

Review Article

The quantum Monte Carlo method

M. D. Towler*

TCM group, Cavendish Laboratory, Cambridge University, J. J. Thomson Ave.,
Cambridge CB3 0HE, UK

Received 21 March 2006, revised 6 July 2006, accepted 10 July 2006

Published online 23 August 2006

PACS 02.70.Ss, 31.10.+z, 31.15.Ar, 31.25.-v, 71.15.-m, 71.15.Nc

Quantum Monte Carlo is an important and complementary alternative to density functional theory when performing computational electronic structure calculations in which high accuracy is required. The method has many attractive features for probing the electronic structure of real atoms, molecules and solids. In particular, it is a genuine many-body theory with a natural and explicit description of electron correlation which gives consistent, highly-accurate results while at the same time exhibiting favourable (cubic or better) scaling of computational cost with system size. This article is intended to provide a brief and hopefully accessible review of some relevant aspects of quantum Monte Carlo together with an outline of our implementation of it in the Cambridge computer code ‘CASINO’ [1, 2].

© 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Contents

- 1 Introduction**
 - 2 QMC algorithms**
 - 2.1 Variational Monte Carlo
 - 2.1.1 Basics
 - 2.1.2 The form of the trial wave function
 - 2.1.3 Optimization of trial wave functions
 - 2.1.4 VMC conclusions
 - 2.2 Diffusion Monte Carlo
 - 3 Miscellaneous issues**
 - 3.1 More about trial wave functions
 - 3.2 Basics set expansions: how to represent the orbitals?
 - 3.3 Pseudopotentials
 - 4 Recent developments**
 - 4.1 All-electron QMC calculations for heavier atoms
 - 4.2 Improved scaling algorithms
 - 5 Applications**
 - 6 The CASINO code**
- References**

* e-mail: mdt26@cam.ac.uk, Phone: +44 (0)1223 337378, Fax: +44 (0)1223 337356

1 Introduction

The continuum Quantum Monte Carlo (QMC) method has been developed to calculate the properties of assemblies of interacting quantum particles. It is generally capable of doing so with great accuracy. The various different techniques which lie within its scope have in common the use of *random sampling*, and this is used because it represents by far the most efficient way to do numerical integrations of expressions involving wave functions in many dimensions. Two particular variants of QMC are in relatively common use, namely variational Monte Carlo (VMC) and diffusion Monte Carlo (DMC) [3, 4]; here we give a brief introduction to both. As we shall see, VMC is simple in concept and is designed just to sample a given trial wave function and calculate the expectation value of the Hamiltonian using Monte Carlo numerical integration. This is more useful than it sounds since the method is variational and thus we can to some extent *optimize* suitably parametrized explicitly correlated wave functions using standard techniques. DMC is one of a class of so-called ‘projector’ methods which attempt the much more difficult job of simultaneously creating and sampling the unknown exact ground state wave function. Other variants, including those aimed at expanding the scope of the method to finite temperature such as path integral Monte Carlo (PIMC) [5, 6], or those designed to find the exact non-relativistic energy overcoming the small fixed-node approximation made in DMC (such as fermion Monte Carlo (FMC) [7–9]) will not be discussed in any detail here. The interested reader is invited to consult the literature for more detailed discussions (the extensive bibliography in Ref. [4] is a good place to start).

In its early days QMC was perhaps best known for its application to the homogeneous electron gas by Ceperley and Alder [10]. The results of these calculations were generally understood to be extremely accurate and were used to develop accurate parametrizations of the local density approximation to density functional theory (DFT) in the early 1980s. However, it is of course perfectly possible to apply the method to real systems with atoms, and for small molecules containing helium and hydrogen QMC gives total energies with a remarkable accuracy greater than 0.01 kcal/mole ($\approx 1.5 \times 10^{-5}$ Ha or 4×10^{-4} eV). In one well-known QMC study of the $\text{H} + \text{H}_2 \rightarrow \text{H}_2 + \text{H}$ potential energy surface tens of thousands of points with accuracies close to this value were computed [11]. Despite such capabilities the technology of QMC is neither mature nor particularly widely used; its routine application to arbitrary finite and periodic systems, particularly those containing heavier atoms, has long been just out of reach and there are still many open methodological and algorithmic problems to interest the computational electronic structure theorist. The situation is clearly changing however, and it ought now to be a matter of routine for people to perform accurate QMC calculations of even quite large systems, albeit starting from wave functions generated from one-electron molecular orbital or band theory. Systems and problems for which an accurate determination of the total energy actually matters, and for which DFT (for example) is not sufficiently accurate, are likely more numerous than is generally believed. To this end, our group in Cambridge University’s Cavendish Laboratory has spent a considerable number of years developing a general-purpose QMC computer program - CASINO [1, 2]. This code is capable of performing both variational and diffusion Monte Carlo calculations on a wide variety of systems, which may be of finite extent (atoms or molecules) or may obey periodic boundary conditions in one, two or three dimensions, modelling what one might respectively call polymers, slabs (or surfaces) and crystalline solids. The code may also be used to study situations where there is no external potential (such as the homogeneous electron gas or the Wigner crystal) and can treat generalized ‘quantum particles’, i.e. fermions or bosons with user-defined charge and mass tensor. We shall describe CASINO in more detail presently.

One of the more attractive features of QMC is the scaling behaviour of the necessary computational effort with system size. This is favourable enough that we can continue to apply the method to systems as large as are treated in conventional DFT, albeit with a considerably bigger pre-factor and thus probably not on the same computers. In fact QMC seems currently to be the most accurate method available for medium-sized and large systems. Other correlated wave function methods based on quantum chemistry’s ‘standard model’ of multideterminant expansions – such as configuration interaction or high-order coupled cluster theory – are capable of similar accuracy for systems containing a few electrons, but as the size of the molecule is increased they quickly become too expensive. Standard QMC calculations

scale as the third power of the system size (the same as DFT) and are capable of treating solids and other periodic systems as well as molecules. The largest calculations done to date on the more expensive periodic systems using the regular algorithm include almost 2000 electrons per cell in the three-dimensional electron gas [12], 1732 electrons (432 atoms) per cell in crystalline silicon [13], and 1024 electrons (128 atoms) per cell in antiferromagnetic nickel oxide [14]. Furthermore the natural observation has been made that provided localized molecular or crystalline orbitals are used in constructing the QMC trial wave function, and provided these orbitals are expanded in a localized basis set, then the scaling of the basic algorithm can be substantially improved over implementations using delocalized functions (such as Bloch orbitals and plane-wave basis sets). This has led to claims of linear scaling QMC in the literature [15, 16], although the definition of 'linear scaling' in this context is controversial. An improved scaling capability based on such ideas, to be discussed in more detail in Section 4.2, has been implemented in the CASINO program and has considerably extended the range of problems that may be studied.

Before we go further, it will be useful to list some other favourable properties of the method:

- For most practical purposes the 'basis set problem' is essentially absent in DMC. Errors due to the use of a finite basis set are expected to be small since the many-electron wave function is not represented directly in terms of a basis set, but rather by the average distribution of an ensemble of particles evolving in (imaginary) time. The sole purpose of the basis set that is in fact employed in DMC is to represent a guiding function required for importance sampling. The final DMC energy depends only weakly on the nodal surface of this guiding function (i.e., the set of points in configuration space at which the function is zero).

- The QMC algorithm is intrinsically parallel and Monte Carlo codes are thus easily adapted to parallel computers and scale linearly with the number of processors. There are no memory or disk bottlenecks even for relatively large systems.

- We can use many-electron wave functions with explicit dependence on interparticle distances and no need for analytic integrability.

- We can calculate ground states, some excited states, chemical reaction barriers and other properties within a single unified framework. The method is size-consistent and variational.

One may ask why one should formulate a method based on the many-electron wave function when so much stress is normally placed on reducing the number of variables in the quantum problem (by using, e.g., density, Green's functions, density matrices or other quantities which depend on fewer independent variables). The main point is that the many-electron wave function satisfies a rather well-known fundamental equation [17]:

$$\hat{H}\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = E\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N). \quad (1)$$

The price we pay for reformulating the problem in terms of the density is that we no longer know the exact equation satisfied by the density. In DFT, the complicated many-body problem is effectively relocated into the definition of the exchange-correlation functional, whose mathematical expression is not currently known and unlikely ever to be known exactly. The inevitable approximations to this quantity substantially reduce the attainable accuracy.

The quantum chemistry community has invested a great deal of effort into calculating accurate approximate solutions to the full many-electron Schrödinger equation for atoms and molecules, but as condensed matter physicists we are also interested in doing this for solids and other condensed phases. So what are our chances of solving the full many-electron Schrödinger equation in an infinite solid? Standard widely-used solid-state texts often deny the possibility of doing this directly in any meaningful way for large crystalline systems. To take a particular example, the well-known textbook by Ashcroft and Mermin [18] states that, 'one has no hope of solving an equation such as [Eq. (1)]' and one must reformulate the problem in such a way as 'to make the one-electron equations least unreasonable'. However the key simplifying physical idea to allow one to use, for example, QMC in crystalline solids is not the use of one-electron orbitals but simply the imposition of periodic boundary conditions. One can then

have an explicitly correlated many-body wave function (i.e., with explicit dependence on the interparticle separations), in a box, embedded in an infinite number of copies of itself. One can then visualize the ‘particles’ sampling the many-body wave function as a periodic array of electrons moving in tandem with each other rather than as individual electrons. It is clear that in order for this to have any chance of being an accurate approximation the range of the electron–electron pair correlation function must be substantially shorter than the repeat distance and the box must be large enough so that the forces on the particles within it are very close to those in the bulk. If not, then we may get substantial ‘finite-size errors’.

This problem is analogous to but not quite the same as the problem of representing an infinite system in DFT calculations. In that case Bloch’s theorem is generally used in the extrapolation to infinite system size so that the problem of calculating an infinite number of one-electron states reduces to the calculation of a finite number (equal to the number of electrons in the primitive cell) of states at an infinite number of k -points in reciprocal space. As the band energies vary continuously and relatively slowly with k , the k -space may thus be ‘sampled’ and if this is done efficiently the calculated energy per cell approaches that in the infinite system. The situation in QMC is a little different since the explicit correlation between electrons means that the problem cannot be reduced to the primitive cell; a one-electron wave function on a $2 \times 2 \times 2$ k -point grid corresponds to a many-electron wave function for a $2 \times 2 \times 2$ supercell in real space. There is a ‘many-body Bloch theorem’ expressing the invariance of the Hamiltonian under translations of all electrons by a primitive lattice vector or of a single electron by a supercell lattice vector [19], and thus there are two k -vectors associated with the periodic many-body wave function. The error analogous to inadequate Brillouin zone sampling might be made smaller either by increasing the size of the simulation cell or by choosing the k -values using ‘special k -point’ techniques [20]. An additional type of finite-size error arises in periodic QMC calculations (though not in DFT) when calculating interactions between particles with long-range Coulomb sums. The difference is that in QMC we deal with instantaneous positions of electron configurations, rather than with the interaction of averaged densities. When using the standard Ewald formulation [21, 22] for these long-range summations, the choice of boundary conditions (equivalent to embedding your supposed hunk of crystal in a perfect conductor) leads to an effective depolarization field which cancels the field due to your notional surface charges. As all periodic copies of the simulation cell contain, for example, the same net dipole due to the random arrangement of electrons with respect to nuclei the interaction of these dipoles (and higher multipoles) with the depolarization field gives rise to ‘Coulomb finite size errors’. These can be substantially reduced by using special techniques [23].

A few years ago in his Nobel prize-winning address Walter Kohn suggested that the many-electron wave function is not a legitimate scientific concept when more than about a thousand particles are involved [24]. It would be pretty disastrous if this meant that QMC could not be used for large systems, so let us try to understand what he means. The main idea behind his statement is that the overlap of any approximate wave function with the exact one will tend exponentially to zero as the number of particles increases unless one uses a wave function in which the number of parameters increases exponentially with system size, and that clearly such a wave function would not be computable for large systems. This is indeed true, and one may easily verify it by calculating the overlap integral directly using VMC [25]. One does not need the exact wave function itself to perform such a calculation, since Kohn’s argument is based solely on the high-dimensionality of the overlap integrals rather than, say, the explicit cancellation of positive and negative regions. One can thus evaluate the overlap between, say, a single-determinant wave function and the same single-determinant function multiplied by a Jastrow correlation function. Even though these objects share the same nodal surface, we still expect to see and indeed do see the result that Kohn predicts. Luckily his objection seems not to be relevant to the sort of QMC calculations discussed here. Certainly the successful DMC calculations of systems containing up to 2000 electrons mentioned earlier suggest as much, but as Kohn himself points out, we are interested in quantities such as the total energy, which can be accurate even when the overlap with the exact wave function goes to zero. To get the energy right it is required only that relatively low-order correlation functions (such as the pair-correlation function) are well-described and QMC seems to manage this very well. Kohn’s arguments were used to motivate density functional theory, but it is possible to argue that, within the standard

Kohn–Sham formulation, DFT suffers from exactly the same overlap ‘catastrophe’. For a large system the overlap of the determinant of Kohn–Sham orbitals with the exact one will go to zero because of the inevitable numerical inaccuracies and the approximations to the exchange-energy functional. Fortunately, as I have suggested, the overlap catastrophe seems to be irrelevant to actually calculating most quantities of interest.

To understand how accurate the total energies must be we note that the main goal is to calculate the energy difference between two arrangements of a set of atoms. The desired result might be the energy required to form a defect, or the energy barrier to some process, or whatever. All electronic structure methods for large systems rely on a cancellation of errors in energy differences. For such error cancellations to occur we require that the error in the energy per atom is proportional to the number of atoms. If this condition was not satisfied then, for example, the cohesive energy would not have a well-defined limit for large systems. Many VMC calculations have demonstrated that the commonly-used forms of many-body wave function lead to errors which are proportional to the number of atoms, and typically give between 70 and 90% of the correlation energy independent of system size. In DMC the error is also proportional to the number of atoms but is capable of recovering up to 100% of the correlation energy in favourable cases. Additional requirements on QMC algorithms are that the number of parameters in the trial wave function must not increase too rapidly with system size and that the wave function be easily computable. Fortunately the number of parameters in a typical QMC trial wave function increases only linearly, or at worst quadratically, with system size and the function can be evaluated in a time which rises as a low power of the system size.

2 QMC algorithms

In this section, we shall look at the basic ideas and algorithms underlying VMC and DMC.

2.1 Variational Monte Carlo

2.1.1 Basics

With variational methods we must ‘guess’ an appropriate many-electron wave function which is then used to calculate the energy as the expectation value of the Hamiltonian operator. In general this wave function will depend on a set of parameters $\{\alpha\}$ which can be varied to optimize the function and minimize either the energy or the statistical variance. The energy thus obtained is an upper bound to the true ground state energy,

$$\frac{\langle \Psi_T(\{\alpha\}) | \hat{H} | \Psi_T(\{\alpha\}) \rangle}{\langle \Psi_T(\{\alpha\}) | \Psi_T(\{\alpha\}) \rangle} = E(\{\alpha\}) \geq E_0. \quad (2)$$

The expectation value of the Hamiltonian \hat{H} with respect to the trial wave function Ψ_T can be written as

$$\langle \hat{H} \rangle = \frac{\int E_L(\mathbf{R}) \Psi_T^2(\mathbf{R}) d\mathbf{R}}{\int \Psi_T^2(\mathbf{R}) d\mathbf{R}}, \quad (3)$$

where \mathbf{R} is a $3N$ dimensional vector giving the coordinates $(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ of the N particles in the system, and $E_L(\mathbf{R}) = \frac{\hat{H}(\mathbf{R}) \Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})}$ is known as the *local energy*.

We can evaluate this expectation value by using the Metropolis algorithm [26] to generate a sequence of configurations \mathbf{R} distributed according to $\Psi_T^2(\mathbf{R})$ and averaging the corresponding local energies,

$$\langle \hat{H} \rangle = \frac{1}{M} \sum_{i=1}^M E_L(\mathbf{R}_i) = \frac{1}{M} \sum_{i=1}^M \frac{\hat{H} \Psi_T(\mathbf{R}_i)}{\Psi_T(\mathbf{R}_i)}. \quad (4)$$

The question of whether or not we get the right answer with this approach is just one of *complexity*; can we create a wave function with enough variational freedom so that the energy approaches the exact (non-relativistic) ground state energy? The answer in general is no. There is no systematic way in which one can improve the wave function until the correct answer is reached, and in general, we shouldn't normally expect to recover much more than 80–90% of the correlation energy in this way (although one can in fact do much better than this for specific individual systems). As we shall see, the final 20% or so can be calculated by feeding the VMC wave function into a projector method such as DMC. This, to my mind, is the main use of VMC and in our laboratory we rarely use it as a method in its own right when performing calculations. With this attitude, it is not generally necessary to kill oneself optimizing wave functions in order to recover an extra 1% of the correlation energy with VMC – it is better to use DMC and let the computer do the work for you. Although the efficiency of the DMC calculations is greatly increased with more accurate trial functions, the final DMC energy does not in principle depend on that part of the wave function that we generally optimize.

2.1.2 The form of the trial wave function

For VMC however it is clear that the choice of the trial function is particularly important as it directly determines the accuracy of the calculation; the answer will approach the true energy from above as we use better and better wave functions. Something else to consider is the 'zero variance principle'. As the trial function approaches an exact eigenstate the local energy $\hat{H}\Psi/\Psi$ approaches a constant, E , everywhere in configuration space (see the Schrödinger equation again!) and hence the variance approaches zero. Through its direct influence on the variance of the energy the accuracy of the trial wave function thus determines the amount of computation required to achieve a specified accuracy. When optimizing wave functions, one can therefore choose to use energy or variance as the objective function to be minimized.

The fact that arbitrary wave function forms can be used is one of the defining characteristics of QMC. We do not need to be able to integrate the wave function analytically as is done for example in quantum chemistry methods with Gaussian basis functions. We just need to be able to *evaluate* it at a point in the configuration space i.e. if the electrons and nuclei have certain fixed positions in space, what is the value of the wave function? This being the case, we can use correlated wave functions which depend explicitly on the distances between particles.

The most commonly-used functional form is known as the Slater–Jastrow wave function [27]. This consists of a single Slater determinant (or sometimes a linear combination of a small number of them) multiplied by a positive-definite Jastrow correlation function which is symmetric in the electron coordinates and depends on the inter-particle distances. The Jastrow factor allows efficient inclusion of both long and short range correlation effects. As we shall see however, the final DMC answer depends only on the nodal surface of the wave function and this cannot be affected by the nodeless Jastrow. In DMC it serves mainly to decrease the amount of computer time required to achieve a given statistical error bar and to improve the stability of the algorithm.

The basic functional form of the Slater–Jastrow function is

$$\Psi(\mathbf{X}) = e^{J(\mathbf{X})} \sum_n c_n D_n(\mathbf{X}), \quad (5)$$

where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ and $\mathbf{x}_i = \{\mathbf{r}_i, \sigma_i\}$ denotes the space-spin coordinates of electron i , $e^{J(\mathbf{X})}$ is the Jastrow factor, the c_n are coefficients, and the $D_n(\mathbf{X})$ are Slater determinants of single-particle orbitals,

$$D(\mathbf{X}) = \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_1(\mathbf{x}_2) & \cdots & \psi_1(\mathbf{x}_N) \\ \psi_2(\mathbf{x}_1) & \psi_2(\mathbf{x}_2) & \cdots & \psi_2(\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_N(\mathbf{x}_1) & \psi_N(\mathbf{x}_2) & \cdots & \psi_N(\mathbf{x}_N) \end{vmatrix}. \quad (6)$$

The orbitals in the determinants are often obtained from self-consistent DFT or Hartree–Fock calculations and are assumed to be products of spatial and spin factors,

$$\psi_\alpha(\mathbf{x}) = \psi_\alpha(\mathbf{r}) \delta_{\sigma,\sigma_\alpha}. \quad (7)$$

Here $\delta_{\sigma,\sigma_\alpha} = 1$ if $\sigma = \sigma_\alpha$ and zero otherwise. If the determinant contains N_\uparrow orbitals with $\sigma_\alpha = \uparrow$ and $N_\downarrow = N - N_\uparrow$ with $\sigma_\alpha = \downarrow$, it is an eigenfunction of \hat{S}_z with eigenvalue $(N_\uparrow - N_\downarrow)/2$. To avoid having to sum over spin variables in QMC calculations, one generally replaces the determinants D_n by products of separate up- and down-spin determinants,

$$\Psi(\mathbf{R}) = e^{J(\mathbf{R})} \sum_n c_n D_n^\uparrow(\mathbf{r}_1, \dots, \mathbf{r}_{N_\uparrow}) D_n^\downarrow(\mathbf{r}_{N_\uparrow+1}, \dots, \mathbf{r}_N), \quad (8)$$

where $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ denotes the spatial coordinates of all the electrons. This function is not antisymmetric under exchange of electrons with opposite spins but it can be shown that it gives the same expectation value as $\Psi(\mathbf{X})$ for any spin-independent operator. Note that the use of wave function forms in QMC which allow one to treat non-collinear spin arrangements and the resultant vector magnetization density is an interesting open problem, and we are currently working on developing such an algorithm [28].

The full Jastrow function that we typically use in CASINO contain one- and two-electron terms and may be inhomogeneous, i.e., depend on the distances of the electrons from the nuclei. The exact functional form is quite complicated and there is no need to go into all the details here (for the curious, they may be found in Ref. [29]). Essentially our Jastrow consists of separate electron–electron (u), electron–nucleus (χ_i), and electron–electron–nucleus (f_i) terms which are expanded in polynomials and are forced to go to zero at some cutoff radii (as they must do in periodic systems). One can get a feel for this from a much simpler one-parameter Jastrow function that might be used for a homogeneous system such as the electron gas:

$$e^{J(\mathbf{R})} \quad \text{with} \quad J(\mathbf{R}) = -\sum_{i>j} u_{\sigma_i,\sigma_j}(r_{ij}) \quad \text{and} \quad u_{\sigma_i,\sigma_j}(r_{ij}) = \frac{A}{r_{ij}} (1 - e^{-r_{ij}/F_{\sigma_i,\sigma_j}}). \quad (9)$$

Here r_{ij} is the distance between electrons i and j , and F is chosen so that the electron–electron cusp conditions are obeyed i.e. $F_{\uparrow\uparrow} = \sqrt{2A}$ and $F_{\uparrow\downarrow} = \sqrt{A}$. The value of A could be optimized using, for example, variance minimization. In the full inhomogeneous Jastrow we generally optimize the coefficients of the various polynomial expansions (which appear linearly in the Jastrow factor) and the cutoff radii of the various terms (which are non-linear). The linearity or otherwise of the various terms clearly has a bearing on their ease of optimization, a subject to which we now turn.

2.1.3 Optimization of trial wave functions

The optimization of the wave function in QMC is clearly a critical step. In addition to the various Jastrow parameters mentioned in the previous section, the CASINO code allows optimization of the coefficients of the determinants of a multi-determinant wave function, various parameters in specialized wave functions used e.g. in electron–hole phases, and even the orbitals in the Slater determinants themselves (in the latter case only for atoms). So clearly the parameters appear in many different contexts, they need to be minimized in the presence of noise, and there can be many of them. This makes the optimization a complicated task in general. Directly optimizing the orbitals in the presence of the Jastrow factor is generally thought to be a good thing, since this in some sense optimizes the nodal surface and in so doing allows improvement of the DMC energy. The best way to do this in systems containing more than one atom remains an open problem however, though some progress has been made [30, 31].

There are many approaches to wave function optimization, but as far as the current version of CASINO is concerned this is achieved by minimizing the variance of the energy,

$$\sigma_E^2(\alpha) = \frac{\int \Psi^2(\alpha) [E_L(\alpha) - E_V(\alpha)]^2 d\mathbf{R}}{\int \Psi^2(\alpha) d\mathbf{R}}, \quad (10)$$

with respect to the set of parameters α . E_V in this expression is the variational energy. There is no reason why one may not optimize the energy directly, and indeed it is generally believed that wave functions corresponding to the minimum energy have more desirable properties. There are however a number of reasons why variance minimization has historically been generally preferred to energy minimization (beyond the trivial fact that the variance has a known lower bound of zero). The most important of these is simply that it has proved easier to design robust, numerically-stable algorithms to minimize the variance than it has for the energy [32, 33]. This is particularly so in large systems.

Beginning with an initial set of parameters α_0 (zeroing polynomial coefficients is usually sufficient), minimization of σ_E^2 is generally carried out via a correlated-sampling approach. First of all a set of some thousands of ‘configurations’ distributed according to $\Psi^2(\alpha_0)$ is generated. A configuration in this sense is just a ‘snapshot’ of the system taken during a VMC run and physically consists of the current electron positions and associated interaction energies written on a line of a file. We then use this information to calculate the objective function – in this case the variance – and proceed to minimize it by varying the parameters. The variance $\sigma_E^2(\alpha)$ is given by the following integral, and this may be approximated by summing over a set of fixed configurations, with variations in the parameters allowed for through the use of weights w :

$$\sigma_E^2(\alpha) = \frac{\int \Psi^2(\alpha_0) w(\alpha) [E_L(\alpha) - E_V(\alpha)]^2 d\mathbf{R}}{\int \Psi^2(\alpha_0) w(\alpha) d\mathbf{R}}, \quad (11)$$

where

$$E_V(\alpha) = \frac{\int \Psi^2(\alpha_0) w(\alpha) E_L(\alpha) d\mathbf{R}}{\int \Psi^2(\alpha_0) w(\alpha) d\mathbf{R}}. \quad (12)$$

The integrals contain weighting factors, $w(\alpha)$, given by

$$w(\alpha) = \frac{\Psi^2(\alpha)}{\Psi^2(\alpha_0)}. \quad (13)$$

The parameters α are then adjusted until $\sigma_E^2(\alpha)$ is minimized. This may be done using a standard algorithm which does an unconstrained minimization (without requiring derivatives) of a sum of m squares of functions which contain n variables, where $m \geq n$.

Note that the point of using the weights here is that we do not have to regenerate the set of configurations every time the parameter values are changed. However, having generated a new set of parameters with this algorithm, we can then carry out a second configuration generation run with these new, more accurate parameters followed by a second optimization, and so on. Generally very few such ‘cycles’ are required before the true minimum is approached.

Thus far we have described the optimization of what is known as the *reweighted variance*. In the limit of perfect sampling, the reweighted variance is equal to the actual variance, and is therefore independent of the configuration distribution, so that the optimized parameters would not change over successive cycles. There is a major problem with it however, and this arises from the fact that the weights may vary rapidly as the parameters change especially for large systems. This can lead to severe instabilities in the numerical procedure. Somewhat surprisingly perhaps, it usually turns out that the best solution to this is to do without the weights at all, in which case we are minimizing the *unreweighted variance*. This turns out to have a number of advantages beyond improving the numerical stability. The self-consistent minimum in the unreweighted variance almost always turns out to give lower energies than the minimum in the reweighted variance. Furthermore our group has recently demonstrated a new scheme which hugely speeds up the optimization of parameters that occur linearly in the Jastrow, which are the most important in the wave functions that we use. The basis of this is that the unreweighted variance can be written analytically as a quartic function of the linear parameters. This function usually has a single minimum in the parameter

space, and as the minima of multidimensional quartic functions may be found very rapidly, the optimization is extraordinarily efficient compared to the regular algorithm. The scheme is described in Ref. [34].

The whole procedure of variance minimization can be, and in CASINO is, thoroughly automated and providing a systematic approach is adopted, optimizing VMC wave functions is not the complicated time-consuming business it once was. This is particularly the case if one requires the optimized wave function only for input into a DMC calculation, in which case one need not be overly concerned with lowering the VMC energy as much as possible.

2.1.4 VMC conclusions

Although VMC can be quite powerful when applied to the right problem, the necessity of guessing the functional form of the trial function limits its accuracy and there is no known way to *systematically* improve it all the way to the exact non-relativistic limit. In practice therefore, the main use of VMC is in providing the optimized trial wave function required as an importance sampling function by the much more powerful DMC technique, which we now describe.

2.2 Diffusion Monte Carlo

Let us imagine that we are ignorant, or have simply not been paying attention in our quantum mechanics class, and that we believe that the wave function of the hydrogen atom looks like a square box centred on the nucleus. If we tried to calculate the expectation value of the Hamiltonian using VMC we would obtain an energy which was substantially in error. What DMC can do, in essence, is to correct the functional form of the guessed square box wave function so that it looks like the correct exponentially-decaying one before calculating the expectation value. This is a nice trick if you can do it, particularly when we have very little practical idea of what the exact ground state wave function looks like (that is, almost always). As one might expect, the algorithm is necessarily rather more involved than that for VMC.

Essentially then, the DMC method is a stochastic projector method for evolving the imaginary-time Schrödinger equation (which you can get by taking the regular time-dependent equation and replacing the time variable t with $\tau = it$):

$$-\frac{\partial \Psi(\mathbf{R}, \tau)}{\partial \tau} = -\frac{1}{2} \nabla^2 \Psi(\mathbf{R}, \tau) + (V(\mathbf{R}) - E_T) \Psi(\mathbf{R}, \tau). \quad (14)$$

Here the real variable τ measures the progress in imaginary time and \mathbf{R} is a $3N$ -dimensional vector of the positions of the N electrons. $V(\mathbf{R})$ is the potential energy operator, E_T is an energy offset which only affects the normalization of the wave function Ψ , and $\nabla = (\nabla_1, \nabla_2, \dots, \nabla_N)$ is the $3N$ -dimensional gradient operator.

This equation has the property that an initial starting state $\Psi(\mathbf{R}, \tau = 0)$ decays towards the ground state wave function. In DMC the time evolution of Eq. (14) may be followed using a stochastic technique in which $\Psi(\mathbf{R}, \tau)$ is represented by an ensemble of $3N$ -dimensional electron configurations (sometimes called ‘walkers’), $\{\mathbf{R}_i\}$. The time evolution of these configurations is governed by the Green’s function of Eq. (14). Within the short time approximation the Green’s function separates into functions representing two processes: random diffusive jumps of the configurations arising from the kinetic term and creation/destruction of configurations arising from the potential energy term.

Unfortunately this simple algorithm suffers from two very serious drawbacks. The first is that we have implicitly assumed that Ψ is a probability distribution, even though its fermionic nature means that it must have positive and negative parts. The second problem is less fundamental but in practice very severe. The required rate of removing or adding configurations diverges when the potential energy diverges, which occurs whenever two electrons or an electron and a nucleus are coincident. This leads to extremely poor statistical behaviour.

These problems are dealt with at a single stroke by introducing an importance sampling transformation. If we consider the *mixed distribution* $f = \Psi_T \Psi$, where Ψ_T is known as the trial or guiding wave function, and substitute into Eq. (14) we obtain

$$-\frac{\partial f(\mathbf{R}, \tau)}{\partial \tau} = -\frac{1}{2} \nabla^2 f(\mathbf{R}, \tau) + \nabla[\mathbf{v}_D(\mathbf{R}) f(\mathbf{R}, \tau)] + (E_L(\mathbf{R}) - E_T) f(\mathbf{R}, \tau), \quad (15)$$

where $\mathbf{v}_D(\mathbf{R})$ is the $3N$ -dimensional *drift velocity* defined by

$$\mathbf{v}_D(\mathbf{R}) = \nabla \ln |\Psi_T(\mathbf{R})| = \frac{\nabla \Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})}, \quad (16)$$

and

$$E_L(\mathbf{R}) = \Psi_T^{-1} (-\frac{1}{2} \nabla^2 + V(\mathbf{R})) \Psi_T \quad (17)$$

is the local energy. This formulation imposes the *fixed-node approximation* [35]. The nodal surface of a wave function is the surface on which it is zero and across which it changes sign. The nodal surface of Ψ is constrained to be the same as that of Ψ_T and therefore f can be interpreted as a probability distribution. The time evolution generates the distribution $f = \Psi_T \Psi$, where Ψ is the best (lowest energy) wave function with the same nodes as Φ_T . The problem of the poor statistical behaviour due to the divergences in the potential energy is also solved because the term $(V(\mathbf{R}) - E_S)$ in Eq. (14) has been replaced by $(E_L(\mathbf{R}) - E_T)$ which is much smoother. Indeed, if Ψ_T was an exact eigenstate then $(E_L(\mathbf{R}) - E_T)$ would be independent of position in configuration space. Although we cannot in practice find the exact Ψ_T it is possible to eliminate the divergences in the local energy by choosing a Ψ_T which has the correct cusp-like behaviour whenever two electrons or an electron and a nucleus are coincident [36]. The fixed-node approximation implies that we solve independently in different nodal pockets, and at first sight it appears that we have to solve the Schrödinger equation in every nodal pocket, which would be an impossible task in large systems. However, the *tiling theorem* for exact fermion ground states [37, 38] asserts that all nodal pockets are in fact equivalent and therefore one only need solve the Schrödinger equation in one of them. This theorem is intimately connected with the existence of a variational principle for the DMC ground state energy [38].

A DMC simulation proceeds as follows. First we pick an ensemble of a few hundred configurations chosen from the distribution $|\Psi_T|^2$ using VMC and the standard Metropolis algorithm. This ensemble is evolved according to the short-time approximation to the Green function of the importance-sampled imaginary-time Schrödinger equation (Eq. (15)), which involves biased diffusion and addition/subtraction steps. The bias in the diffusion is caused by the importance sampling which directs the sampling towards parts of configuration space where $|\Psi_T|$ is large. After a period of equilibration the excited state contributions will have largely died out and the configurations start to trace out the probability distribution $f(\mathbf{R}) / \int f(\mathbf{R}) d\mathbf{R}$. We can then start to accumulate averages, in particular the DMC energy, E_D , which is given by

$$E_D = \frac{\int f(\mathbf{R}) E_L(\mathbf{R}) d\mathbf{R}}{\int f(\mathbf{R}) d\mathbf{R}} \approx \sum_i E_L(\mathbf{R}_i). \quad (18)$$

This energy expression would be exact if the nodal surface of Ψ_T was exact, and the fixed-node error is second order in the error in the nodal surface of Ψ_T (when a variational theorem exists [38]). The accuracy of the fixed node approximation can be tested on small systems and normally leads to very satisfactory results. The trial wave function limits the final accuracy that can be obtained because of the fixed-node approximation and it also controls the statistical efficiency of the algorithm. Like VMC, the DMC algorithm satisfies a zero-variance principle, i.e., the variance of the energy goes to zero as the trial wave function goes to an exact eigenstate.

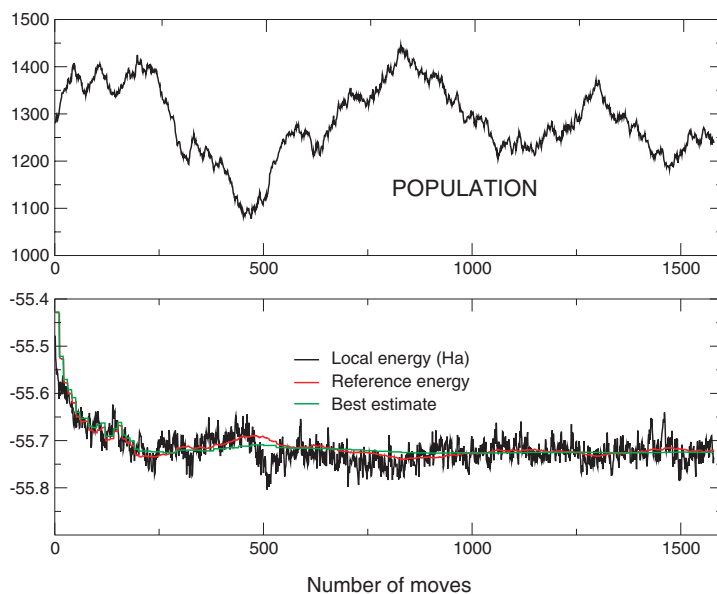


Fig. 1 DMC simulation of solid antiferromagnetic NiO. In the lower panel, the noisy black line is the local energy after each move, the green line is the current best estimate of the DMC energy, and the red line is E_T in Eq. (15) which is varied to control the population of configurations through a feedback mechanism. As the simulation equilibrates the best estimate of the energy, initially equal to the VMC energy, decreases significantly then approaches a constant – the final DMC energy. The upper panel shows the variation in the population of the ensemble during the simulation as walkers are created or destroyed.

3 Miscellaneous issues

In this section I will discuss some practical issues related to VMC and DMC.

3.1 More about trial wave functions

Single-determinant Slater–Jastrow wave functions often work very well in QMC calculations since the orbital part alone provides a pretty good description of the system. In the ground state of the carbon pseudo-atom, for example, a single Hartree–Fock determinant retrieves about 98.2% of the total energy. The remaining 1.8%, which at the VMC level must be recovered by the Jastrow factor, is the correlation energy and in this case it amounts to 2.7 eV – clearly important for an accurate description of chemical bonding. By definition a determinant of Hartree–Fock orbitals gives the lowest energy of all single-determinant wave functions and DFT orbitals are often very similar to them. These orbitals are not optimal when a Jastrow factor is included, but it turns out that the Jastrow factor does not change the detailed structure of the optimal orbitals very much, and the changes are well described by a fairly smooth change to the orbitals. This can be conveniently included in the Jastrow factor itself.

How though might we improve on the Hartree–Fock/DFT orbitals in the presence of the Jastrow factor? CASINO is capable of directly optimizing the atomic orbitals in a single atom by optimizing a parametrized function that is added to the self-consistent orbitals [39]. This was found to be useful only in certain cases. In atoms one often sees an improvement in the VMC energy but not in DMC, indicating that the Hartree–Fock nodal surface is close to optimal even in the presence of a correlation function. Unfortunately direct optimization of both the orbitals and Jastrow factor cannot easily be done for large polyatomic systems because of the computational cost of optimizing large numbers of parameters, and so it is difficult to know how far this observation extends to more complex systems. One promising tech-

nique [30, 31] is to optimize the potential that generates the orbitals rather than the orbitals themselves. Another possible way to improve the orbitals over the Hartree–Fock form, suggested by Grossman and Mitas [40], is to use a determinant of the natural orbitals which diagonalize the one-electron density matrix. It is not immediately clear why this should be expected to work in QMC however – the motivation appears to be that the convergence of configuration interaction expansions is improved by using natural orbitals instead of Hartree–Fock orbitals. The calculation of reasonably accurate natural orbitals is unfortunately computationally demanding, and this makes such an approach less attractive for large systems.

It should be noted that all such techniques which move the nodal surface of the trial function (and hence potentially improve the DMC energy) make wave function optimization with fixed configurations more difficult. The nodal surface deforms continuously as the parameters are changed, and in the course of this deformation the fixed set of electron positions of one of the configurations may end up being on the nodal surface. As the local energy $\hat{H}\Psi/\Psi$ diverges on the nodal surface, the unweighted variance of the local energy of a fixed set of configurations also diverges, making it difficult to locate the global minimum of the variance. A discussion of what one might do about this can be found in Ref. [34].

In some cases it is necessary to use multi-determinant wave functions to preserve important symmetries of the true wave function. In other cases a single determinant may give the correct symmetry but a significantly better wave function can be obtained by using a linear combination of a few determinants. Multi-determinant wave functions have been used successfully in QMC studies of small molecular systems and even in periodic calculations such as the recent study of the neutral vacancy in diamond due to Hood et al. [41]. However other studies have shown that while using multideterminant functions gives an improvement in VMC, this sometimes does not extend to DMC, indicating that the nodal surface has not been improved [39].

It is widely believed that a direct expansion in determinants (as used in, for example, configuration interaction calculations) converges very slowly because of the difficulty in describing the strong correlations which occur when electrons are close to one another. These correlations result in cusps in the wave function when two electrons are coincident, which are not well approximated by a finite sum of smooth functions [42]. However, this is not the whole story, and Prendergast et al. [43] have pointed out that the cusp is energetically less important, and that the slow convergence of determinant expansions has a lot to do with the description of medium-range correlations. In any case the number of determinants required to describe the wave function to some fixed accuracy increases exponentially with the system size; for some molecular cases billions of determinants have been used. Ordinarily one might think that an expansion which required so many terms is not a very good expansion, because the basis functions look nothing like the function that is being expanded, but this viewpoint has historically not been popular in the quantum chemistry community. As far as QMC is concerned, this would seem to rule out the possibility of retrieving a significant extra fraction of the correlation energy with QMC in large systems via an expansion in determinants. Methods in which only local correlations are taken into account might be helpful, but overall an expansion in determinants is not a promising direction to pursue for making QMC trial wave functions for large systems.

One approach which might be more useful is the *backflow* technique. Backflow correlations were originally derived from a current conservation argument by Feynman [44], and Feynman and Cohen [45] to provide a picture of the excitations in liquid ^4He and the effective mass of a ^3He impurity in ^4He . In a modern context they can also be derived from an imaginary-time evolution argument [46, 47]. In the backflow trial function the electron coordinates \mathbf{r}_i appearing in the Slater determinants of Eq. (8) are replaced by *quasiparticle coordinates*,

$$\bar{\mathbf{r}}_i = \mathbf{r}_i + \sum_{\substack{j=1 \\ (j \neq i)}}^N \eta(r_{ij}) (\mathbf{r}_i - \mathbf{r}_j), \quad (19)$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. The optimal function $\eta(r_{ij})$ may be determined variationally, and in so doing the nodal surface is shifted. Backflow thus represents another practical possibility for relaxing the constraints

of the fixed-node approximation in DMC. Kwon, Ceperley, and Martin [46, 48] found that the introduction of backflow significantly lowered the VMC and DMC energies of the two and three-dimensional uniform electron gas at high densities. The use of backflow has also been investigated for metallic hydrogen [49]. A full inhomogeneous backflow algorithm for real polyatomic systems has been implemented in the CASINO 2.0 program [50], and first results for the Ne atom and Ne⁺ ion are very promising [39]. One interesting thing that we found is that energies obtained from VMC with backflow approached those of DMC without backflow. VMC with backflow may thus represent a useful level of theory since it is significantly less expensive than DMC.

Unfortunately the use of backflow wave functions significantly increases the cost of QMC calculations. This is largely because every element of the Slater determinant has to be recomputed each time an electron is moved, whereas only a single column of the Slater determinant has to be updated after each move when the basic Slater–Jastrow wave function is used. The basic scaling of the algorithm with backflow is thus N^4 rather than N^3 . Backflow functions also introduce more parameters into the trial wave function, making the optimization procedure more difficult and costly. However the reduction in the variance normally observed with backflow greatly improves the statistical efficiency of QMC calculations, i.e., the number of moves required to obtain a fixed error in the energy is smaller. In our Ne atom calculations [39], for example, it was observed that the computational cost per move in VMC and DMC increased by a factor of between four and seven, but overall the time taken to complete the calculations increased only by a factor of two to three. Finally, it should be noted that backflow is expected to improve the QMC estimates of all expectation values, not just the energy, so on the whole it appears to be a *good thing*.

3.2 Basis set expansions: how to represent the orbitals?

The importance of using good quality single-particle orbitals in building up the Slater determinants in the trial wave function is clear. The determinant part accounts for by far the most significant fraction of the variational energy. However, the evaluation of the single-particle orbitals and their first and second derivatives can sometimes take up more than half of the total computer time, and consideration must therefore be given to obtaining accurate orbitals which can be evaluated rapidly at arbitrary points in space. It is not difficult to see that the most critical thing is to expand the single-particle orbitals in a basis set of *localized* functions. This ensures that beyond a certain system size, only a fixed number of the localized functions will give a significant contribution to a particular orbital at a particular point. The cost of evaluating the orbitals does not then increase rapidly with the size of the system. Note that ‘localized basis functions’ can (1) be strictly zero beyond a certain radius, or (2) can decrease monotonically and be pre-screened before the calculation starts, so that only those functions which could be significant in a particular region are considered for evaluation.

An alternative procedure is to tabulate the orbitals and their derivatives on a grid, and this is feasible for small systems such as atoms, but for periodic solids or larger molecules the storage requirements quickly become enormous. This is an important consideration when using parallel computers as it is much more efficient to store the single-particle orbitals on every node. Historically a very large proportion of condensed matter electronic structure theorists have used plane-wave basis sets in their DFT calculations. However in QMC, plane-wave expansions are normally extremely inefficient because they are not localized in real space; every basis function contributes at every point, and the required number of functions increases linearly with system size. Only if there is a short repeat length in the problem are plane waves not totally unreasonable. Note that this does not mean that all plane-wave DFT codes are useless for generating trial wave functions for CASINO; a post-processing utility can be used to reexpand a function expanded in plane-waves in another localized basis before the wave function is input into CASINO. The usual thing here is to use some form of localized spline functions on a grid such as the ‘blip’ functions used by Mike Gillan’s group [51] and implemented in CASINO by Dario Alfè [52].

Another pretty good way to do this is to expand the orbitals in a basis of Gaussian-type functions. These are localized, quick to evaluate, and are available from a wide-range of sophisticated software

packages. Such a large expertise has been built up within the quantum chemistry community with Gaussians that there is a significant resistance to using any other type of basis. A great many Gaussian-based packages have been developed by quantum chemists for treating molecules. The most well-known of these are the various versions of the GAUSSIAN package [53]. In addition to the regular single determinant methods, these codes include various techniques involving multi-determinant correlated wave functions (although sadly, not QMC!). This makes them very flexible tools for developing accurate molecular trial wave functions. For Gaussian basis sets with periodic boundary conditions, the CRYSTAL program [54] can perform all-electron or pseudopotential Hartree–Fock and DFT calculations both for molecules and for systems with periodic boundary conditions in one, two or three dimensions, which makes it very useful as a tool for generating trial functions for CASINO.

3.3 Pseudopotentials

Pseudopotentials or effective core potentials are commonly used in electronic structure calculations to remove the inert core electrons from the problem and to improve the computational efficiency. Although QMC scales very favourably with system size it has been estimated that the scaling of all-electron calculations with the atomic number Z is approximately $Z^{5.5-6.5}$ which is generally considered to rule out applications to atoms with Z greater than about ten. We have in fact pushed all-electron QMC calculations to $Z = 54$ using techniques to be described in the next section [55] although we were eventually forced to stop when smoke was observed coming out of the side of the computer [56]. The use of a pseudopotential serves to reduce the effective value of Z and although errors are inevitably introduced, the gain in computational efficiency is sufficient to make applications to heavy atoms feasible.

Accurate pseudopotentials for single-particle theories such as DFT or Hartree–Fock theory are well developed, but pseudopotentials for correlated wave function techniques such as QMC present additional challenges. The presence of core electrons causes two related problems. The first is that the shorter length scale variations in the wave function near a nucleus of large Z require the use of a small time step. This problem can be significantly reduced (in VMC at least) by the use of acceleration schemes [57, 58]. The second problem is that the fluctuations in the local energy tend to be large near the nucleus because both the kinetic and potential energies are large.

The central idea of pseudopotential theory is to create an effective potential which reproduces the effects of both the nucleus and the core electrons on the valence electrons. This is done separately for each of the different angular momentum states, so the pseudopotential contains angular momentum projectors and is therefore a non-local operator.

It is convenient to divide the pseudopotential for each atom into a local part $V_{\text{loc}}^{\text{ps}}(r)$ common to all angular momenta and a correction, $V_{\text{nl},l}^{\text{ps}}(r)$, for each angular momentum l . The electron-ion potential energy term in the full many-electron Hamiltonian of the atom then takes the form

$$V_{\text{loc}} + \hat{V}_{\text{nl}} = \sum_i V_{\text{loc}}^{\text{ps}}(r_i) + \sum_i \hat{V}_{\text{nl},i}^{\text{ps}}, \quad (20)$$

where $\hat{V}_{\text{nl},i}^{\text{ps}}$ is a non-local operator which acts on an arbitrary function $g(\mathbf{r}_i)$ as follows

$$\hat{V}_{\text{nl},i}^{\text{ps}} g(\mathbf{r}_i) = \sum_l V_{\text{nl},l}^{\text{ps}}(r_i) \sum_{m=-l}^l Y_{lm}(\Omega_{r_i}) \int Y_{lm}^*(\Omega_{r'_i}) \int g(\mathbf{r}'_i) d\Omega'_i, \quad (21)$$

where the angular integration is over the sphere passing through the r_i . This expression can be simplified by choosing the z -axis along \mathbf{r}_i , noting that $Y_{lm}(0, 0) = 0$ for $m \neq 0$, and using the definition of the spherical harmonics to give

$$\hat{V}_{\text{nl},i}^{\text{ps}} g(\mathbf{r}_i) = \sum_l V_{\text{nl},l}^{\text{ps}}(r_i) \frac{2l+1}{4\pi} \int P_l[\cos(\theta'_i)] g(\mathbf{r}'_i) d\Omega'_i, \quad (22)$$

where P_l denotes a Legendre polynomial.

It is not currently possible to construct pseudopotentials for heavy atoms entirely within a QMC framework, although progress in this direction was made by Acioli and Ceperley [59]. It is therefore currently necessary to use pseudopotentials generated within some other framework. Possible schemes include Hartree–Fock theory and local DFT, where there is a great deal of experience in generating accurate pseudopotentials. There is evidence to show that Hartree–Fock pseudopotentials give better results within QMC calculations than DFT ones, although DFT ones work quite well in many cases. The problem with DFT pseudopotentials appears to be that they already include a (local) description of correlation which is quite different from the QMC description. Hartree–Fock theory, on the other hand, does not contain any effects of correlation. The QMC calculation puts back the valence–valence correlations but neglects core–core correlations (which have only an indirect and small effect on the valence electrons) and core–valence correlations. Core–valence correlations are significant when the core is highly polarizable, such as in alkali-metal atoms. The core–valence correlations may be approximately included by using a ‘core polarization potential’ (CPP) which represents the polarization of the core due to the instantaneous positions of the surrounding electrons and ions. Another issue is that relativistic effects are important for heavy elements. It is still, however, possible to use a QMC method for solving the Schrödinger equation with the scalar relativistic effects obtained within the Dirac formalism incorporated within the pseudopotentials. The combination of Dirac–Hartree–Fock pseudopotentials and CPPs appears to work well in many QMC calculations. CPPs have been generated for a wide range of elements (see, e.g., Ref. [60]).

Many Hartree–Fock pseudopotentials are available in the literature, mostly in the form of sets of parameters for fits to Gaussian basis sets. Unfortunately many of them diverge at the origin, which can lead to significant time step errors in DMC calculations [61]. We concluded that none of the available sets are ideal for QMC calculations and that it would be helpful if we generated an on-line periodic table of smooth non-divergent Hartree–Fock pseudopotentials (with relativistic corrections). This project has now been completed by Trail and Needs, and is described in detail in Refs. [62, 63].

4 Recent developments

In this Section 1 will describe some recent improvements to the basic algorithms that improve the ability of QMC to (1) treat heavier atoms with all-electron calculations, and (2) to treat larger systems by improving the scaling behaviour. Both these features are implemented in the CASINO code.

4.1 All-electron QMC calculations for heavier atoms

At a nucleus the exact wave function has a cusp so that the divergence in the potential energy is cancelled by an equal and opposite divergence in the kinetic energy. If this cusp is represented accurately in the QMC trial wave function therefore, then the fluctuations in the local energy referred to in the previous section will be greatly reduced. Now if numerical orbitals are used it is relatively easy to produce an accurate representation of the cusp. However, as we have already remarked, such representations cannot really be used for large polyatomic systems because of the excessive storage requirements. Alternatively if the wave function is formed from determinants of single-particle orbitals expanded, for example, in a Gaussian basis set, then there can be no cusp in the wave function since Gaussians have zero gradient at $r = 0$. The local energy thus diverges at the nucleus. In practice one finds that the local energy has wild oscillations close to the nucleus which can lead to numerical instabilities in DMC calculations. To solve this problem we can make small corrections to the single particle orbitals close to the nuclei which impose the correct cusp behaviour. Such corrections need to be applied at each nucleus for every orbital which is larger than a given tolerance at that nucleus.

It is likely that a number of other researchers have developed such schemes, but within the literature we are only aware of the scheme developed by Manten and Lüchow [64], which is rather different from ours [65]. Our scheme is based on the idea of making the one-electron part of the local energy

for each orbital, $\hat{H}_{\text{oe}}\phi/\phi$, finite at the nucleus. \hat{H}_{oe} is given by

$$\hat{H}_{\text{oe}} = -\frac{1}{2}\nabla^2 - \frac{Z}{r}, \quad (23)$$

where r is the distance to the nucleus of charge Z . The scheme need only be applied to the s -component of orbitals centred at the nuclear position in question. Inside some radius r_c we replace the orbital expanded in Gaussians by $\phi = \text{sgn}[\psi(r=0)] \exp[p]$, where $\text{sgn}[\psi(r=0)]$ denotes the sign of the Gaussian orbital at $r=0$ and p is a polynomial in r . Therefore $\ln|\phi| = p$ and the local energy is given by

$$E_L = \frac{\hat{H}_{\text{oe}}\phi}{\phi} = -\frac{p'}{r} - \frac{p''}{2} - \frac{p'^2}{2} - \frac{Z}{r}. \quad (24)$$

We impose five constraints, that $p(r_c)$, $p'(r_c)$, and $p''(r_c)$ are continuous, that $p'(0) = -Z$ (to satisfy the cusp condition), and that $E_L(0)$ is chosen to minimize the maximum of the square of the deviation of $E_L(r)$ from an 'ideal curve' of local energy versus radius.

To see the cusp corrections in action, let us first look at a hydrogen atom where the basis set has been made to model the cusp very closely by using very sharp Gaussians with high exponents. Visually (top left in Fig. 2) the fact that the orbital does not obey the cusp condition is not immediately apparent. If we zoom in on the region close to the nucleus (top right) we see the problem: the black line is the orbital expanded in Gaussians, the red line is the cusp-corrected orbital. The effect on the gradient and local energy is clearly significant. This scheme has been implemented within the CASINO code both for finite and for periodic systems, and produces a significant reduction in the computer time required to achieve a specified error bar, as one can appreciate from Fig. 3.

In order to understand our capability to do all-electron DMC calculations for heavier atoms, and to understand how the necessary computer time scales with atomic number, we performed calculations for various noble gas atoms [55]. By ensuring that the electron–nucleus cusps were accurately represented it

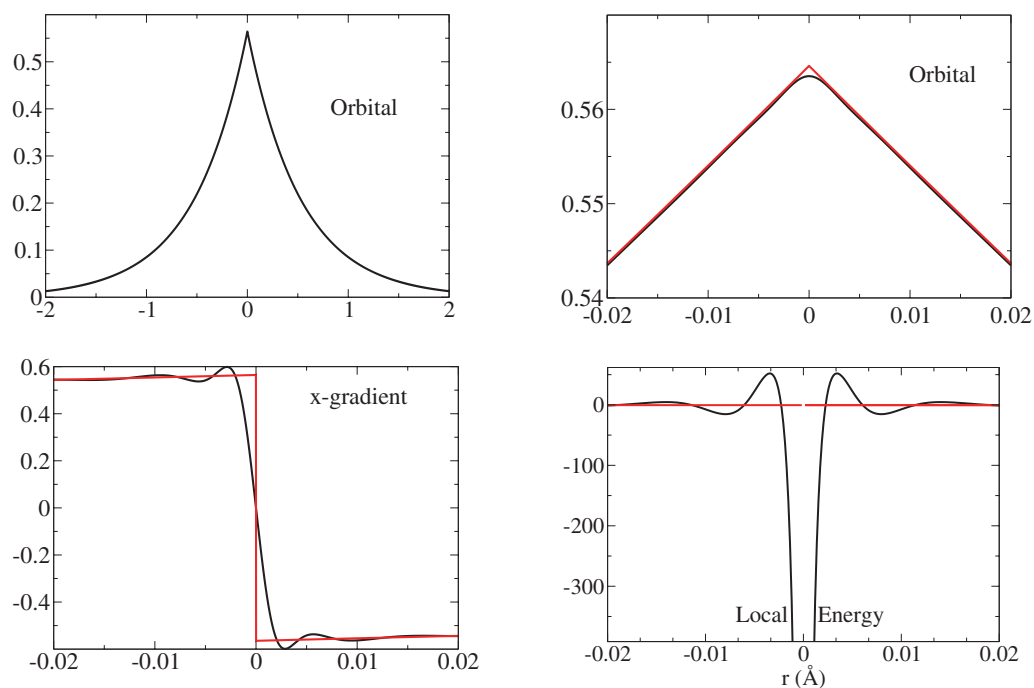


Fig. 2 Cusp corrections in the hydrogen atom.

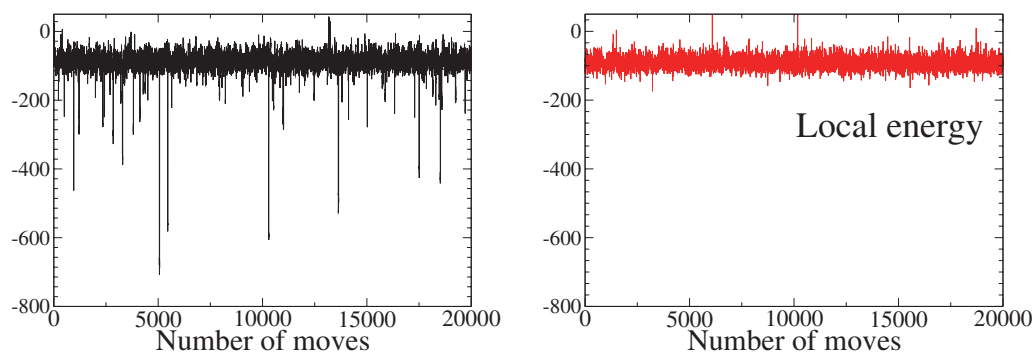


Fig. 3 Local energy as a function of move number in a VMC calculation for a carbon monoxide molecule with a standard reasonably good Gaussian basis set. The cusp corrections are imposed only in the figure on the right. The reduction in the local energy fluctuations with the new scheme is clearly apparent.

proved perfectly possible to produce converged DMC energies with acceptably small error bars for atoms up to xenon ($Z = 54$).

4.2 Improved scaling algorithms

Let us now consider in more detail how QMC calculations scale with system size, and what one might do in order to improve the scaling behaviour. QMC methods are stochastic and therefore yield mean values with an associated statistical error bar. We might want to calculate the energy of some system and compare it with the energy of a different arrangement of the atoms. The desired result might be a defect formation energy, an energy barrier, or an excitation energy. These are evidently energy differences which become *independent of the system size* when the system is large enough. To perform such a calculation we therefore require an error bar ΔE on the energy of the system which is independent of system size, a feature denoted here by $\Delta E = \mathcal{O}(1)$. There are other quantities such as cohesive energies, lattice constants, and elastic constants, for example, in which both energy and error bar may be defined *per atom* or *per formula unit*, in which case the error bar on the whole system is allowed to scale linearly with system size, i.e., $\Delta E = \mathcal{O}(N)$.

How does the computational cost C of a QMC calculation, yielding an error $\Delta E = \mathcal{O}(1)$, scale with the system size, measured by the number of electrons N ? The result for the standard algorithm with localized basis sets is $C = AN^3 + \varepsilon N^4$, where ε is very small [4]. In current solid simulations $N \leq 2000$, and the first term in this expression dominates, giving an N^3 scaling for the standard algorithm: double the system size and the cost goes up eightfold. What is the best scaling we could possibly achieve? As is well known, the best possible scaling for conventional (non-stochastic) single-particle methods such as DFT is $\mathcal{O}(N)$ [66]. A considerable effort has been made over the previous decade to design DFT codes which (a) scale linearly with system size, (b) are faster than the regular cubic scaling algorithm for reasonable system sizes, and (c) are *as accurate* as codes using the regular algorithm, with the latter two problems being the most difficult. In wave function-based QMC, these additional problems do not occur; with the improved scaling algorithms described here the speed benefit is immediate and there is essentially no loss of accuracy. However, for the scaling one cannot do better than $\mathcal{O}(N^2)$ in general, unless the desired quantity is expressible as an energy per atom. Why is this so? One still has the ‘near-sightedness’ in the many-body problem which is exploited in linear scaling DFT algorithms, but the difference is the stochastic nature of QMC. The statistical noise in the energy adds incoherently over the particles, so the variance in the mean energy increases as N (and thus the error bar as \sqrt{N}). Since the variance is inversely proportional to the number of statistically independent configurations in the calculation, we see that to obtain $\Delta E = \mathcal{O}(1)$ we must therefore evaluate the energy of $\mathcal{O}(N)$ configurations, each of which costs $\mathcal{O}(N)$ operations. This accounts for the ‘extra’ power of N in the cost of a QMC calculation. How-

ever, $\mathcal{O}(N^2)$ scaling is still a vast improvement over $\mathcal{O}(N^3)$ scaling when N can be of the order of a few thousand, and clearly the scaling is improved further for properties which can be expressed in terms of energies per atom. The primary task is thus to reduce the AN^3 term to AN^2 . The operations which make up this term are (1) evaluation of the orbitals in the Slater determinants, (2) evaluation of the Jastrow factor, and (3) evaluation of Coulomb interactions between particles.

The first of these operations is by far the most costly. As in $\mathcal{O}(N)$ -DFT methods, the solution is to use *localized orbitals* instead of the delocalized single-particle orbitals that arise naturally from standard DFT calculations. The number of such orbitals contributing at a point in space is *independent of N* which leads to the required improvement in scaling. Two different groups using the CASINO code have shown that this approach is extremely effective, namely Williamson, Hood, Grossman, and Reboredo [15, 67], and Alfè and Gillan [68]. An impartial evaluation of the two different methods [69] showed that the latter was superior, and this was the approach finally adopted for the production version of CASINO.

For the Jastrow factor all that is required to achieve the improved scaling is that it be truncated at some distance which is independent of system size. Because the correlations are essentially local it is natural to truncate the Jastrow factor at the radius of the exchange-correlation hole. Of course, truncating the Jastrow factor does not affect the final answer obtained within DMC because it leaves the nodal surface of the wave function unchanged, although if it is truncated at too short a distance the statistical noise increases. The scaling of the Coulomb interactions can be improved using an accurate scheme which exploits the fact that correlation is short-ranged to replace the long-range part by its Hartree contribution (in the style of the Modified Periodic Coulomb (MPC) interaction [23]).

For extremely large systems, the notionally εN^4 term might begin to be significant. This arises from N updates of the matrix of cofactors of the inverse Slater matrix (required when computing the ratio of new to old determinants after each electron move), each of which takes a time proportional to N^2 , plus the extra factor of N from the statistical noise. In CASINO this operation has been significantly streamlined through the use of sparse matrix techniques and we have not yet found a system where it contributes substantially to the overall CPU time.

Taken together the localization algorithms described above should speed up continuum fermion QMC calculations significantly for large systems, but we can view it in another light – as an *embedding* algorithm in which a QMC calculation could be embedded within a DFT one. The idea is to use the higher accuracy of QMC where it is most needed, such as around a defect site or in the neighbourhood of a molecule attached to a solid surface. Developments along the lines of those described here might allow such QMC/DFT embedding calculations to be performed for the first time. This is quite simple in VMC although a practical DMC embedding scheme would be more difficult.

5 Applications

Time and space preclude me from presenting a long list of applications, but here is an unfair comparison of the worst DFT functional with VMC and DMC for some cohesive energies of tetrahedrally-bonded semiconductors. Many other applications can be found in Ref. [4].

6 The CASINO code

CASINO [1, 2] is a program package originally developed in Cambridge in the groups of Richard Needs and Mike Towler. Its purpose is to perform quantum Monte Carlo electronic structure calculations for finite and periodic systems. The philosophy behind it involves generality, speed, portability and ease-of-use. Generality in this sense means that one ought to be able to create a trial wave function for any system, expanded in any of a variety of different basis sets, and use it as input to a CASINO QMC calculation. Clearly the wave functions must be generated by an external electronic structure program, and this must in the past have been persuaded to write out the wave function in a format that CASINO understands, either all by itself, or through the transformation of its standard output using a separate CASINO

Table 1 Cohesive energies of tetrahedrally bonded semiconductors calculated within the LSDA, VMC and DMC methods and compared with experimental values. The energies for Si, Ge, and C are quoted in eV per atom while those for BN are in eV per two atoms. Refs.: *a.* Farid and Needs [70], and references therein. *b.* Rajagopal et al. [19], *c.* Li, Ceperley, and Martin [95], *d.* Fahy, Wang, and Louie [71]. Zero-point energy corrections of 0.18 eV for C and 0.06 eV for Si have been added to the published values for consistency with the other data in the table. *e.* Malatesta, Fahy, and Bachelet [72], *f.* Hood et al. [41], *g.* Leung et al. [73], *h.* Estimated by Knittle et al. [74] from experimental results on hexagonal BN.

method	Si	Ge	C	BN
LSDA	5.28 ^a	4.59 ^a	8.61 ^a	15.07 ^e
VMC	4.38(4) ^c	3.80(2) ^b	7.27(7) ^d	12.85(9) ^e
	4.82(7) ^d		7.36(1) ^f	
	4.48(1) ^g			
DMC	4.63(2) ^g	3.85(2) ^b	7.346(6) ^f	
exp.	4.62(8) ^a	3.85 ^a	7.37 ^a	12.9 ^h

utility. This is one of the main reasons that producing a QMC code is somewhat labour intensive. Maintaining these interfaces as codes evolve, and persuading their owners that this is a good idea in the first place, is a difficult and sometimes frustrating task. It is nevertheless part of the philosophy that CASINO should support a reasonably wide range of the most popular electronic structure codes, and at the present time this list includes CRYSTAL95/98/03 [54], GAUSSIAN94/98/03 [53], CASTEP [75], ABINIT [76], PWSCF [77], ONETEP [78], TURBOMOLE [79] and JEEP.

The most important current capabilities of CASINO are as follows:

- It can do variational Monte Carlo calculations (including wave function optimization through minimization of the variance or the energy) and diffusion Monte Carlo calculations (branching DMC or pure DMC).
- It may be applied to finite systems such as atoms and molecules and also to systems with periodic boundary conditions in one, two or three dimensions (polymers, slabs/surfaces, crystalline solids) with arbitrary crystal structure.
- Arbitrary quantum particles (fermion/bosons) with user-defined spin, charge and mass tensor may be used in any combination.
- It uses flexible Slater–Jastrow many-electron wave functions where the Slater part may consist of multiple determinants of spin orbitals.
- The code may use orbitals expanded in a variety of basis sets in the determinantal part of the many-electron trial wave function: (1) *s*, *p*, *d*, *f*, *g* Gaussian basis functions centred on atoms or elsewhere (aperiodic or periodic systems) with cusp corrections in the case of all-electron calculations, (2) plane-waves (periodic systems), (3) blip functions, i.e., cubic splines on a regular grid (aperiodic or periodic systems) generated by post-processing the results of a plane-wave calculation, (4) atomic calculations with numerical orbitals interpolated from a radial grid.
- There are predefined defaults for a variety of 2D/3D electron phases with fluid or crystal wave functions, and electron–hole phases with fluid, crystal, or pairing wave functions, all with arbitrary cell shape, spin polarization, density and particle mass ratio. Excited states of these systems may be treated.
- Improved scaling behaviour is attainable through use of localized orbitals and localized basis functions.
- Both ground and excited state energies may be computed.
- The code can compute expectation values of quantities other than the energy such as density, spin density, spin density matrix, one- and two-electron density matrix, pair-correlation function, localization tensor, structure factors, and electric dipole moment.
- Each atom in the system can be treated as all-electron or it may have its core electrons replaced with a non-local pseudopotentials with *s*, *p*, *d*, non-locality and, if desired, corresponding core-polarization potentials.

- Spin-polarized systems such as magnetic solids may be treated, as can systems with non-collinear spins (albeit for a restricted set of cases).
- There is a full implementation of backflow correlations for both homogeneous and inhomogeneous systems.
- A variety of efficient wave function optimization algorithms are implemented.
- Electron–electron interactions in periodic systems may be evaluated using either the standard Ewald interaction, our ‘modified periodic Coulomb interaction’ [23] which is faster and has smaller Coulomb finite size effects, or directly from the structure factor.

And from a computational point of view, one may also note that:

- The source code is written in strict compliance with the Fortran90 standard using modern software design techniques. It is supposed to be easy to use, easy to install, and easy to read and understand. It contains a self-documenting help system and comes with a helpful manual and examples.
- The code has been parallelized using the MPI standard and has been tested in parallel on a large variety of multiprocessor hardware, such as the Hitachi SR2201, Cray T3E, SGI Origin 2000, SGI Altix, IBM SP3, Fujitsu Primepower, Alpha servers, and SunFire Galaxy machines along with standard Linux PC clusters. It is also set up for workstation use on DEC Alphas, SGI Octane and O2, Linux PC with various compilers. Installed MPI libraries are not required on single processor machines and the code should compile and run out of the box on most machines. The speed of the code scales essentially linearly with the number of processors on a parallel computer.

It is worth sketching a brief history of the CASINO code. Its development was inspired by a Fortran77 development code (known simply as ‘the QMC code’) written in the early 1990s in Cambridge by Richard Needs and Guna Rajagopal, assisted by many helpful discussions with Matthew Foulkes. This was later extended by Andrew Williamson up to 1995 and then by Mike Towler and Paul Kent up to 1998. Various different versions of this were able to treat fcc solids, single atoms and the homogeneous electron gas. By the late 1990s it was clear that a modern general code capable of treating arbitrary systems (e.g. at least atoms, molecules, polymers, slabs, crystals, and electron phases) was required, not only for the use of the Cambridge QMC group, but for public distribution. At that time, a user-friendly general publically available code did not exist, at least for periodic systems, and it was felt to be a good thing to create one to allow other researchers to join in the fun. So beginning in 1999 a new Fortran90 code, CASINO, was gradually developed in the group of Richard Needs initially by Mike Towler, considerably assisted from 2002 by Neil Drummond and from 2004 by Pablo Lopez Rios. Some routines from the old code were retained, translated and reused, although most were gradually replaced. Various additional contributions have been made over the years by Andrew Porter, Randy Hood, Dario Alfè, Andrew Williamson, Gavin Brown, Chris Pickard, Rene Gaudoin, Ben Wood, Zoltan Radnai and others. The code continues to be actively developed.

7 Discussion and conclusions

In this article I have provided a very brief review of the VMC and DMC methods, highlighted some recent innovations, and introduced our implementation in the CASINO code. I have tried to make the case that QMC is a useful addition to the toolbox of the computational electronic structure theorist.

That said, it’s important to be honest about the problems with QMC. If it is such a wonderful technique, why doesn’t everyone use it? Until fairly recently there was a document on the web published by David Ceperley’s group entitled ‘Top Ten List of reasons why quantum Monte Carlo is not generally used in chemistry’ which might serve as a basis for discussion. The Top Ten list contained the following twelve reasons:

1. We need forces, dummy!
2. Try getting O₂ to bind at the variational level.
3. How many graduate students lives have been lost optimizing wavefunctions?
4. It is hard to get 0.01 eV accuracy by throwing dice.
5. Most chemical problems have more than 50 electrons.

6. Who thought LDA or HF pseudopotentials would be any good?
7. How many spectra have you seen computed by QMC?
8. QMC is only exact for energies.
9. Multiple determinants. We can't live with them, we can't live without them.
10. After all, electrons are fermions.
11. Electrons move.
12. QMC isn't included in Gaussian 90. Who programs anyway?

This apparently first appeared on the web in 1996, so it might be worth examining whether any progress has been made in these areas in the last ten years. An attempt at a commentary follows.

1. *We need forces, dummy!*

Of course it is true that for QMC to be considered a general method, one ought to be able to calculate forces (i.e., derivatives of the energy with respect to nuclear displacements) in order to optimize structures or to perform dynamical simulations. In fact almost all QMC calculations up to the present time have been done within the Born–Oppenheimer approximation. The nuclear positions are thus fixed during the calculation and the wave function depends parametrically on the nuclear coordinates. The fixed nuclear positions are normally taken from geometry optimizations done with alternative methods such as DFT, on the principle that DFT is more reliable for geometries than for total energy differences. Calculating forces using a stochastic algorithm is a difficult thing to do. A straightforward application of the Hellmann–Feynman theorem (where the force is given by the gradient of the potential energy surface with respect to nuclear positions) leads to estimators with a very large variance. While a convincing general algorithm has yet to be demonstrated for QMC calculations, some progress has been made. The literature contains a variety of interesting contributions to this problem, which may be roughly classified into three groups:

- finite differences using correlated sampling techniques which take advantages of correlations between statistical samples to reduce the overall statistical error, e.g. [80].
- methods which use the standard Hellmann–Feynman formula. This is not normally useful in QMC as the Hellmann–Feynman estimator (at least with bare nuclei) has an infinite variance associated with it. One way of getting round this is to use some renormalized expression for the force, see e.g. [81]. Furthermore, the Hellmann–Feynman expression does not give the exact derivative of the DMC energy if the nodal surface is not exact and depends on the nuclear positions. This is due to an additional nodal term rising from the action of the kinetic-energy operator on the discontinuity in the derivative of the wave function at inexact nodal surfaces [82].
- direct attempts to calculate analytic derivatives, see e.g. [83].

Currently one cannot calculate accurate forces for large systems with QMC. However, some progress has been made and one has to be reasonably optimistic that a better general method for calculating forces will be devised in the near future. Also worth mentioning is the interesting recent attempt to design a coupled QMC and DFT molecular dynamics algorithm by Grossman and Mitas [84].

2. *Try getting O₂ to bind at the variational level.*

I haven't managed to do so either. However, one can choose not to be overly concerned with QMC calculations done at the variational level. The binding energy of the oxygen molecule comes out very accurately in DMC [85, 86].

3. *How many graduate students lives have been lost optimizing wavefunctions?*

To give a feel for the time scale involved in optimizing wave functions, I can tell you about the weekend recently when I added the entire G2-1 set [87, 88] to the examples included with the CASINO distribution. This is a standard set of 55 molecules with various experimentally well-characterized properties intended for benchmarking of different quantum chemistry methods see e.g. Ref. [89]. Grossman has published the results of DMC calculations of these molecules using pseudopotentials [85], while we are doing the same with all-electron calculations [86]. It took a little over three days using only a few single-processor workstations to create all 55 sets of example files from scratch including optimizing the Jas-

trow factors for each molecule. While if one concentrated very hard on each individual case one might be able to pull a little more energy out of a VMC simulation, the optimized Jastrow factors are all perfectly good enough to be used as input to DMC simulations. I suggest that the process is sufficiently automated these days that graduate students are better employed elsewhere; certainly we have not suffered any fatalities here in Cambridge.

4. *It is hard to get 0.01 eV accuracy by throwing dice.*

With modern computers and efficient computer codes there are a great many systems where one can get sufficient accuracy in a reasonable time. Obviously this becomes increasingly difficult for heavier atoms and large systems, but as discussed previously, satisfying the electron-nuclear cusp condition accurately in all-electron calculations or using pseudopotentials helps a lot.

5. *Most chemical problems have more than 50 electrons.*

QMC calculations for several thousand electrons (per simulation cell) have been published, and this number will only increase with the new improved scaling techniques currently being introduced, and with the increasing power of available computational hardware.

6. *Who thought LDA or HF pseudopotentials would be any good?*

Very accurate QMC results using LDA/HF pseudopotentials have been published. Such pseudopotentials seem to work best for *sp*-bonded systems, and it is not clear that particularly good results can be obtained in systems containing, for example, transition elements. As previously explained, our group has developed an on-line periodic table containing a new set of smooth non-divergent Dirac–Hartree–Fock pseudopotentials which seem to be particularly satisfactory for QMC calculations [62, 63].

7. *How many spectra have you seen computed by QMC?*

Almost none. The calculation of excited state properties is a difficult problem for most other methods; one can sometimes compute the energies of individual excited states pretty accurately with QMC, but the calculation of complete spectra is too difficult for the moment. It is very difficult in general to treat frequency-dependent properties.

8. *QMC is only exact for energies.*

The problem here is to find unbiased expectation values of operators that do not commute with the Hamiltonian. In standard DMC, the time evolution generates the mixed distribution $f(\mathbf{R}) = \Psi_T(\mathbf{R})\Psi_0(\mathbf{R})$, where $\Psi_0(\mathbf{R})$ is the best (lowest energy) wave function with the same nodes as the trial function $\Psi_T(\mathbf{R})$. To obtain the exact energy with respect to the distribution $\Psi_0^2(\mathbf{R})$ it is necessary to project by means of a mixed estimator, which in practice just involves summing the local energies over the positions generated in the DMC simulation:

$$E_{\text{DMC}} = \frac{\int \Psi_0(\mathbf{R}) \hat{H} \Psi_T(\mathbf{R}) d\mathbf{R}}{\int \Psi_0(\mathbf{R}) \Psi_T(\mathbf{R}) d\mathbf{R}} \approx \frac{1}{M} \sum_{i=1}^M E_L(\mathbf{R}_m). \quad (25)$$

If another operator \hat{O} commutes with the Hamiltonian, then any eigenstate of \hat{H} is (or may be chosen to be) an eigenstate of \hat{O} and the same mixed estimator can be used to work out its expectation value. If \hat{O} and \hat{H} do not commute, then the mixed estimator does not give the ground state eigenvalue of \hat{O} . It is a good approximation to it however, and normally lies somewhere between the variational estimate and the exact value.

In general, the best thing to do to improve the expectation value of an operator that does not commute with the Hamiltonian is to improve the trial wave function. One may additionally use the extrapolation method [90] which approximates the true estimator as a linear combination of the mixed and variational ones, but it is not aesthetically pleasing and may not be particularly accurate in all cases. Note however that if the operator is just a simple local function of \mathbf{R} and not, for example, a differential operator, then

various methods can be used to obtain the ground state eigenvalue exactly. Among them are the future walking algorithm (clearly explained in Ref. [3]) based on estimating the ratio Ψ_0/Ψ_T using the asymptotic offspring of the DMC branching term, and the related time-correlation method [91]. Both of these methods are somewhat delicate however, and lead to signal-to-noise ratios that decay to zero at large time. Neither are likely to prove useful for large systems. More recently introduced bilinear methods [92] show promise but have not yet been fully developed into a practical general scheme. Probably the best scheme for the moment is the 'reptation QMC' method of Baroni and Moroni [93]. As mentioned earlier, VMC with backflow seems to be almost as good as DMC in some cases, and may represent an attractive alternative for computing expectation values other than the energy.

9. *Multiple determinants. We can't live with them, we can't live without them.*

As we have seen, multiple determinants are occasionally required to preserve important wave function symmetries, but more often one sees that using a linear combination of a few determinants can give a significantly lower DMC energy (this is absolutely not guaranteed in DMC however). If one does see an energy lowering it is because the nodal surface of this wave function is in some sense closer to the true one than that of the single determinant trial function. This is a problem not least because most electronic structure codes are not capable of generating optimized multiple determinant wave functions using e.g. the multi-configuration self-consistent field (MCSCF) method. Of the eight codes in the list supported by CASINO, only two (GAUSSIAN and TURBOMOLE) support multideterminant calculations, and to my knowledge there is no periodic code that does so. In an earlier section we observed that the number of determinants likely to be required in large systems effectively rules this out as a general solution. Orbital optimization and the backflow method are likely to prove more useful.

10. *After all, electrons are fermions.*

The solution to the fermion sign problem, which in this context means finding a way to bypass the fixed-node approximation in DMC, is one of the most interesting and important problems in computational physics. A solution to this would allow one to provide exact numerical solutions to the many-electron Schrödinger equation for fermionic systems (such as those containing electrons). It might seem strange to say that a stochastic or Monte Carlo method could solve such a problem exactly, since there is always a statistical error associated with the result, but we rely on the central limit theorem to provide an estimate of the probability that the exact results lies within a given interval. If we need to have the error to be less than this amount, this implies that the computer time will scale proportional to one over the square of the error.

We have seen that in DMC random walks serve to filter out the higher energy modes of some initial distribution, so that for very large imaginary time, the probability distribution of the random walkers is given by the ground state. The problem is that the ground state in question is actually the *bosonic* ground state which has the same sign everywhere in configuration space. It is only by imposing an additional boundary condition for the electronic Hamiltonian in the form of the fixed nodal surface that we are able to prevent collapse to the bosonic state, and maintain the positive and negative regions which are required for an antisymmetric wave function. The simplest methods which allow diffusing configurations to cross nodes and change the nodal surface, such as release-node Monte Carlo [94], are very inefficient because the 'signal' quickly becomes overwhelmed by the 'noise'. Methods which overcome this problem, such as fermion Monte Carlo [7–9] have been shown to be apparently stable for small test systems, but suffer from scaling problems with system size. It should never be forgotten that a genuine solution to the fermion sign problem in the context of continuum Monte Carlo methods must not only give the exact answer, but must also scale as some low-order polynomial with system size, and work generally rather than just for certain special cases. No proposed solution (and there are many in the literature) has all these characteristics. Recall that in some sense we already know how to solve the Schrödinger equation exactly for any system. One can in principle just do, for example, a full CI expansion in a complete basis set. The only trouble is it would take forever to compute and so this is not a viable solution.

That said, the fixed-node approximation is a good one, and one can normally recover well in excess of 95% of the correlation energy from a DMC calculation. For most problems this turns out to be accurate enough. Another approach – which is not a genuine solution to the sign problem but which might be the best practical thing to do – would be to optimize the nodal surface in some way, e.g., by optimizing the orbitals in the Slater part of the trial wave function (recalling that in some cases more than one determinant might be required for symmetry purposes) or by using backflow correlations. Research is ongoing in our group regarding the best way to do this.

11. *Electrons move.*

Quite so.

12. *QMC isn't included in Gaussian 90. Who programs anyway?*

Nor in Gaussian94, '98 or '03. Neil Drummond, Pablo Lopez Rios and I plus some of our collaborators like programming.

So, quite a lot of progress has been made in the theory and practical implementation of quantum Monte Carlo over the last few years, but certainly many interesting problems remain to be solved. For its most important purpose of calculating highly accurate total energies, the method works well and currently has no serious competitors for medium-sized and large systems. The CASINO program [1, 2] has been designed to allow researchers to explore the potential of QMC in arbitrary molecules, polymers, slabs and crystalline solids and in various model systems including standard electron and electron–hole phases such as the homogenous electron gas and Wigner crystal.

Acknowledgements I would like to thank the Royal Society (www.royalsoc.ac.uk, for the award of a research fellowship, to acknowledge the underlying contributions to this article of Richard Needs, Neil Drummond, Pablo Lopez Rios and Matthew Foulkes, and to thank all our various collaborators.

References

- [1] R. J. Needs, M. D. Towler, N. D. Drummond, and P. Lopez Rios, CASINO version 2.0 User Manual, University of Cambridge, Cambridge (2006).
- [2] CASINO web site: www.tcm.phy.cam.ac.uk/~mdt26/casino2.html
- [3] B. L. Hammond, W. A. Lester, Jr., and P. J. Reynolds, Monte Carlo Methods in Ab Initio Quantum Chemistry (World Scientific, Singapore, 1994).
- [4] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).
- [5] D. M. Ceperley and E. L. Pollock, *Phys. Rev. Lett.* **56**, 351 (1986).
- [6] C. Pierleoni, D. M. Ceperley, B. Bernu, and W. R. Magro, *Phys. Rev. Lett.* **73**, 2145 (1994).
- [7] M. H. Kalos and K. E. Schmidt, *J. Stat. Phys.* **89**, 425 (1997).
- [8] F. Pederiva and M. H. Kalos, *Comput. Phys. Commun.* **122**, 445 (1999).
- [9] M. H. Kalos and F. Pederiva, *Phys. Rev. Lett.* **85**, 3547 (2000).
- [10] D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).
- [11] Y. S. M. Wu, A. Kuppermann, and J. B. Anderson, *Phys. Chem. Chem. Phys.* **1**, 929 (1999).
- [12] G. Ortiz, M. Harris, and P. Ballone, *Phys. Rev. Lett.* **82**, 5317 (1999).
- [13] D. Alfè, M. J. Gillan, M. D. Towler, and R. J. Needs, *Phys. Rev. B* **70**, 214102 (2004).
- [14] M. D. Towler, R. Maezono, and R. J. Needs, to be published.
- [15] A. J. Williamson, R. Q. Hood, and J. C. Grossman, *Phys. Rev. Lett.* **87**, 246406 (2001).
- [16] S. Manten and A. Lüchow, *J. Chem. Phys.* **119**, 1307 (2003).
- [17] E. Schrödinger, *Ann. Phys.* **79**, 361 (1926).
- [18] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders, 1976), p. 330.
- [19] G. Rajagopal, R. J. Needs, A. J. James, S. D. Kenny, and W. M. C. Foulkes, *Phys. Rev. B* **51**, 10591 (1995).
- [20] G. Rajagopal, R. J. Needs, S. D. Kenny, W. M. C. Foulkes, and A. J. James, *Phys. Rev. Lett.* **73**, 1959 (1994).
- [21] P. P. Ewald, *Ann. Phys.* **64**, 25 (1921).

- [22] M. P. Tosi, in: *Solid State Physics*, Vol. 16, edited by H. Ehrenreich and D. Turnbull (Academic, New York, 1964), p. 1.
- [23] L. M. Fraser, W. M. C. Foulkes, G. Rajagopal, R. J. Needs, S. D. Kenny, and A. J. Williamson, *Phys. Rev. B* **53**, 1814 (1996).
A. J. Williamson, G. Rajagopal, R. J. Needs, L. M. Fraser, W. M. C. Foulkes, Y. Wang, and M.-Y. Chou, *Phys. Rev. B* **55**, R4851 (1997).
- [24] W. Kohn, *Rev. Mod. Phys.* **71**, 1253 (1999).
- [25] N. D. Drummond, First Year Ph.D. Report, University of Cambridge.
- [26] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. M. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- [27] R. J. Jastrow, *Phys. Rev.* **98**, 1479 (1955).
- [28] Z. Radnai, N. D. Drummond, M. D. Towler, and R. J. Needs, unpublished.
- [29] N. D. Drummond, M. D. Towler, and R. J. Needs, *Phys. Rev. B* **70**, 235119 (2004).
- [30] S. Fahy, in: *Quantum Monte Carlo Methods in Physics and Chemistry*, Nato Science Series C: Mathematical and Physical Sciences, Vol. 525, edited by M. P. Nightingale and C. J. Umrigar (Kluwer Academic, Dordrecht, The Netherlands, 1999), p. 101.
- [31] C. Filippi and S. Fahy, *J. Chem. Phys.* **112**, 3523 (2000).
- [32] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, *Phys. Rev. Lett.* **60**, 1719 (1988).
- [33] P. R. C. Kent, R. J. Needs, and G. Rajagopal, *Phys. Rev. B* **59**, 12344 (1999).
- [34] N. D. Drummond and R. J. Needs, *Phys. Rev. B* **72**, 085124 (2005).
- [35] J. B. Anderson, *J. Chem. Phys.* **63**, 1499 (1975).
- [36] T. Kato, *Commun. Pure Appl. Math.* **10**, 151 (1957).
- [37] D. Ceperley, *J. Stat. Phys.* **63**, 1237 (1991).
- [38] W. M. C. Foulkes, R. Q. Hood, and R. J. Needs, *Phys. Rev. B* **60**, 4558 (1999).
- [39] N. D. Drummond, P. Lopez Rios, A. Ma, J. R. Trail, G. Spink, M. D. Towler, and R. J. Needs, *J. Chem. Phys.* **124**, 224104 (2006).
- [40] J. C. Grossman and L. Mitas, *Phys. Rev. Lett.* **74**, 1323 (1995).
- [41] R. Q. Hood, P. R. C. Kent, R. J. Needs, and P. R. Briddon, *Phys. Rev. Lett.* **91**, 076403 (2003).
- [42] W. Kutzelnigg and J. D. Morgan, III, *J. Phys. Chem.* **96**, 4484 (1992).
- [43] D. Prendergast, M. Nolan, C. Filippi, S. Fahy, and J. C. Greer, *J. Chem. Phys.* **115**, 1626 (2001).
- [44] R. P. Feynman, *Phys. Rev.* **94**, 262 (1954).
- [45] R. P. Feynman and M. Cohen, *Phys. Rev.* **102**, 1189 (1956).
- [46] Y. Kwon, D. M. Ceperley, and R. M. Martin, *Phys. Rev. B* **48**, 12037 (1993).
- [47] M. Holzmann, D. M. Ceperley, C. Pierleoni, and K. Esler, *Phys. Rev. E* **68**, 046707 (2003).
- [48] Y. Kwon, D. M. Ceperley, and R. M. Martin, *Phys. Rev. B* **58**, 6800 (1998).
- [49] C. Pierleoni, D. M. Ceperley, and M. Holzmann, *Phys. Rev. Lett.* **93**, 146402 (2004).
- [50] P. Lopez Rios, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, *Phys. Rev. E*, in press.
- [51] E. Hernandez, M. J. Gillan, and C. M. Goringe *Phys. Rev. B* **55**, 13485 (1997).
- [52] D. Alfè and M. J. Gillan, *Phys. Rev. B* **70**, 161101 (2004).
- [53] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, *Gaussian 03, Revision B.03* (Gaussian, Inc., Pittsburgh PA, 2003).
- [54] V. R. Saunders, R. Dovesi, C. Roetti, R. Orlando, C. M. Zicovich-Wilson, N. M. Harrison, K. Doll, B. Civalieri, I. Bush, Ph. D'Arco and M. Llunell, *CRYSTAL2003 User's Manual* (University of Torino, Torino, 2003).
- [55] A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, *Phys. Rev. E* **71**, 066704 (2005).
- [56] A practice which has recently been outlawed in our department by new University anti-smoking legislation. My thanks to the anonymous referee for supplying me with this joke.

- [57] C. J. Umrigar, *Phys. Rev. Lett.* **71**, 408 (1993).
- [58] M. L. Stedman, W. M. C. Foulkes, and M. Nekovee, *J. Chem. Phys.* **109**, 2630 (1998).
- [59] P. H. Acioli and D. M. Ceperley, *J. Chem. Phys.* **100**, 8169 (1994).
- [60] E. L. Shirley and R. M. Martin, *Phys. Rev. B* **47**, 15413 (1993).
- [61] C. W. Greeff and W. A. Lester, Jr., *J. Chem. Phys.* **109**, 1607 (1998).
- [62] J. R. Trail and R. J. Needs, *J. Chem. Phys.* **122**, 174109 (2005).
- [63] J. R. Trail and R. J. Needs, *J. Chem. Phys.* **122**, 014112 (2005).
- [64] S. Manten and A. Lüchow, *J. Chem. Phys.* **115**, 5362 (2001).
- [65] A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, *J. Chem. Phys.* **122**, 224322 (2005).
- [66] S. Goedecker, *Rev. Mod. Phys.* **71**, 1085 (1999).
- [67] F. A. Reboredo and A. J. Williamson, *Phys. Rev. B* **71**, 121105 (2005).
- [68] D. Alfè and M. J. Gillan, *J. Phys.: Condens. Mater.* **16**, L305–L311 (2004).
- [69] N. D. Drummond, unpublished.
- [70] B. Farid and R. J. Needs, *Phys. Rev. B* **45**, 1067 (1992).
- [71] S. Fahy, X. W. Wang, and S. G. Louie, *Phys. Rev. B* **42**, 3503 (1990).
- [72] A. Malatesta, S. Fahy, and G. B. Bachelet, *Phys. Rev. B* **56**, 12201 (1997).
- [73] W.-K. Leung, R. J. Needs, G. Rajagopal, S. Itoh, and S. Ihara, *Phys. Rev. Lett.* **83**, 2351 (1999).
- [74] E. Knittle, R. Wentzcovitch, R. Jeanloz, and M. L. Cohen, *Nature* **337**, 349 (1989).
- [75] M. D. Segall, P. L. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark, and M. C. Payne, *J. Phys.: Condens. Mater.* **14**, 2717 (2002).
- [76] First-principles computation of material properties: the ABINIT software project, X. Gonze, J.-M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G.-M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, Ph. Ghosez, J.-Y. Raty, and D. C. Allan, *Comput. Mater. Sci.* **25**, 478–492 (2002).
- [77] S. Baroni, A. Dal Corso, S. de Gironcoli, and P. Giannozzi, <http://www.pwscf.org>.
- [78] C. K. Skylaris, A. A. Mostofi, P. D. Haynes, O. Diéguez, and M. C. Payne, *Phys. Rev. B* **66**, 035119 (2002).
A. A. Mostofi, P. D. Haynes, C. K. Skylaris, and M. C. Payne, *J. Chem. Phys.* **119**, 8842 (2003).
- [79] R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel, *Chem. Phys. Lett.* **162**, 165 (1989).
- [80] C. Filippi and C.J. Umrigar, *Phys. Rev. B* **61**, 16291 (2000).
- [81] R. Assaraf and M. Caffarel, *J. Chem. Phys.* **113**, 4028 (2000).
- [82] K. C. Huang, R. J. Needs, and G. Rajagopal, *J. Chem. Phys.* **112**, 4419 (2000).
- [83] J. Vrbik and S. M. Rothstein, *J. Chem. Phys.* **96**, 2071 (1992).
P. Belohorec, S. M. Rothstein, and J. Vrbik, *J. Chem. Phys.* **98**, 6401 (1993).
- [84] J. C. Grossman and L. Mitas, *Phys. Rev. Lett* **94**, 056403 (2005).
- [85] J. C. Grossman, *J. Chem. Phys.* **117**, 1434 (2002).
- [86] M. D. Towler, unpublished.
- [87] J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, *J. Chem. Phys.* **90**, 5622 (1989).
- [88] L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari, and J. A. Pople, *J. Chem. Phys.* **93**, 2537 (1990).
- [89] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, *J. Chem. Phys.* **106**, 1063 (1997).
- [90] P. A. Whitlock, D. M. Ceperley, G. V. Chester, and M. H. Kalos, *Phys. Rev. B* **19**, 5598 (1979).
- [91] D. M. Ceperley and B. Bernu, *J. Chem. Phys.* **89**, 6316 (1988).
- [92] F. Arias de Saavedra and M. H. Kalos, *Phys. Rev. E* **67**, 026708 (2003).
S. W. Zhang and M. H. Kalos, *J. Stat. Phys.* **70**, 515 (1993).
- [93] S. Baroni and S. Moroni, *Phys. Rev. Lett.* **82**, 4745 (1999).
- [94] D. M. Ceperley and B. J. Alder, *J. Chem. Phys.* **81**, 5833 (1984).
- [95] X.-P. Li, D. M. Ceperley, and R. M. Martin, *Phys. Rev. B* **44**, 10929 (1991).