{T(\mathbf{R}' \leftarrow \mathbf{R}) \Psi_T^2(\mathbf{R})}{T(\mathbf{R} \leftarrow \mathbf{R}') \Psi_T^2(\mathbf{R}')} \right\} . \tag{5}$$

It can easily be verified that this algorithm satisfies the *detailed balance* condition

$$\Psi_T^2(\mathbf{R}) T(\mathbf{R}' \leftarrow \mathbf{R}) A(\mathbf{R}' \leftarrow \mathbf{R}) = \Psi_T^2(\mathbf{R}') T(\mathbf{R} \leftarrow \mathbf{R}') A(\mathbf{R} \leftarrow \mathbf{R}'). \tag{6}$$

Hence $p(\mathbf{R})$ is the equilibrium configuration distribution of this Markov process and, so long as the transition probability is ergodic (i.e., it is possible to reach any point in configuration space in a finite number of moves), it can be shown that the process will converge to this equilibrium distribution. Once equilibrium has been reached, the configurations are distributed as $p(\mathbf{R})$, but successive configurations along the random walk are in general correlated.

The variational energy is estimated as

$$E_V \simeq \frac{1}{M} \sum_{i=1}^{M} E_L(\mathbf{R}_i), \tag{7}$$

where M configurations $\mathbf{R}_i$ have been generated after equilibration. The serial correlation of the configurations and therefore local energies $E_L(\mathbf{R}_i)$ complicates the calculation of the statistical error on the energy estimate: see section 8.2. Other expectation values may be evaluated in a similar manner to the energy.

Equation (2) is an importance sampling transformation of equation (1). Equation (2) exhibits the zero variance property: as the trial wave function approaches an exact eigenfunction ($\Psi_\mathrm{T} \to \phi_i$), the local energy approaches the corresponding eigenenergy, E_i, everywhere in configuration space. As Ψ_T is improved, E_L becomes a smoother function of $\mathbf{R}$ and the number of sampling points, M, required to achieve an accurate estimate of E_V is reduced.

VMC is a simple and elegant method. There are no restrictions on the form of trial wave function which can be used and it does not suffer from a fermion sign problem. However, even if the underlying physics is well understood it is often difficult to prepare trial wave functions of equivalent accuracy for two different systems, and therefore the VMC estimate of the energy difference between them will be biased. We use the VMC method mostly to optimise parameters in trial wave functions (see section 4) and our main calculations are performed with the more sophisticated DMC method, which is described in the next section.

2.2. The DMC method

In DMC the operator $\exp(-t\hat{H})$ is used to project out the ground state from an initial state. This can be viewed as solving the imaginary-time Schrödinger equation, which for electrons is

$$-\frac{\partial}{\partial t}\Phi(\mathbf{R}, t) = \left(\hat{H} - E_\mathrm{T}\right)\Phi(\mathbf{R}, t) = \left(-\frac{1}{2}\nabla_\mathbf{R}^2 + V(\mathbf{R}) - E_\mathrm{T}\right)\Phi(\mathbf{R}, t) , \quad (8)$$

where t is a real variable measuring the progress in imaginary time, V is the potential energy (assumed to be local for the time being), and E_T is an arbitrary energy offset known as the reference energy. Throughout this article we use Hartree atomic units where $m_e = \hbar = |e| = 4\pi\epsilon_0 = 1$, where m_e is the mass of the electron and e is its charge. Equation (8) can be solved formally by expanding $\Phi(\mathbf{R}, t)$ in the eigenstates ϕ_i of the Hamiltonian, which leads to

$$\Phi(\mathbf{R}, t) = \sum_i \exp[-(E_i - E_\mathrm{T})t]\, c_i\phi_i(\mathbf{R}) . \quad (9)$$

For long times one finds

$$\Phi(\mathbf{R}, t \to \infty) \simeq \exp[-(E_0 - E_\mathrm{T})t]\, c_0\phi_0(\mathbf{R}) , \quad (10)$$

which is proportional to the ground state wave function, ϕ_0.

The Hamiltonian is the sum of kinetic and potential terms: $\hat{H} = -(1/2)\nabla_\mathbf{R}^2 + V(\mathbf{R})$. Suppose for a moment that we can interpret the initial state, $\sum_i c_i\phi_i$, as a probability distribution. If we neglect the potential term then the imaginary-time Schrödinger equation (8) reduces to a diffusion equation in the configuration space. If, on the other hand, we neglect the kinetic term, (8) reduces to a rate equation. It should not be surprising that a short time slice of the imaginary-time evolution can be simulated by taking a population of configurations and subjecting them to random hops to simulate the diffusion process, and "birth" and "death" of configurations to simulate the rate

process. By "birth" and "death" we mean replicating some configurations and deleting others at the appropriate rates, a process which is often referred to as "branching".

Unfortunately the wave function cannot in general be interpreted as a probability distribution. A wave function for two or more identical fermions must have positive and negative regions, as should an excited state of any system. One can construct algorithms which are formally exact using two distributions of configurations with positive and negative weights [48], but they are inefficient and the scaling of the computational cost with system size is unclear.

The fixed-node approximation [49, 50] provides a way to evade the sign problem. This approximation is equivalent to placing an infinite repulsive potential barrier on the nodal surface of the trial wave function which is sufficiently strong to force the wave function to be zero on the nodal surface. (The nodal surface is the $3N - 1$ dimensional surface on which the wave function is zero and across which it changes sign.) In effect we solve the Schrödinger equation exactly within each volume enclosed by the nodal surface, subject to the boundary condition that the wave function is zero on the nodal surface. The infinite repulsive potential barrier has no effect if the trial nodal surface is placed correctly but, if it is not, the energy is always raised. It follows that the DMC energy is always less than or equal to the VMC energy with the same trial wave function, and always greater than or equal to the exact ground-state energy.

The fixed-node DMC algorithm described above is extremely inefficient and a vastly superior algorithm can be obtained by introducing an importance sampling transformation [51, 52]. Consider the mixed distribution,

$$f(\mathbf{R}, t) = \Psi_{\mathrm{T}}(\mathbf{R})\Phi(\mathbf{R}, t) \,, \tag{11}$$

which has the same sign everywhere if and only if the nodal surface of $\Phi(\mathbf{R}, t)$ equals that of $\Psi_{\mathrm{T}}(\mathbf{R})$. Substituting in equation (8) for Φ we obtain

$$-\frac{\partial f}{\partial t} = -\frac{1}{2}\nabla_{\mathbf{R}}^2 f + \nabla_{\mathbf{R}} \cdot [\mathbf{v}f] + [E_{\mathrm{L}} - E_{\mathrm{T}}]f \,, \tag{12}$$

where the $3N$-dimensional drift velocity is defined as

$$\mathbf{v}(\mathbf{R}) = \Psi_{\mathrm{T}}^{-1}(\mathbf{R})\nabla_{\mathbf{R}}\Psi_{\mathrm{T}}(\mathbf{R}) \,. \tag{13}$$

The three terms on the right-hand side of equation (12) correspond to diffusion, drift, and branching processes, respectively. The importance sampling transformation has several consequences. First, the density of configurations is increased where $|\Psi_{\mathrm{T}}|$ is large, so that the more important parts of the wave function are sampled more often. Second, the rate of branching is now controlled by the local energy which is normally a much smoother function than the potential energy. This is particularly important for the Coulomb interaction, which diverges when particles are coincident. The importance sampling transformation, together with an algorithm that imposes $f(\mathbf{R}, t) \geq 0$, ensures that Ψ_{T} and $\Phi(\mathbf{R}, t)$ have the same nodal surfaces, as can be seen in equation (11). The importance sampling transformation also reduces the statistical error bar on the estimate of the energy and leads to a zero variance property analogous to that in VMC.

The importance-sampled imaginary-time Schrödinger equation may be written in integral form:

$$f(\mathbf{R}, t) = \int G(\mathbf{R} \leftarrow \mathbf{R}', t - t') f(\mathbf{R}', t') \, d\mathbf{R}' , \tag{14}$$

where the Green's function $G(\mathbf{R} \leftarrow \mathbf{R}', t - t')$ is a solution of equation (12) satisfying the initial condition $G(\mathbf{R} \leftarrow \mathbf{R}', 0) = \delta(\mathbf{R} - \mathbf{R}')$. The exact Green's function can be sampled using the Green's function Monte Carlo (GFMC) algorithm developed by Kalos and coworkers [53, 54, 52, 55, 56].

Let us interpret $f(\mathbf{R}, t)$ as the probability distribution of a discrete population of P configurations with positive weights:

$$f(\mathbf{R}, t) \approx \sum_{p=1}^{P} w_p(t) \, \delta[\mathbf{R} - \mathbf{R}_p(t)] , \tag{15}$$

where the pth configuration at time t has position $\mathbf{R}_p(t)$ in configuration space and weight $w_p(t)$, and the "approximately equal" sign implies that this representation of $f(\mathbf{R}, t)$ is formally correct from a coarse-grained perspective, or if one performs an ensemble average of the right-hand side of equation (15). Using equation (14), the evolution of $f(\mathbf{R}, t)$ to time $t + \tau$ yields

$$\begin{aligned} f(\mathbf{R}, t + \tau) &\approx \sum_{p=1}^{P} w_p(t) \, G[\mathbf{R} \leftarrow \mathbf{R}_p(t), \tau] \\ &\approx \sum_{p=1}^{P} w_p(t + \tau) \, \delta[\mathbf{R} - \mathbf{R}_p(t + \tau)] . \end{aligned} \tag{16}$$

The dynamics of the configurations and their weights is governed by the Green's function.

The GFMC algorithm is computationally expensive, but considerably faster calculations can be made using an approximate Green's function which becomes exact in the limit of infinitely small time steps. Within the short-time approximation

$$G(\mathbf{R} \leftarrow \mathbf{R}', \tau) \simeq G_{\text{st}}(\mathbf{R} \leftarrow \mathbf{R}', \tau) = G_{\text{D}}(\mathbf{R} \leftarrow \mathbf{R}', \tau) G_{\text{B}}(\mathbf{R} \leftarrow \mathbf{R}', \tau) , \tag{17}$$

where

$$G_{\text{D}}(\mathbf{R} \leftarrow \mathbf{R}', \tau) = \frac{1}{(2\pi\tau)^{3N/2}} \exp\left(-\frac{[\mathbf{R} - \mathbf{R}' - \tau\mathbf{v}(\mathbf{R})]^2}{2\tau}\right) \tag{18}$$

is the drift-diffusion Green's function and

$$G_{\text{B}}(\mathbf{R} \leftarrow \mathbf{R}', \tau) = \exp\left(-\frac{\tau}{2} [E_{\text{L}}(\mathbf{R}) + E_{\text{L}}(\mathbf{R}') - 2E_{\text{T}}]\right) \tag{19}$$

is the branching factor.

The process described by $G_{\text{D}}(\mathbf{R} \leftarrow \mathbf{R}', \tau)$ is simulated by making each configuration $\mathbf{R}'$ in the population drift through a distance $\tau\mathbf{v}(\mathbf{R}')$, then diffuse by a random distance drawn from a Gaussian distribution of variance τ. Each configuration is then copied or deleted in such a fashion that, on average, $G_{\text{B}}(\mathbf{R} \leftarrow \mathbf{R}', \tau)$ configurations continue from the new position $\mathbf{R}$. When using the short time approximation, configurations occasionally attempt to cross the nodal surface but such moves may simply be rejected.

The short time approximation leads to a dependence of DMC results on the time step. It is important to investigate the size of the time step dependence, and it is common practice to extrapolate the energy to zero time step: see figure 5. It turns out that G_{st} does not precisely satisfy the detailed-balance condition, but it is standard practice to reinstate detailed balance by incorporating an accept-reject step. The importance-sampled fixed-node fermion DMC algorithm was first used by Ceperley and Alder in their ground-breaking study of the homogeneous electron gas (HEG) [3].

It can be seen that the reference energy E_T appears in the branching factor of equation (19). By adjusting the reference energy during the simulation we may keep the total population close to a target value, preventing the population from either increasing exponentially or dying out. An example of the behaviour of the total population and the reference energy can be seen in figure 1.

Another important aspect of practical implementations is that the particles are normally moved one at a time in both the VMC and DMC algorithms. For any given timestep, the probability of accepting single-particle moves is much larger than that of accepting an entire configuration move, resulting in shorter correlation times of the set of local energies, and the efficiency of configuration-space sampling is therefore considerably improved [57].

The initial configurations are normally taken from a VMC calculation and equilibrated within DMC for a period of imaginary time. The importance-sampled DMC algorithm generates configurations asymptotically distributed according to $f(\mathbf{R}) = \Psi_T(\mathbf{R})\phi_0(\mathbf{R})$, where ϕ_0 is the ground state of the Schrödinger equation subject to the fixed-node boundary condition. Noting that $\hat{H}\phi_0 = E_0\phi_0$ everywhere (except on the nodal surface where $\phi_0 = 0$) the fixed-node DMC energy can be written as

$$E_D \equiv E_0 = \frac{\langle\phi_0|\hat{H}|\Psi_T\rangle}{\langle\phi_0|\Psi_T\rangle} = \frac{\int f(\mathbf{R})E_L(\mathbf{R})\,d\mathbf{R}}{\int f(\mathbf{R})\,d\mathbf{R}} \tag{20}$$

which can be evaluated as a weighted average of the set of local energies sampled in the calculation.

Some example DMC data are shown in figure 1.

3. Trial wave functions

Trial wave functions are of central importance in VMC and DMC calculations because they introduce importance sampling and control both the statistical efficiency and accuracy obtained. The accuracy of a DMC calculation depends on the nodal surface of the trial wave function via the fixed-node approximation, while in VMC the accuracy depends on the entire trial wave function. VMC energies are therefore more sensitive to the quality of the trial wave function than DMC energies.
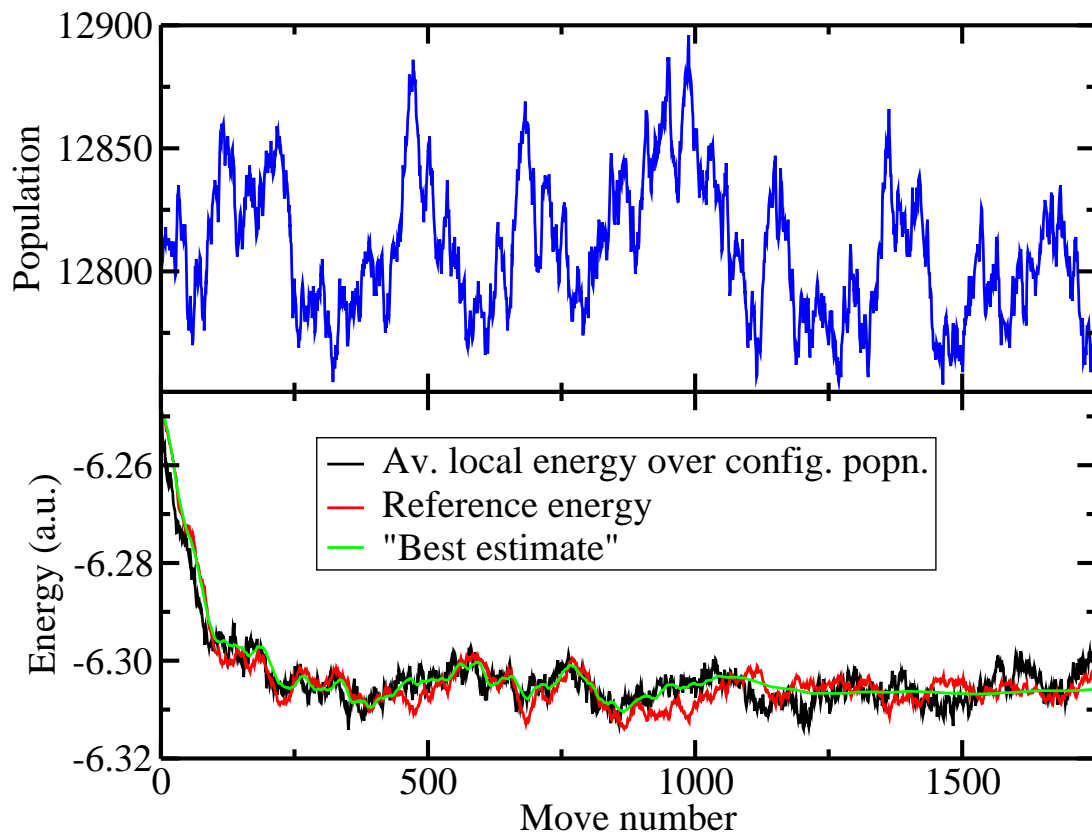
Figure 1. DMC data for a silane (SiH_4) molecule, with the ions represented by pseudopotentials. The upper panel shows the fluctuations in the population of configurations arising from the branching process used to simulate equation (19). The reference energy, E_T, is altered during the run to return the population towards a target value of 12800. The total energy is shown in the lower panel as a function of the move number. The black line shows the instantaneous value of the local energy averaged over the current population of configurations, the red line is the reference energy E_T, and the green line is the best estimate of the DMC energy as the simulation progresses. The configurations at move number zero are from the output of a VMC simulation, and the energy decays rapidly from its initial VMC value of about -6.250 a.u. and reaches a plateau with a DMC energy of about -6.305 a.u. The data up to move 1000 are deemed to form the equilibration phase, and are discarded.

3.1. Slater-Jastrow wave functions

QMC calculations require a compact trial wave function which can be evaluated rapidly. Most studies of electronic systems have used the Slater-Jastrow form, in which a pair of up- and down-spin determinants is multiplied by a Jastrow correlation factor,

$$\Psi_{SJ}(\mathbf{R}) = e^{J(\mathbf{R})} \det\left[\psi_n(\mathbf{r}_i^\uparrow)\right] \det\left[\psi_n(\mathbf{r}_j^\downarrow)\right] , \qquad (21)$$

where e^J is the Jastrow factor and $\det\left[\psi_n(\mathbf{r}_i^\uparrow)\right]$ is a determinant of single-particle orbitals for the up-spin electrons. The quality of the single-particle orbitals is very important,

and they are often obtained from density functional theory (DFT) or Hartree-Fock (HF) calculations. Note that the spin variables themselves do not appear in equation (21). Formally the sum over spin variables in the expectation values in equations (1) and (20) has already been performed and the single determinant with spin variables is replaced by two determinants of up- and down-spin orbitals whose arguments are the up- and down-spin electron coordinates $\mathbf{R}_\uparrow$ and $\mathbf{R}_\downarrow$, respectively. This is explained in more detail in reference [1].

The Jastrow factor is taken to be symmetric under the interchange of identical particles and its positivity means that it does not alter the nodal surface of the trial wave function. The Jastrow factor introduces correlation by making the wave function depend explicitly on the particle separations. The optimal Jastrow factor is normally small when particles with repulsive interactions (for example, two electrons) are close to one another and large when particles with attractive interactions (for example, an electron and a positron) are close to one another.

The Jastrow factor can also be used to ensure that the trial wave function obeys the Kato cusp conditions [58], which leads to smoother behaviour in the local energy $E_{\mathrm{L}}(\mathbf{R})$. When two particles interacting via the Coulomb potential approach one another, the potential energy diverges, and therefore the exact wave function Ψ must have a cusp so that the local kinetic energy $-(1/2)\Psi^{-1}\nabla^2\Psi$ supplies an equal and opposite divergence. It seems very reasonable to enforce the cusp conditions on trial wave functions because they are obeyed by the exact wave function. Imposition of the cusp conditions is in fact very important in both VMC and DMC calculations because divergences in the local energy lead to poor statistical behaviour and even instabilities in DMC calculations due to divergences in the branching factor.

Figure 2 shows the local energies generated during two VMC runs for a silane molecule in which the Si^{4+} and H^+ ions are described by smooth pseudopotentials. In figure 2(a) the trial wave function consists of a product of up- and down-spin Slater determinants of molecular orbitals. The Kato cusp conditions for electron-electron coalescences are therefore not satisfied and the local energy shows very large positive spikes when two electrons are close together. Figure 2(b) shows the effect of adding a Jastrow factor which satisfies the electron-electron cusp conditions. The large positive spikes in the local energy are removed and the mean energy is lowered. Some small spikes remain, and the frequency and size of the positive and negative spikes are roughly equal. These spikes arise from electrons approaching the nodes of the trial wave function, where the local kinetic energy diverges positively on one side of the node and negatively on the other side.

The basic Jastrow factor that we use for systems of electrons and ions contains the sum of homogeneous, isotropic electron-electron terms u, isotropic electron-nucleus terms χ centred on the nuclei, and isotropic electron-electron-nucleus terms f, also centred on the nuclei [59]. We use a Jastrow factor of the form $\exp[J(\mathbf{R})]$, where

$$J(\{\mathbf{r}_i\}, \{\mathbf{r}_I\}) = \sum_{i>j}^{N} u(r_{ij}) + \sum_{I=1}^{N_{\mathrm{ions}}} \sum_{i=1}^{N} \chi_I(r_{iI}) + \sum_{I=1}^{N_{\mathrm{ions}}} \sum_{i>j}^{N} f_I(r_{iI}, r_{jI}, r_{ij}) \,, \quad (22)$$
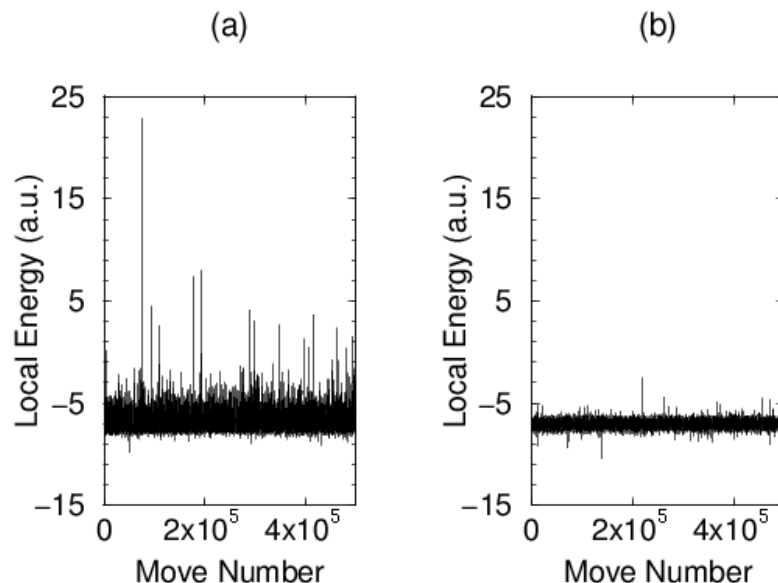
Figure 2. Local energy of a silane (SiH_4) molecule from a VMC calculation (a) using a Slater-determinant trial wave function and (b) including a Jastrow factor.

N is the number of electrons, N_{ions} is the number of ions, $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$, $\mathbf{r}_{iI} = \mathbf{r}_i - \mathbf{r}_I$, $\mathbf{r}_i$ is the position of electron i and $\mathbf{r}_I$ is the position of nucleus I. The functions u, χ, and f are represented by power expansions with optimisable coefficients. Different coefficients are used for terms involving different spins.

When using periodic boundary conditions, we often add a plane-wave term in the electron-electron separations, $p(\mathbf{r}_{ij})$, which describes similar sorts of correlation to the u term. The $u(r_{ij})$ term, however, is cut off at a distance less than or equal to the Wigner-Seitz radius of the simulation cell, and the p term adds variational freedom in the corners of the simulation cell. Occasionally we add a plane-wave expansion in electron position, $q(\mathbf{r}_i)$, and also occasionally add three-body electron-electron-electron terms.

We have recently developed a more general form of Jastrow factor [60] which allows the inclusion of higher order terms than those of equation (22), such as terms involving the distances between four or more particles. An example of the application of such a Jastrow factor to the H_2 molecule is shown in figure 3. The molecular orbital was calculated within Hartree-Fock theory and VMC calculations were performed using Jastrow factors of increasing complexity. The Jastrow factor of equation (22) includes electron-nucleus (e-N *etc.*), e-e and e-e-N terms, but the additional reductions in energy from including the e-N-N and e-e-N-N terms are clearly visible in figure 3.
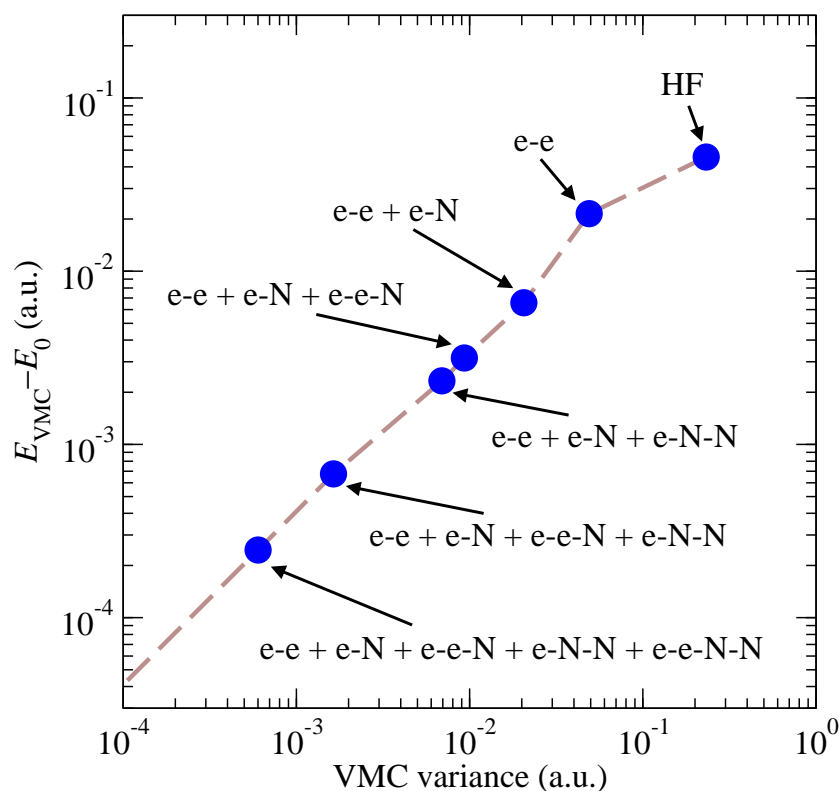
Figure 3. The difference between the VMC energy and the exact ground state energy against the variance of the VMC local energies on logarithmic scales for H_2 at a bond length of 1.397453 a.u. obtained using Jastrow factors of increasing complexity. "HF" indicates a wave function consisting of a molecular orbital obtained from a Hartree-Fock calculation and "e-e-N" denotes a term in the Jastrow factor involving the three distances between two electrons and one proton, *etc.*

3.2. Pairing wave functions

Slater-Jastrow wave functions are not appropriate for all systems. For example, the strongly attractive interaction between electrons and holes within an effective-mass theory leads to the formation of excitons, which are not well described by a Slater-Jastrow wave function. A more appropriate wave function is formed from the antisymmetrised product of identical electron-hole pairing functions ψ, multiplied by a Jastrow factor,

$$\Psi_{\text{SP}}(\mathbf{R}) = e^{J(\mathbf{R})} \det \left[\psi(\mathbf{r}_i^\uparrow, \mathbf{r}_j^\downarrow) \right].\tag{23}$$

It is also possible to include additional orbitals for unpaired particles within this wave function.

3.3. Multi-determinant wave functions

Multi-determinant expansions have been used with considerable success over many decades within the quantum chemistry community. The trial wave function can be

written as

$$\Psi_{\mathrm{MD}}(\mathbf{R}) = e^{J(\mathbf{R})} \sum_n c_n \det\left[\psi_n(\mathbf{r}_i^\uparrow)\right] \det\left[\psi_n(\mathbf{r}_j^\downarrow)\right] , \tag{24}$$

where the c_n are coefficients. This method provides a systematic approach to improving the trial wave function, and there have been numerous applications of multi-determinant trial wave functions in QMC calculations for small molecules. Such trial wave functions can capture near-degeneracy effects (also known as *static correlation*). Multi-determinant wave functions are not in general suitable for large systems because the number of determinants required to retrieve a given fraction of the correlation energy increases exponentially with system size. An exception to this occurs if only a small region of the system requires a multi-determinant description. An example of a DMC calculation of this type is the study of the electronic states formed by the strongly interacting dangling bonds at a neutral vacancy in diamond by Hood *et al.* [28].

3.4. Backflow wave functions

Additional correlation effects can be incorporated in the trial wave function using backflow transformations [61, 62]. Consider a solid ball falling through a classical liquid. The incompressible liquid is pushed out of the way and it fills in behind the ball to form a characteristic flow pattern. One can imagine that similar correlations occur as a quantum particle moves through a quantum fluid, as shown in figure 4. Much of this correlation can be captured in a Jastrow factor which, however, preserves the nodal surface of the wave function. The backflow motion gives an additional contribution which leaves its imprint on the nodes. Quantum backflow was discussed by Feynman and coworkers [61, 62] for excitations in ^{4}He and the effective mass of a ^{3}He impurity in liquid ^{4}He. Backflow wave functions have been used successfully in QMC studies of liquid He [9, 10], the electron gas [63, 64, 4], hydrogen systems [32], and various inhomogeneous systems [57, 65, 66].

The backflow wave functions we use [57] can be written as

$$\Psi_{\mathrm{BF}}(\mathbf{R}) = e^{J(\mathbf{R})} \det\left[\psi_i(\mathbf{r}_j^\uparrow + \boldsymbol{\xi}_j(\mathbf{R}))\right] \det\left[\psi_i(\mathbf{r}_j^\downarrow + \boldsymbol{\xi}_j(\mathbf{R}))\right] . \tag{25}$$

For a system of N electrons and N_{ion} ions we write the backflow displacement for electron i in the form

$$\boldsymbol{\xi}_i = \sum_{j \neq i}^{N} \eta_{ij} \mathbf{r}_{ij} + \sum_{I}^{N_{\mathrm{ion}}} \mu_{iI} \mathbf{r}_{iI} + \sum_{j \neq i}^{N} \sum_{I}^{N_{\mathrm{ion}}} (\Phi_i^{jI} \mathbf{r}_{ij} + \Theta_i^{jI} \mathbf{r}_{iI}) . \tag{26}$$

In this expression $\eta_{ij} = \eta(r_{ij})$ is a function of electron-electron separation, $\mu_{iI} = \mu(r_{iI})$ is a function of electron-ion separation, and $\Phi_i^{jI} = \Phi(r_{iI}, r_{jI}, r_{ij})$ and $\Theta_i^{jI} = \Theta(r_{iI}, r_{jI}, r_{ij})$. We parameterise the functions η, μ, Φ, and Θ using power expansions with optimisable coefficients [57].
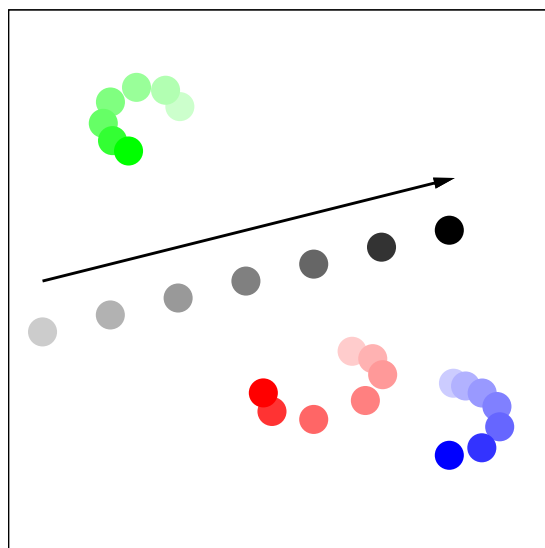
Figure 4. Effect of the motion of an electron (black, with the arrow showing the direction of motion) on the backflow-transformed coordinates of three opposite-spin electrons (red, green and blue). Circles with the same colour intensity correspond to the same instant in the motion.

3.5. Other wave functions

The wave function types of equations (21), (23), (24), and (25) can be combined in various ways within the CASINO code [45] so that, for example, it is possible to use Slater-Jastrow-pairing-backflow wave functions, *etc.* Of course the range of possible wave functions could be extended by, for example, including Pfaffian wave functions [67, 68], *etc.*

4. Optimisation of trial wave functions

Optimising trial wave functions is a very important part of QMC calculations which can consume large amounts of human and computing resources. With modern stochastic methods it is possible to optimise hundreds or even thousands of parameters in the wave function. The parameters which can be optimised include those in the Jastrow factor, the coefficients of determinants in a multi-determinant wave function, the parameters in the backflow functions, and the parameters in single-particle and pairing orbitals.

The trial wave function used in a DMC calculation should ideally be optimised within DMC, but reliable and efficient methods to achieve this are still under development [69, 70]. Minimisation of the DMC energy has been performed "by hand" for small numbers of parameters [5, 8]. Wave function optimisation within CASINO is performed by minimising the VMC energy or its variance.

Optimising wave functions by minimising the variance of the energy is an old idea dating back to the 1930s. The first application within Monte Carlo methods may have

been by Conroy [71], but the method was popularised within QMC by the work of Umrigar and coworkers [72]. It is now generally believed that it is better to minimise the VMC energy than its variance, but it has proved more difficult to develop robust and efficient algorithms for this purpose. Since the trial wave function forms used cannot generally represent energy eigenstates exactly, except in trivial cases, the minima in the energy and variance do not coincide. Energy minimisation should therefore produce lower VMC energies, and although it does not necessarily follow that it produces lower DMC energies, experience indicates that, more often than not, it does.

4.1. Variance minimisation

The variance of the VMC energy is

$$\sigma^2(\boldsymbol{\alpha}) = \frac{\int [\Psi_{\mathrm{T}}^{\boldsymbol{\alpha}}]^2 [E_{\mathrm{L}}^{\boldsymbol{\alpha}} - E_{\mathrm{V}}^{\boldsymbol{\alpha}}]^2 \, d\mathbf{R}}{\int [\Psi_{\mathrm{T}}^{\boldsymbol{\alpha}}]^2 \, d\mathbf{R}} \;, \tag{27}$$

where $\boldsymbol{\alpha}$ denotes the set of variable parameters. The minimum possible value of $\sigma^2(\boldsymbol{\alpha})$ is zero, which is obtained if and only if $\Psi_{\mathrm{T}}^{\boldsymbol{\alpha}}$ is an exact eigenstate of $\hat{H}$. In practice the trial wave function forms used are incapable of representing the exact eigenstates. Nevertheless, the minimum value of $\sigma^2(\boldsymbol{\alpha})$ is still expected to correspond to a reasonable set of wave function parameters.

Minimisation of $\sigma^2(\boldsymbol{\alpha})$ is carried out via a correlated sampling approach in which a set of configurations distributed according to $[\Psi_{\mathrm{T}}^{\boldsymbol{\alpha}_0}]^2$ is generated, where $\boldsymbol{\alpha}_0$ is an initial set of parameter values [73]. $\sigma^2(\boldsymbol{\alpha})$ is then evaluated as

$$\sigma^2(\boldsymbol{\alpha}) = \frac{\int [\Psi_{\mathrm{T}}^{\boldsymbol{\alpha}_0}]^2 \, w_{\boldsymbol{\alpha}_0}^{\boldsymbol{\alpha}} \, [E_{\mathrm{L}}^{\boldsymbol{\alpha}} - E_{\mathrm{V}}^{\boldsymbol{\alpha}}]^2 \, d\mathbf{R}}{\int [\Psi_{\mathrm{T}}^{\boldsymbol{\alpha}_0}]^2 \, w_{\boldsymbol{\alpha}_0}^{\boldsymbol{\alpha}} \, d\mathbf{R}} \;, \tag{28}$$

where the integrals contain weights, $w_{\boldsymbol{\alpha}_0}^{\boldsymbol{\alpha}}$, given by

$$w_{\boldsymbol{\alpha}_0}^{\boldsymbol{\alpha}}(\mathbf{R}) = \frac{[\Psi_{\mathrm{T}}^{\boldsymbol{\alpha}}]^2}{[\Psi_{\mathrm{T}}^{\boldsymbol{\alpha}_0}]^2} \;, \tag{29}$$

and E_{V} is evaluated using

$$E_{\mathrm{V}} = \frac{\int [\Psi_{\mathrm{T}}^{\boldsymbol{\alpha}_0}]^2 \, w_{\boldsymbol{\alpha}_0}^{\boldsymbol{\alpha}} \, E_{\mathrm{L}}^{\boldsymbol{\alpha}} \, d\mathbf{R}}{\int [\Psi_{\mathrm{T}}^{\boldsymbol{\alpha}_0}]^2 \, w_{\boldsymbol{\alpha}_0}^{\boldsymbol{\alpha}} \, d\mathbf{R}} \;. \tag{30}$$

After generating the initial set of configurations, the optimisation proceeds using standard techniques to locate the new parameter values which minimise $\sigma^2(\boldsymbol{\alpha})$. With perfect sampling $\sigma^2(\boldsymbol{\alpha})$ is independent of the initial parameter values $\boldsymbol{\alpha}_0$. For real (finite) sampling, however, one runs into problems because the values of $w_{\boldsymbol{\alpha}_0}^{\boldsymbol{\alpha}}$ for different configurations can vary by many orders of magnitude if $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}_0$ differ substantially. During the minimisation procedure a few configurations (often only one) acquire very large weights and the estimate of the variance is reduced almost to zero by a poor set of parameter values. This optimisation scheme is therefore often unstable, and in practice modified versions of it are used.

The above scheme can be made much more stable by altering the weights $w_{\boldsymbol{\alpha}_0}^{\boldsymbol{\alpha}}$. A robust procedure is to set all the weights $w_{\boldsymbol{\alpha}_0}^{\boldsymbol{\alpha}}$ in equation (28) to unity, which is

reasonable because the minimum value of $\sigma^2(\boldsymbol{\alpha}) = 0$ is still obtained only if $E_{\mathrm{L}}(\mathbf{R})$ is a constant independent of $\mathbf{R}$, which holds only for eigenstates of the Hamiltonian. We call this the "unreweighted variance" minimisation method. The procedure is cycled until the parameters converge to their optimal values (within the statistical noise). For a number of model systems it was found that the trial wave functions generated by unreweighted variance minimisation iterated to self-consistency have a lower variational energy than wave functions optimised by reweighted variance minimisation [74].

We also have a particularly fast algorithm for optimising the linear parameters in a Jastrow factor [74]. If the Jastrow factor of equation (22) can be written in the form

$$J(\mathbf{R}) = \sum_n \alpha_n f_n(\mathbf{R}) \,, \tag{31}$$

then it can be shown that the unreweighted variance of the VMC energy is a quartic function of the linear parameters α_n. This has two advantages: (i) the unreweighted variance can be evaluated extremely rapidly at a cost which depends only on the number of parameters and is independent of the number of particles; and (ii) the unreweighted variance along a line in parameter space is a quartic polynomial. This is useful because it allows the exact global minimum of the unreweighted variance along the line to be computed analytically by solving the cubic equation obtained by setting the derivative equal to zero.

The unreweighted variance minimisation method works well for optimising Jastrow factors, but it often performs poorly when parameters which alter the nodal surface of Ψ_{T} are optimised. The problem is that the local energy E_{L} generally diverges for a configuration on the nodal surface. As the parameter values are changed during a minimisation cycle the nodal surface can move through a configuration, resulting in a very large (positive or negative) value of E_{L}, which adversely affects the optimisation. Such an effect would not occur when using the weights $w_{\boldsymbol{\alpha}_0}^{\boldsymbol{\alpha}}$ because they go to zero on the nodal surface. We have developed two schemes which solve this problem. In the first scheme we limit the weights by replacing them with $\min(w_{\boldsymbol{\alpha}_0}^{\boldsymbol{\alpha}}, W)$, so that the weight goes to zero on the nodal surface but can never become larger than a chosen value W. In the second scheme we use a weight which goes smoothly to zero as E_{L} deviates from an estimate of the energy.

Unreweighted variance minimisation belongs to a wider class of wave-function optimisation methods which are based on minimising a measure of the spread of the set of local energies. Another measure of spread that we have used with considerable success for wave-function optimisation is the mean absolute deviation of the local energies of a set of configurations from the median energy,

$$\mathcal{M} = \frac{\int [\Psi_{\mathrm{T}}^{\boldsymbol{\alpha}_0}(\mathbf{R})]^2 |E_{\mathrm{L}}^{\boldsymbol{\alpha}}(\mathbf{R}) - E_{\mathrm{m}}^{\boldsymbol{\alpha}}| \, d\mathbf{R}}{\int [\Psi_{\mathrm{T}}^{\boldsymbol{\alpha}_0}(\mathbf{R})]^2 \, d\mathbf{R}} \,. \tag{32}$$

In this expression, $E_{\mathrm{m}}^{\boldsymbol{\alpha}}$ is the median value of the local energies evaluated with the parameter values $\boldsymbol{\alpha}$. This is useful for optimising parameters that affect the nodal surface, because outlying local energies are less significant.

4.2. Energy minimisation

A well-known method for finding approximations to the eigenstates of a Hamiltonian is to express the wave function as a linear combination of basis states g_i,

$$\Psi_{\mathrm{T}}(\mathbf{R}) = \sum_{i=1}^{p} \beta_i \, g_i(\mathbf{R}) \, , \tag{33}$$

calculate the matrix elements $H_{ij} = \langle g_i | \hat{H} | g_j \rangle$ and $S_{ij} = \langle g_i | g_j \rangle$, and solve the two-sided eigenproblem $\sum_j H_{ij}\beta_j = E \sum_j S_{ij}\beta_j$ by standard diagonalisation techniques. One can also do this in QMC [75], although the statistical noise in the matrix elements leads to slow convergence with respect to the number of configurations used to evaluate the integrals.

Nightingale and Melik-Alaverdian [76] reformulated the diagonalisation procedure as a least-squares fit rather than integral evaluation, which leads to much faster convergence with the number of configurations. Let us assume that the set $\{g_i\}$ spans an invariant subspace of $\hat{H}$, which means that the result of acting $\hat{H}$ on any member of the set $\{g_i\}$ can be expressed as a linear combination of the $\{g_i\}$, *i.e.*,

$$\hat{H}g_i(\mathbf{R}) = \sum_{i=1}^{p} \mathcal{E}_{ij} g_j(\mathbf{R}) \quad \forall \, i \, . \tag{34}$$

The eigenstates and associated eigenvalues of $\hat{H}$ can then be obtained by diagonalising the matrix $\mathcal{E}_{ij}$. Within a Monte Carlo approach we could evaluate the $g_i(\mathbf{R})$ and $\hat{H}g_i(\mathbf{R})$ for p uncorrelated configurations generated by a VMC calculation and solve the resulting set of linear equations for the $\mathcal{E}_{ij}$. For problems of interest, however, the assumption that the set $\{g_i\}$ span an invariant subspace of $\hat{H}$ does not hold and there exists no set of $\mathcal{E}_{ij}$ which solves equation (34). If we took p configurations and solved the set of p linear equations, the values of $\mathcal{E}_{ij}$ would depend on which configurations had been chosen. To overcome this problem, a number of configurations $M \gg p$ is sampled to obtain an overdetermined set of equations which can be solved in a least-squares sense using singular value decomposition. In fact Nightingale and Melik-Alaverdian recommended that equation (34) be divided by $\Psi_{\mathrm{T}}(\mathbf{R})$ so that in the limit of perfect sampling the scheme corresponds precisely to standard diagonalisation.

The method of Nightingale and Melik-Alaverdian works very well for linear variational parameters as in equation (33). The natural generalisation to parameters which appear non-linearly in Ψ_{T} is to consider a first-order Taylor expansion of the trial wave function about the initial parameter values,

$$\Psi_T^{\boldsymbol{\alpha}} = \Psi_T^{\boldsymbol{\alpha}_0} + \sum_{i=1}^{p} \frac{\partial \Psi_T^{\boldsymbol{\alpha}_0}}{\partial \alpha_i} \left(\alpha_i - \alpha_i^0 \right) + \mathcal{O}\left[(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^2 \right] \, , \tag{35}$$

where, by comparison with equation (33), g_i can be identified with the derivative of the initial trial wave function with respect to the ith parameter, $\beta_i = \alpha_i - \alpha_i^0$, and the initial trial wave function represents an additional basis function g_0 with a fixed coefficient $\beta_0 = 1$.

In its simplest form this algorithm turns out to be highly unstable because neglecting the second-order contribution in equation (35) is often inadequate. Umrigar and coworkers [77, 78] showed how this method can be stabilised. The details of the stabilisation procedures are quite involved and we refer the reader to the original papers [77, 78] for the details. The stabilised algorithm works well and is quite robust. The VMC energies given by this method are usually lower than those obtained from any of the variance-based algorithms described in section 4.1, although the difference is often small.

5. QMC calculations within periodic boundary conditions

QMC calculations for extended systems may be performed using cluster models or periodic boundary conditions, just as in other techniques. Periodic boundary conditions are preferred because they give smaller finite size effects. One can also use the standard supercell approach for systems that lack three-dimensional periodicity in which a cell containing, for example, a point defect and a small part of the host crystal, is repeated periodically throughout space. Just as in other electronic structure methods, one must ensure that the supercell is large enough for the interactions between defects in different supercells to be small.

When using standard single-particle-like theories within periodic boundary conditions such as density functional theory, the charge density and potentials are taken to have the periodicity of a chosen unit cell or supercell. The single particle orbitals can then be chosen to obey Bloch's theorem and the results for the infinite system are obtained by summing quantities obtained from the different Bloch wave vectors within the first Brillouin zone. This procedure can also be applied within HF calculations, although the Coulomb interaction couples the Bloch wave vectors in pairs. In calculations with the wave functions described in section 3, these simplifications do not arise and QMC calculations are performed at a single k-point. A single k-point normally gives a poor representation of the infinite-system result, so that larger non-primitive simulation cells are often used. It is also possible to perform QMC calculations at a set of different k-points [79, 80] and average the results [81], which can substantially reduce the size-dependence of the results, especially for metals

Many-body techniques such as QMC also suffer from finite size errors arising from long-ranged interactions, most notably the Coulomb interaction. Coulomb interactions are normally included within periodic boundary conditions calculations using the Ewald interaction. Long-ranged interactions induce long-ranged correlation effects, and if the simulation cell is not large enough these effects are described incorrectly. Such effects are absent in local DFT calculations because the interaction energy is written in terms of the electronic charge density, but HF calculations show very strong effects of this kind and various ways to accelerate the convergence have been developed. The finite size effects arising from the long-ranged interaction can be divided into potential and kinetic energy contributions [82, 83]. The potential energy component can be removed

from the calculations by replacing the Ewald interaction by the so-called model periodic Coulomb (MPC) interaction [84, 85, 86]. Recent work has added substantially to our understanding of finite size effects, and theoretical expressions have been derived for them [82, 83], but at the moment it seems that they cannot entirely replace extrapolation procedures.

Kwee *et al.* [87] have developed an alternative approach for estimating finite size errors in QMC calculations. DMC results for the three-dimensional HEG are used to obtain a system-size-dependent local density approximation (LDA) functional. The correction to the total energy is given by the difference between the DFT energies for the finite-sized and infinite systems. This approach appears promising, although it does rely on the LDA giving a reasonable description of the system.

6. Pseudopotentials in QMC calculations

The computational cost of a DMC calculation increases with the atomic number Z of the atoms as roughly $Z^{5.5}$ [88, 89] which makes calculations with $Z > 10$ extremely expensive. This problem can be solved by using pseudopotentials to represent the effect of the atomic core on the valence electrons. The use of non-local pseudopotentials within VMC is quite straightforward [90, 91], but DMC poses an additional problem because the use of a non-local potential is incompatible with the fixed-node boundary condition. To circumvent this difficulty an additional approximation is made. In the "locality approximation" [92] the non-local part of the pseudopotential $\hat{V}_{\mathrm{nl}}$ is taken to act on the trial wave function rather than the DMC wave function, *i.e.*, $\hat{V}_{\mathrm{nl}}$ is replaced by $\Psi_{\mathrm{T}}^{-1}\hat{V}_{\mathrm{nl}}\Psi_{\mathrm{T}}$. The leading-order error term in the locality approximation is proportional to $(\Psi_{\mathrm{T}} - \phi_0)^2$ [92], where ϕ_0 is the exact fixed-node ground state wave function, although it can be of either sign, so that the variational property of the algorithm is lost. Casula *et al.* [93, 94] have introduced a fully variational "semi-localisation" scheme for dealing with non-local pseudopotentials within DMC, which also shows superior numerical stability to the locality approximation.

Currently it is not possible to generate pseudopotentials entirely within a QMC framework, and therefore they are obtained from other sources. There is evidence that HF theory provides better pseudopotentials than DFT for use within QMC calculations [95], and we have developed smooth relativistic HF pseudopotentials for H to Ba and Lu to Hg, which are suitable for use in QMC calculations [96, 97, 98]. Another set of pseudopotentials for use in QMC calculations has been developed by Burkatzki *et al.* [99]. In the few cases where reliable tests have been performed [100, 101], the pseudopotentials of Refs. [96, 97, 98] and those of [99] have produced almost identical results, although those of references [96, 97, 98] are a little more efficient as they have smaller core radii.

7. DMC calculations for excited states

DMC can be applied to excited states as the fixed node constraint ensures convergence to the lowest energy state compatible with the nodal surface of the trial wave function. DMC therefore gives the exact energy of *any* state if the nodal surface is exact, and it gives an approximate energy with an approximate nodal surface. An important difference from the ground state case is that the existence of a variational principle for excited state energies cannot in general be guaranteed, and it depends on the symmetry of the trial wave function [102]. In practice DMC works quite well for excited states [20, 21, 103, 104, 25, 26, 105]. Ceperley and Bernu [106] have devised a method which combines DMC and the variational principle to calculate the eigenvalues of several different excited states simultaneously. However, this method suffers from stability problems in large systems.

8. Sources of error and statistical analysis

8.1. Sources of error in DMC calculations

The potential sources of errors in DMC calculations may be summarised as follows.
- Statistical errors. The standard error in the mean is proportional to $1/\sqrt{M}$, where M is the number of particles moves. It therefore costs a factor of 100 in computer time to reduce the statistical error bars by a factor of 10. On the other hand, a random error is much better than a systematic one as its size can normally be reliably estimated.
- Fixed-node error. This is the central approximation of the DMC technique, and is normally the limiting factor in the accuracy of the results.
- Time-step bias. The short time approximation leads to a bias in the f distribution and hence in expectation values. This bias is often significant and can be of either sign, but it can be largely removed by performing calculations for different time steps and extrapolating to zero time step or by simply choosing a small enough time step. An example of time-step extrapolation is shown in figure 5.
- Population control bias. The f distribution is represented by a finite population of configurations which fluctuates due to branching. The population may be controlled in various ways, but this introduces a population control bias which is positive and falls off as the reciprocal of the population. In practice the population control bias is normally so small that it is difficult to detect [107, 5].
- Finite size errors within periodic boundary conditions calculations. It is important to correct for finite size effects carefully, as mentioned in section 5.
- The pseudopotential approximation inevitably introduces errors. In DMC there is an additional error arising from the localisation [92] or semi-localisation [94] of the non-local pseudopotential operator. The localisation error appears to be quite small in the cases for which it has been tested [65].
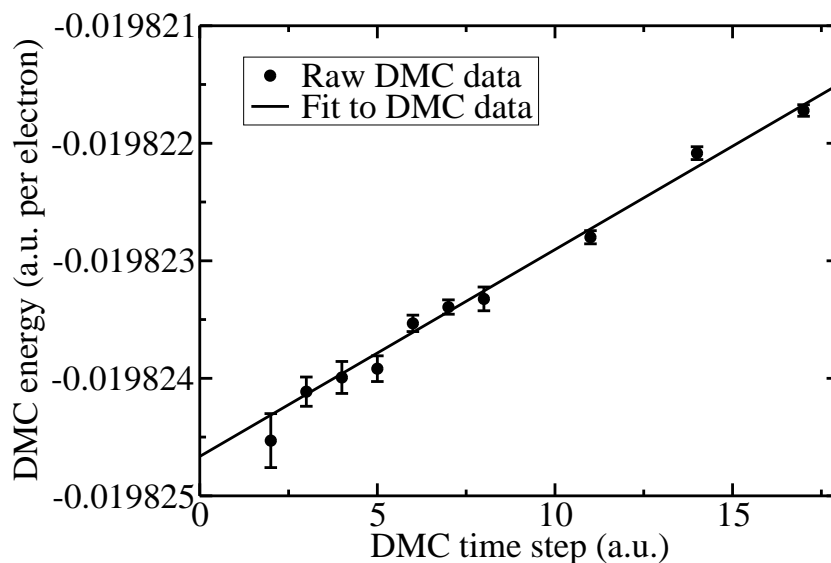
Figure 5. DMC energy against time step for a 64-electron ferromagnetic 2D hexagonal Wigner crystal at density parameter $r_s = 50$ a.u. with a Slater-Jastrow wave function. The solid line is a linear fit to the data.

8.2. Practical methods for handling statistical errors in QMC results

Two main practical problems are encountered when dealing with errors in the QMC data: the data are serially correlated and the underlying probability distributions are non-Gaussian. The probability distribution of the local energies has $|E - E_0|^{-4}$ tails, where E_0 is a constant. These tails arise from singularities in the local energy such as the divergence at the nodal surface [96, 97], as shown in figure 6. In consequence, although the mean energy and its variance are well defined, the variance of the variance is infinity. For other quantities the problem may be even more severe; for example, the probability distributions for the Pulay terms in the forces described in section 10.2 decay as $|F - F_0|^{-5/2}$, so that the variance of the force is infinity [108]. Reasonably robust estimates of the errors can still be made, although it has to be accepted that they are not as well founded as for Gaussian statistics.

The data produced by VMC and DMC calculations are correlated from one step to the next. The problem is very important in DMC because short time steps are used to reduce the effect of the approximation in the Green's function. The simulation effectively produces only one independent data point per correlation time, so that the estimate of the statistical error obtained on the assumption that the data points are independent is too small. We use the "blocking method" to obtain an estimate of the error. In this approach adjacent data points are averaged to form block averages [109]. This procedure is carried out recursively so that after each blocking transformation the number of data points is reduced by one half. An example of blocking is shown in figure 7. The computed value of the standard error Δ_k increases with the number of blocking transformations k until a limiting value is reached when the block length starts to exceed the correlation time. The standard error in the mean is estimated by the
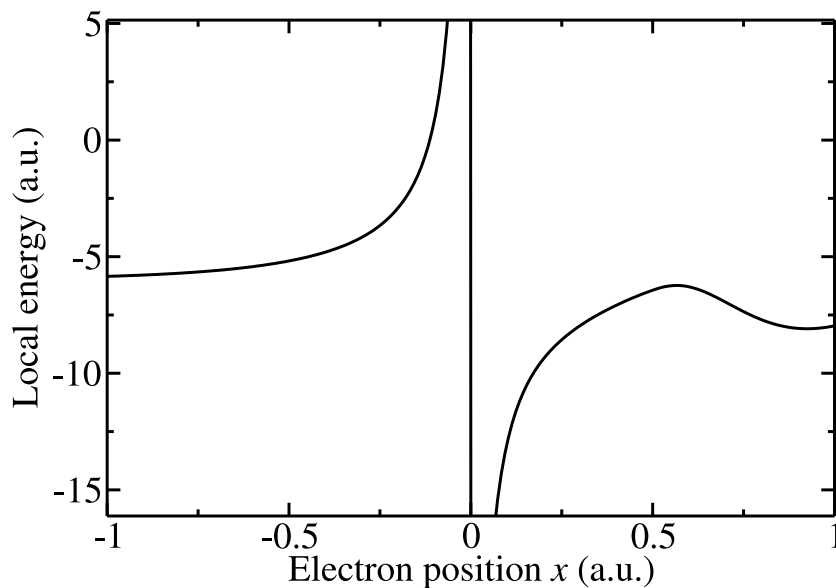
Figure 6. Variation in the local energy E_L of a silane (SiH_4) molecule as an electron moves through the nodal surface at $x = 0$. The local energy diverges as $1/x$.

value of Δ on the plateau. Because the sizes of the error bars on QMC expectation values are themselves approximate estimates, apparent outliers in QMC data can be more common than one might expect on the basis of Gaussian statistics.
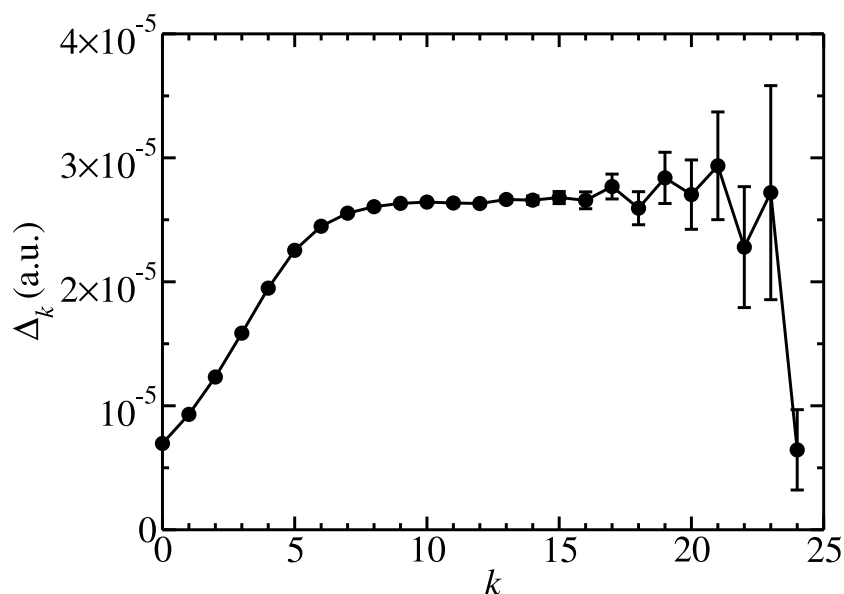
Figure 7. Blocking analysis of data for an (all-electron) lithium atom. The blocking analysis indicates that the true standard error in the mean is about $\Delta = 2.6 \times 10^{-5}$ a.u., which is reached at about blocking transformation $k = 10$, while the raw value is $\Delta_0 = 7.0 \times 10^{-6}$ a.u.

9. Evaluating other expectation values

As mentioned in section 1, VMC and DMC can be used to calculate expectation values of many time-independent operators, not just the Hamiltonian. Typical quantities of interest are particle densities, pair correlation functions, and one- and two-body density matrices, all of which can be evaluated using the CASINO code. It is not possible to obtain unbiased expectation values directly from the DMC distribution, $f(\mathbf{R})$, for operators which do not commute with the Hamiltonian (which includes all of the quantities mentioned in the previous sentence). Unbiased (within the fixed-node approximation) estimates can be obtained as pure expectation values,

$$\langle \hat{A} \rangle = \frac{\int \phi_0(\mathbf{R}) \hat{A} \phi_0(\mathbf{R}) \, d\mathbf{R}}{\int \phi_0^2(\mathbf{R}) \, d\mathbf{R}} \ . \tag{36}$$

Pure expectation values can be obtained using a variety of methods: the approximate (but often very accurate) extrapolation technique [55], the future walking technique [110, 111] which is formally exact but statistically poorly behaved, and the reptation QMC technique of Baroni and Moroni [112], which is formally exact and well behaved, but quite expensive. The extrapolation technique can be used for any operator, but the future walking and reptation techniques are limited to spatially local multiplicative operators.

Here we shall illustrate the use of the extrapolation technique [55] to calculate the charge density of a Wigner crystal. The pure estimate of the charge density ρ is approximated as

$$\rho_{\text{ext}} \simeq 2\rho_{\text{DMC}} - \rho_{\text{VMC}}. \tag{37}$$

The errors in both the VMC and DMC charge densities ρ_{VMC} and ρ_{DMC} are linear in the error in the trial wave function, but the error in the extrapolated estimate ρ_{ext} is quadratic in the error in the wave function.

At low densities the HEG freezes into a Wigner crystal to minimise the electrostatic repulsion between electrons. The charge density of a 2D Wigner crystal [8, 113] close to the crystallisation density is shown in figure 8. VMC, DMC and extrapolated results are shown for two different trial wave functions. It can be seen that the dependence of the extrapolated estimate on the trial wave function is much smaller than for the raw VMC and DMC estimates, so we may have more confidence in the extrapolated estimate of the charge density.

10. Energy differences and energy derivatives

In electronic structure theory one is almost always interested in the differences in energy between systems. All electronic structure methods for complex systems rely for their accuracy on the cancellation of errors in energy differences. In DMC this helps with all the sources of error mentioned in section 8 except the statistical errors. Fixed-node errors tend to cancel because the DMC energy is an upper bound, but even though
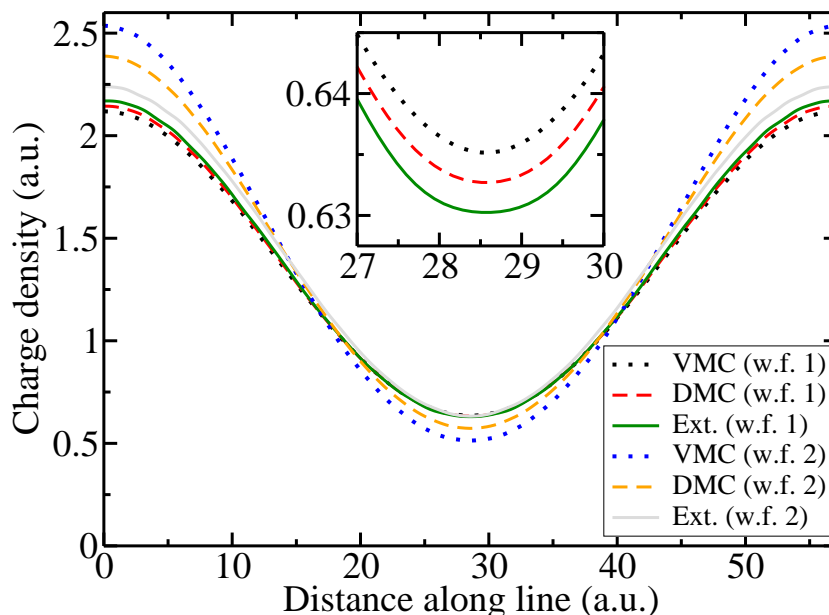
Figure 8. Charge density of a triangular antiferromagnetic Wigner crystal at density parameter $r_s = 30$ a.u., plotted along a line between a pair of nearest-neighbour lattice sites. Two different wave functions are used: wave function 1 was optimised by variance minimisation, while wave function 2 was optimised by energy minimisation. The inset shows the extrapolation with wave function 1 at the minimum in greater detail.

DMC often retrieves 95% or more of the correlation energy, non-cancellation of nodal errors is the most important source of error in DMC results.

10.1. Energy differences in QMC

Correlated sampling methods allow the computation of the energy difference between two similar systems with a smaller statistical error than those obtained for the individual energies [73]. Correlated sampling is relatively straightforward in VMC, and a version of it is described in section 4.1 in the context of optimising wave functions by variance minimisation.

10.2. Energy derivatives (forces) in QMC

Atomic forces are useful for relaxing the structures of molecules and solids, calculating their vibrational properties, and for performing molecular dynamics (MD) simulations. It has proved difficult to develop accurate and efficient methods for calculating atomic forces within QMC, although considerable progress has been made in recent years. Difficulties have arisen in obtaining accurate expressions for DMC forces which can readily be evaluated and in the statistical properties of the expressions, which are not as advantageous as those for the energy.

According to the Hellmann-Feynman theorem (HFT), the derivative of the energy

with respect to a parameter λ in the Hamiltonian is

$$E' = \frac{\int \Psi \, \hat{H}' \, \Psi \, d\mathbf{R}}{\int \Psi \, \Psi \, d\mathbf{R}} \, , \tag{38}$$

where the primes denote derivatives with respect to λ. This expression is valid when Ψ is an exact eigenstate of $\hat{H}$.

Unfortunately the HFT is not normally applicable within QMC because the wave functions are approximate. Exact expressions for the VMC and DMC forces must therefore contain additional Pulay terms which depend on Ψ'_{T}. To define the force properly it is therefore necessary to define and evaluate Ψ'_{T}.

The DMC algorithm solves for the ground state of the fixed-node Hamiltonian exactly and therefore the HFT holds. Unfortunately the fixed-node Hamiltonian is different from the physical Hamiltonian because it contains an additional infinite potential barrier on the nodal surface of Ψ_{T} which forces the DMC wave function ϕ_0 to go to zero. As λ varies, the nodal surface, and hence the infinite potential barrier, moves, giving a contribution to $\hat{H}'$ [114, 115, 116] which depends on Ψ_{T} and Ψ'_{T} and is classified as a Pulay term.

The Pulay terms arising from the derivative of the mixed estimate of the energy of equation (20) contain ϕ'_0, the derivative of the DMC wave function. This quantity cannot readily be evaluated, and the approximation

$$\frac{\phi'_0}{\phi_0} \simeq \frac{\Psi'_{\mathrm{T}}}{\Psi_{\mathrm{T}}} \tag{39}$$

has normally been used [117, 118, 119, 120, 121, 122, 116, 123, 124]. However, it leads to errors of first order in $(\Psi_{\mathrm{T}} - \phi_0)$ and $(\Psi'_{\mathrm{T}} - \phi'_0)$; therefore its accuracy depends sensitively on the quality of Ψ_{T} and Ψ'_{T}, and in practice this approximation is often inadequate.

The pure DMC energy,

$$E_{\mathrm{D}} = \frac{\int \phi_0 \hat{H} \phi_0 \, d\mathbf{R}}{\int \phi_0 \phi_0 \, d\mathbf{R}} \, , \tag{40}$$

is equal to the mixed DMC energy. Forces may also be calculated within pure DMC, and although this is more expensive it brings significant advantages. The derivative E'_{D} contains the derivative of the DMC wave function, ϕ'_0. However, Badinski *et al.* [116] showed that ϕ'_0 can be eliminated from the pure DMC expression, giving the exact result

$$E'_{\mathrm{D}} = \frac{\int \phi_0 \phi_0 \, \phi_0^{-1} \hat{H}' \phi_0 \, d\mathbf{R}}{\int \phi_0 \phi_0 \, d\mathbf{R}} - \frac{1}{2} \frac{\int \phi_0 \phi_0 \, \Psi_{\mathrm{T}}^{-2} |\nabla_{\mathbf{R}} \Psi_{\mathrm{T}}| \Psi'_{\mathrm{T}} \, d\mathbf{S}}{\int \phi_0 \phi_0 \, d\mathbf{R}} \, , \tag{41}$$

where $d\mathbf{S}$ denotes an element of the nodal surface. Unfortunately it is not straightforward to evaluate integrals over the nodal surface. The nodal surface integral can be converted into a volume integral in which ϕ'_0 does not appear using an approximation with an error of order $(\Psi_{\mathrm{T}} - \phi_0)^2$, giving

$$E'_{\mathrm{D}} = \frac{\int \phi_0 \phi_0 \left[\phi_0^{-1} \hat{H}' \phi_0 + \Psi_{\mathrm{T}}^{-1} \left(\hat{H} - E_{\mathrm{D}} \right) \Psi'_{\mathrm{T}} \right] d\mathbf{R}}{\int \phi_0 \phi_0 \, d\mathbf{R}} + \tag{42}$$

$$\frac{\int \Psi_{\mathrm{T}} \Psi_{\mathrm{T}} \left(E_{\mathrm{L}} - E_{\mathrm{D}} \right) \Psi_{\mathrm{T}}^{-1} \Psi'_{\mathrm{T}} \, d\mathbf{R}}{\int \Psi_{\mathrm{T}} \Psi_{\mathrm{T}} \, d\mathbf{R}} + \mathcal{O}[(\Psi_{\mathrm{T}} - \phi_0)^2] \, . \tag{43}$$

This expression is readily calculable if one generates configurations distributed according to the pure (ϕ_0^2) and variational (Ψ_T^2) distributions. The approximation is in the Pulay terms, which are smaller in pure than in mixed DMC and, in addition, the approximation in equation (42) is second order compared with the first-order error in equation (39). Equation (42) satisfies the zero variance condition; if Ψ_T and Ψ_T' are exact the variance of the force obtained from equation (42) is zero. Equation (42) has been used to obtain very accurate forces in small molecules [124, 108]. The calculation of accurate DMC forces is still in its infancy, but it does appear that equation (42) offers a very promising way forward.

11. Conclusions

QMC methods provide a framework for computing the properties of correlated quantum systems to high accuracy within polynomial time [2], facilitating applications to large systems. They can be applied to fermions and bosons with arbitrary inter-particle potentials and external fields. These intrinsically parallel methods are ideal for utilising current and next-generation massively parallel computers. Their accuracy, generality and wide applicability suggest that they will play an important role in improving our understanding of the behaviour of large assemblies of quantum particles.

It is believed [125] that a complete solution to the fermion sign problem may be impossible, and any exact fermion method may be exponentially slow on a classical computer. Accurate quantum chemistry techniques such as the "gold standard" coupled cluster with single and double excitations and perturbative triples [CCSD(T)] have been applied with considerable success to correlated electron problems but, although they are also polynomial time algorithms, their cost increases much more rapidly with system size than for QMC methods. DFT methods have proved extremely useful in describing correlated electron systems, but there are many examples where the accuracy of current density functionals has proved wanting. It is important to remember that trial wave functions for QMC calculations could be improved by developing new wave function forms and better optimisation methods, whereas improving approximate DFT methods requires the development of better density functionals, which seems likely to be a much harder problem.

These considerations motivate the development of approximate QMC methods such as those described in this review. Although the basics of the DMC algorithm used by Ceperley and Alder in 1980 [3] have remained unchanged, enormous progress has been made in using more complex trial wave functions and in optimising the many parameters in them. There is every reason to believe that the current high rate of progress will continue for many years to come. Although these QMC methods will remain approximate, it is already clear that sophisticated computer packages [44] such as the CASINO code [45, 98] can deliver highly accurate results.

12. Acknowledgements

We would like to thank all of our collaborators who have contributed so much to our QMC project. Much of this work has been supported by the Engineering and Physical Sciences Research Council (EPSRC) of the UK. NDD acknowledges support from the Leverhulme Trust and Jesus College, Cambridge, and MDT acknowledges support from the Royal Society. Computing resources were provided by the Cambridge High Performance Computing Service.

References

[1] Foulkes W M C, Mitas L, Needs R J and Rajagopal G 2001 *Rev. Mod. Phys.* **73** 33
[2] The scaling of DMC is expected to become exponential in very large systems because of increasing correlation in the data, see Nemec N 2009 *Diffusion Monte Carlo computational cost scales exponentially for large systems* arXiv:0906.0501. VMC does not suffer from this problem.
[3] Ceperley D M and Alder B J 1980 *Phys. Rev. Lett.* **45** 566
[4] Zong F H, Lin C and Ceperley D M 2002 *Phys. Rev.* E **66** 036703
[5] Drummond N D, Radnai Z, Trail J R, Towler M D and Needs R J 2004 *Phys. Rev.* B **69** 085116
[6] Tanatar B and Ceperley D M 1989 *Phys. Rev.* B **39** 5005
[7] Attaccalite C, Moroni S, Gori-Giorgi P and Bachelet G B 2002 *Phys. Rev. Lett.* **88** 256601
[8] Drummond N D and Needs R J 2009 *Phys. Rev. Lett.* **102** 126402
[9] Casulleras J and Boronat J 2000 *Phys. Rev. Lett.* **84** 3121
[10] Holzmann M, Bernu B and Ceperley D M 2006 *Phys. Rev.* B **74** 104510
[11] Carlson J 2007 *Nuclear Physics* A **787** 516
[12] Carlson J, Chang S-Y, Pandharipande V R and Schmidt K E 2003 *Phys. Rev. Lett.* **91** 050401
[13] Astrakharchik G E, Boronat J, Casulleras J and Giorgini S 2004 *Phys. Rev. Lett.* **93** 200404
[14] Carlson J and Reddy S 2008 *Phys. Rev. Lett.* **100** 150403
[15] Healy S B, Filippi C, Kratzer P, Penev E and Scheffler M 2001 *Phys. Rev. Lett.* **87** 016105
[16] Filippi C, Healy S B, Kratzer P, Pehlke E and Scheffler M 2002 *Phys. Rev. Lett.* **89** 166102
[17] Kim Y-H, Zhao Y, Williamson A, Heben M J and Zhang S B 2006 *Phys. Rev. Lett.* **96** 016102
[18] Ghosal A, Guclu A D, Umrigar C J, Ullmo D and Baranger H U 2006 *Nature Physics* **2** 336
[19] Mitas L and Martin R M 1994 *Phys. Rev. Lett.* **72** 2438
[20] Williamson A J, Hood R Q, Needs R J and Rajagopal G 1998 *Phys. Rev.* B **57** 12140
[21] Towler M D, Hood R Q and Needs R J 2000 *Phys. Rev.* B **62** 2330
[22] Needs R J and Towler M D 2003 *Int. J. Mod. Phys. B* **17** 5425
[23] Wagner L and Mitas L 2003 *Chem. Phys. Lett.* **370** 412
[24] Wagner L K and Mitas L 2007 *J. Chem. Phys.* **126** 034105
[25] Williamson A J, Grossman J C, Hood R Q, Puzder A and Galli G 2002 *Phys. Rev. Lett.* **89** 196803
[26] Drummond N D, Williamson A J, Needs R J and Galli G 2005 *Phys. Rev. Lett.* **95** 096801
[27] Leung W-K, Needs R J, Rajagopal G, Itoh S and Ihara S 1999 *Phys. Rev. Lett.* **83** 2351
[28] Hood R Q, Kent P R C, Needs R J and Briddon P R 2003 *Phys. Rev. Lett.* **91** 076403
[29] Alfè D and Gillan M J 2005 *Phys. Rev.* B **71** 220101
[30] Alfè D, Alfredsson M, Brodholt J, Gillan M J, Towler M D and Needs R J 2005 *Phys. Rev.* B **72** 014114
[31] Natoli V, Martin R M and Ceperley D M 1993 *Phys. Rev. Lett.* **70** 1952
[32] Delaney K T, Pierleoni C and Ceperley D M 2006 *Phys. Rev. Lett.* **97** 235702
[33] Maezono R, Ma A, Towler M D and Needs R J 2007 *Phys. Rev. Lett.* **98** 025701
[34] Pozzo M and Alfè D 2008 *Phys. Rev.* B **77** 104103
[35] Manten S and Lüchow A 2001 *J. Chem. Phys.* **115** 5362

[36] Grossman J C 2002 *J. Chem. Phys.* **117** 1434
[37] Aspuru-Guzik A, El Akramine O, Grossman J C, Lester W A Jr 2004 *J. Chem. Phys.* **120** 3049
[38] Gurtubay I G, Drummond N D, Towler M D and Needs R J 2006 *J. Chem. Phys.* **124** 024318
[39] Gurtubay I G and Needs R J 2007 *J. Chem. Phys.* **127** 124306
[40] Hood R Q, Chou M-Y, Williamson A J, Rajagopal G, Needs R J and Foulkes W M C 1997 *Phys. Rev. Lett.* **78** 3350
[41] Hood R Q, Chou M-Y, Williamson A J, Rajagopal G and Needs R J 1998 *Phys. Rev.* B **57** 8972
[42] Nekovee M, Foulkes W M C and Needs R J 2001 *Phys. Rev. Lett.* **87** 036401
[43] Nekovee M, Foulkes W M C and Needs R J 2003 *Phys. Rev.* B **68** 235108
[44] `www.qmcwiki.org/index.php/Research_resources`
[45] Needs R J, Towler M D, Drummond N D and López Ríos P 2009 CASINO version 2.3 User Manual, University of Cambridge, Cambridge, UK
[46] Ortiz G, Ceperley D M and Martin R M 1993 *Phys. Rev. Lett.* **17**, 2777
[47] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *J. Chem. Phys.* **21** 1087
[48] Kalos M H, Colletti L and Pederiva F 2005 *J. Low Temp. Phys.* **138** 747
[49] Anderson J B 1975 *J. Chem. Phys.* **63** 1499
[50] Anderson J B 1976 *J. Chem. Phys.* **65** 4121
[51] Grimm R C and Storer R G 1971 *J. Comput. Phys.* **7** 134
[52] Kalos M H, Levesque D and Verlet L 1974 *Phys. Rev.* A **9** 257
[53] Kalos M H 1962 *Phys. Rev.* **128** 1791
[54] Kalos M H 1967 *J. Comput. Phys.* **2** 257
[55] Ceperley D M and Kalos M H, in *Monte Carlo Methods in Statistical Physics*, 2nd ed., edited by Binder K (Springer, Berlin, Germany, 1986), p. 145
[56] Schmidt K E and Kalos M H, in *Applications of Monte Carlo Methods in Statistical Physics*, 2nd ed., edited by Binder K (Springer, Berlin, Germany, 1987), p. 125
[57] López Ríos P, Ma A, Drummond N D, Towler M D and Needs R J 2006 *Phys. Rev.* E **74** 066701
[58] Kato T 1957 *Comm. Pure Appl. Math.* **10** 151
[59] Drummond N D, Towler M D and Needs R J 2004 *Phys. Rev.* B **70** 235119
[60] López Ríos P and Needs R J *unpublished*
[61] Feynman R P 1954 *Phys. Rev.* **94** 262
[62] Feynman R P and Cohen M 1956 *Phys. Rev.* **102** 1189
[63] Kwon Y, Ceperley D M and Martin R M 1993 *Phys. Rev.* B **48** 12037
[64] Kwon Y, Ceperley D M and Martin R M 1998 *Phys. Rev.* B **58** 6800
[65] Drummond N D, López Ríos P, Ma A, Trail J R, Spink G, Towler M D and Needs R J 2006 *J. Chem. Phys.* **124** 224104
[66] Brown M D, Trail J R, López Ríos P and Needs R J 2007 *J. Chem. Phys.* **126** 224110
[67] Bajdich M, Mitas L, Drobný G, Wagner L K and Schmidt K E 2006 *Phys. Rev. Lett.* **96** 130201
[68] Bajdich M, Mitas L, Wagner L K and Schmidt K E 2008 *Phys. Rev.* B **77** 115112
[69] Lüchow A, Petz R and Scott T C 2007 *J. Chem. Phys.* **126** 144110
[70] Reboredo F A, Hood R Q and Kent P R C 2009 *Phys. Rev.* B **79** 195117
[71] Conroy H 1964 *J. Chem. Phys.* **41** 1331
[72] Umrigar C J, Wilson K G and Wilkins J W 1998 *Phys. Rev. Lett.* **60** 1719
[73] Dewing M and Ceperley D M 2002 *Methods for Coupled Electronic-Ionic Monte Carlo* in *Recent Advances in Quantum Monte Carlo Methods, Part II*, ed by Lester W A, Rothstein S M and Tanaka S (World Scientific, Singapore)
[74] Drummond N D and Needs R J 2005 *Phys. Rev.* B **72** 085124
[75] Riley K E and Anderson J B 2003 *Mol. Phys.* **101** 3129
[76] Nightingale M P and Melik-Alaverdian V 2001 *Phys. Rev. Lett.* **87** 043401
[77] Umrigar C J, Toulouse J, Filippi C, Sorella S and Hennig R G 2007 *Phys. Rev. Lett.* **98** 110201
[78] Toulouse J and Umrigar C J 2007 *J. Chem. Phys.* **126** 084102

[79] Rajagopal G, Needs R J, Kenny S, Foulkes W M C and James A 1994 *Phys. Rev. Lett.* **73** 1959
[80] Rajagopal G, Needs R J, James A, Kenny S D and Foulkes W M C 1995 *Phys. Rev.* B **51** 10591
[81] Lin C, Zong F H and Ceperley D M 2001 *Phys. Rev.* E **64** 016702
[82] Chiesa S, Ceperley D M, Martin R M and Holzmann M 2006 *Phys. Rev. Lett.* **97** 076404
[83] Drummond N D, Needs R J, Sorouri A and Foulkes W M C 2008 *Phys. Rev.* B **78** 125106
[84] Fraser L M, Foulkes W M C, Rajagopal G, Needs R J, Kenny S D and Williamson A J 1996 *Phys. Rev.* B **53** 1814
[85] Williamson A J, Rajagopal G, Needs R J, Fraser L M, Foulkes W M C, Wang Y and Chou M-Y 1997 *Phys. Rev.* B **55** 4851
[86] Kent P R C, Hood R Q, Williamson A J, Needs R J, Foulkes W M C and Rajagopal G 1999 *Phys. Rev.* B **59** 1917
[87] Kwee H, Zhang S and Krakauer H 2008 *Phys. Rev. Lett.* **100** 126404
[88] Ceperley D M 1986 *J. Stat. Phys.* **43** 815
[89] Ma A, Drummond N D, Towler M D and Needs R J 2005 *Phys. Rev.* E **71** 066704
[90] Fahy S, Wang X W and Louie S G 1998 *Phys. Rev. Lett.* **61** 1631
[91] Fahy S, Wang X W and Louie S G 1990 *Phys. Rev.* B **42** 3503
[92] Mitáš L, Shirley E L and Ceperley D M 1991 *J. Chem. Phys.* **95** 3467
[93] Casula M, Filippi C and Sorella S 2005 *Phys. Rev. Lett.* **95** 100201
[94] Casula M 2006 *Phys. Rev.* B **74** 161102
[95] Greeff C W and Lester W A Jr 1998 *J. Chem. Phys.* **109** 1607
[96] Trail J R and Needs R J 2005 *J. Chem. Phys.* **122** 014112
[97] Trail J R and Needs R J 2005 *J. Chem. Phys.* **122** 174109
[98] www.tcm.phy.cam.ac.uk/~mdt26/casino2_pseudopotentials.html
[99] Burkatzki M, Filippi C and Dolg M 2007 *J. Chem. Phys.* **126** 234105; *ibid.* 2008 **129** 164115
[100] Trail J R and Needs R J 2008 *J. Chem. Phys.* **128** 204103
[101] Santra B, Michaelides A, Fuchs M, Tkatchenko A, Filippi C and Scheffler M 2008 *J. Chem. Phys.* **129** 194111
[102] Foulkes W M C, Hood R Q and Needs R J 1999 *Phys. Rev.* B **60** 4558
[103] Porter A R, Al-Mushadani O K, Towler M D and Needs R J 2001 *J. Chem. Phys.* **114** 7795
[104] Porter A R, Towler M D and Needs R J 2001 *Phys. Rev.* B **64** 035320
[105] Bande A, Lüchow A, Della Sala F and Görling G 2006 *J. Chem. Phys.* **124** 114114
[106] Ceperley D M and Bernu B 1988 *J. Chem. Phys.* **89** 6316
[107] Umrigar C J, Nightingale M P and Runge K J 1993 *J. Chem. Phys.* **99** 2865
[108] Badinski A, Haynes P D, Trail J R and Needs R J 2009 to appear in *J. Phys.: Condensed Matter*
[109] Flyvbjerg H and Petersen H G 1989 *J. Chem. Phys.* **91** 461
[110] Liu K S, Kalos M H and Chester G V 1974 *Phys. Rev.* A **10** 303
[111] Barnett R N, Reynolds P J and Lester W A Jr 1991 *J. Comput. Phys.* **96** 258
[112] Baroni S and Moroni S 1999 *Phys. Rev. Lett.* **82** 4745
[113] Drummond N D and Needs R J 2009 *Phys. Rev.* B **79** 085414
[114] Huang K C, Needs R J and Rajagopal G 2000 *J. Chem. Phys.* **112** 4419
[115] Schautz F and Flad H-J 2000 *J. Chem. Phys.* **112** 4421
[116] Badinski A, Haynes P D and Needs R J 2008 *Phys. Rev.* B **77** 085111
[117] Reynolds P J, Barnett R N, Hammond B L, Grimes R M and Lester W A Jr 1986 *Int. J. Quant. Chem.* **29** 589
[118] Assaraf R and Caffarel M 1999 *Phys. Rev. Lett.* **83** 4682
[119] Casalegno M, Mella M and Rappe A M 2003 *J. Chem. Phys.* **118** 7193
[120] Assaraf R and Caffarel M 2003 *J. Chem. Phys.* **119** 10536
[121] Lee M W, Mella M and Rappe A M 2005 *J. Chem. Phys.* **122** 244103
[122] Badinski A and Needs R J 2007 *Phys. Rev.* E **76** 036707
[123] Badinski A and Needs R J 2008 *Phys. Rev.* B **78** 035134
[124] Badinski A, Trail J R and Needs R J 2008 *J. Chem. Phys.* **129** 224101

[125] Troyer M and Wiese U-J 2005 *Phys. Rev. Lett.* **94** 170201