

# Quantum Monte Carlo, or, how to solve the many-particle Schrödinger equation accurately whilst retaining favourable scaling with system size

M.D. Towler

*TCM group, Cavendish Laboratory, Cambridge University,  
J.J. Thomson Ave., Cambridge CB3 0HE, U.K.*

I introduce and discuss the quantum Monte Carlo method, a state-of-the-art computer simulation technique capable of solving the equations of quantum mechanics with extremely high accuracy whilst remaining tractable for systems with relatively large numbers of constituent particles. The CASINO code, developed in our group in Cambridge over many years, is briefly reviewed and results obtained with it are used to illuminate the discussion.

## I. INTRODUCTION

Of course we all know how to solve the Schrödinger equation for a given quantum-mechanical system as accurately as we like. Just do a full configuration interaction calculation<sup>1</sup> with an overwhelmingly large basis set. Simple. The only problem with such an approach is that - for more than a few tens of electrons - the calculation will take *forever* and this requires more patience than most people have.

The main problem in computational electronic structure theory, therefore, amounts to getting the required computer time to scale nicely as a function of system size without a significant accompanying loss of accuracy. In other words, how can we approximately solve the Schrödinger equation using a highly-accurate method that increases in computational cost as a low-order polynomial with the size of the system? Currently the *only* known way to do this, using a method that remains tractable for systems containing up to at least a few thousand electrons, is the *continuum quantum Monte Carlo* method. With some relatively simple tricks this can be made to scale as the square of the system size. A quantum chemistry ‘gold-standard’ technique with the requisite accuracy such as the coupled-cluster method CCSD(T) scales in most circumstances as the seventh power of the system size. By restricting one’s attention to atoms and very small molecules it is possible to pretend this doesn’t matter, but it does, as anyone who wishes to study solids, surfaces, realistic industrial chemical reactions, or biochemistry will readily appreciate. If you double the size of the system, the cost of a QMC calculation goes up by a factor of 4 or maybe 8; that of a CCSD(T) calculation by a factor of 128. No contest.

Given this overwhelmingly favourable scaling one might assume that QMC would be in widespread use throughout the quantum chemistry community. However, it is not and one might wish to speculate as to why this is the case. One obvious reason is that, despite the favourable scaling, the method has a large pre-factor - that is, it is just an expensive technique. In the worst cases the computational cost might be, for example, up to a thousand times more than a low-accuracy technique with comparable scaling such as density functional theory (DFT)<sup>2</sup>. This might sound disastrous but it turns out that QMC is ideally suited to parallel computing and can exploit modern multiprocessor machines with significantly more efficiency than most other methods. So use a one-hundred processor machine and it’s only ten times slower than DFT. Clearly this is beginning to sound reasonable. The widespread lack of availability of parallel machines in the past is one reason for the limited take-up of QMC but we are surely not so far off the time when it will be routine to have such a thing under one’s desk, and as we move into the petascale era it pays to be prepared.

From the usual quantum chemistry perspective one may derive a quantum of solace from the fact that the mainstream accurate techniques mostly fall into a relatively well-defined hierarchy with all levels systematically derivable from simple Hartree-Fock theory, and the fact that all these methods are implemented in standard widely-available software packages such as the Gaussian09 code<sup>3</sup>. In general, in this field, analytically integrable basis sets - usually Gaussians - are preferred when representing the wave function, with an implied distaste for the messiness of numerical integration. Unfortunately QMC is not directly related to the usual hierarchy of techniques, nor is it implemented in standard commercial codes. Furthermore, the way it works is very far from the quantum chemistry comfort zone; it is effectively a clever but brute force way of doing numerical integration of high-dimensional wave functions usually wedded to optimization techniques for improving the shape of the integrand. ‘Quantum Monte Carlo’ is actually a catch all-term for various methods having in common only the use of *random sampling*, which turns out to be by far the most efficient way to do many-dimensional numerical integration. Exaggerating only slightly, a stochastic technique such as QMC has the disconcerting property of being *accurate* but not precise (it gives numbers essentially in agreement with experiment but possessing a finite error bar), in contrast to the more usual techniques

still tractable for large systems which are *precise* but not accurate (they give the wrong answer to sixteen decimal places). Nevertheless we all derive comfort from precision, and for good reason. QMC calculations may have to be run for a considerable time to get the error bar down to an acceptable value, particularly in cases where one has to compute *differences* between two similar but imprecise numbers (the error bar in the numerical derivative computed by a two-point finite difference might be around a hundred times the error bar in the individual energies).

Where the method is used at all, two particular varieties of QMC are commonly employed - namely variational Monte Carlo (VMC) and diffusion Monte Carlo (DMC)<sup>4,5</sup>. In this article I shall give a short introduction to both. As will become apparent, VMC is a conceptually very simple technique; it can calculate arbitrary expectation values by sampling a given fixed many-electron wave function using the standard Monte Carlo integration algorithm. As the method turns out to be variational it is also possible, to some extent, to *optimize* suitably-parametrized wave functions using standard techniques. Not insisting that the wave function be expanded in analytically integrable functions confers the advantage that such wave functions can be *explicitly correlated*, that is, they can have a direct dependence on interparticle distances and thus a vastly more compact representation. This is not the final answer, of course, as no-one knows any practical way to parameterize arbitrarily accurate wave functions; the problem here is clearly one of *complexity*. However there exists a class of methods, collectively called ‘projector Monte Carlo’, which in principle solve quantum problems exactly - they attempt the much more difficult job of simultaneously creating and sampling the unknown exact ground state wave function. Diffusion Monte Carlo, or DMC, is one such method. In principle such techniques depend on guessed properties of the many-body wave function only in their computational efficiency, if at all. In particular, since the wave function is not represented in terms of a basis set but by the average distribution of a time-varying ensemble of particles, then for most practical purposes the ‘basis set problem’ is greatly minimized in DMC. Although a basis set is usually used its sole purpose is to represent a *guiding function* required for importance sampling - that is, grouping the sampling points in regions where they are most required - and 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). It should be noted for completeness that the use of wave functions implies zero temperature; for finite temperature simulations one is required to use density matrix formalisms which are somewhat underdeveloped in this field.

People not familiar with this technique might wonder what sorts of quantum-mechanical systems it can be or has been applied to. Specialists in density functional theory should certainly have heard of it, as its most famous early application was in Ceperley and Alder’s essentially exact QMC calculations<sup>6</sup> of the correlation energy of the homogeneous electron gas in 1980. The first popular local density approximations (LDAs) to the correlation energy density interpolated the accurate QMC values obtained at various densities whilst reproducing the exactly known limiting behavior at high and low densities. Various different analytic approaches led to a number of widely-used LDAs for the correlation functional such as Vosko-Wilk-Nusair (VWN)<sup>7</sup> and Perdew-Zunger (PZ81)<sup>8</sup>. One could thus argue that QMC was largely responsible for the huge growth of the DFT industry during the 1980s and beyond. In this review however, I shall shift the focus away from QMC calculations of model systems towards real continuum problems involving real atoms (as opposed to ‘pretend atoms’ without separate nuclei and electrons which are sometimes used). It took some time before such applications became routine. Significant calculations of ‘chemical’ systems began to appear in the late 1980s and early 1990s (though completely general and publically available software had to wait for the twenty-first century). The accuracy of the technique was amply demonstrated; in molecules containing light atoms such as H and He total energies accurate to greater than 0.01 kcal/mole ( $\approx 1.5 \times 10^{-5}$  Ha!) were produced, for example, on tens of thousands of points on the  $\text{H} + \text{H}_2 \rightarrow \text{H}_2 + \text{H}$  potential energy surface<sup>9</sup>. A list of interesting references to systems and physical problems that have been treated with QMC since then might include equations of state in solids<sup>10–13</sup>, solid-state structural phase transitions<sup>14</sup>, binding of molecules and their excitation energies<sup>15–19</sup>, studies of exchange and correlation<sup>20–23</sup>, optical band gaps of nanocrystals<sup>24,25</sup>, defects in semiconductors<sup>26–28</sup>, transition metal oxide chemistry<sup>29–31</sup>, band structures of insulators<sup>32–34</sup>, quantum dots<sup>35</sup>, reconstruction of crystalline surfaces<sup>36</sup>, molecules on surfaces<sup>37,38</sup>, and pairing in ultra-cold atomic gases<sup>39–41</sup>. DFT, of course, continues to be the pre-eminent technique for treating problems in computational electronic structure theory. However, it has now been convincingly demonstrated that systems and problems for which an accurate determination of the total energy actually matters, and for which DFT is not sufficiently accurate, are more numerous than is generally believed. QMC calculations are particularly useful in trying to understand when this is the case and, at the time of writing, it is now believed to be the most accurate available technique which remains applicable to medium-sized and large systems.

This increase in accuracy can be traced to our insistence on retaining the 3N-dimensional many-electron wave function as our basic quantity despite the emphasis normally placed on reducing the dimensionality of the quantum problem. For example the density used in DFT depends, of course, on only *three* independent variables - a somewhat clear improvement on  $6 \times 10^{23}$ . The fundamental point in favour of QMC is that the many-electron wave function satisfies the rather well-known and essentially simple fundamental differential equation<sup>42</sup> discovered by Professor

Schrödinger in 1926, namely  $\hat{H}\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = E\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ . If we wish to reformulate the problem in terms of the density then it is an unfortunate fact that the exact equation satisfied by the ground-state density is *completely unknown*. In DFT, the complicated many-body problem is effectively subsumed in the definition of the exchange-correlation functional whose correct mathematical expression is unlikely ever to be known exactly. The inevitable approximations to this quantity, from the various LDAs referred to above up to the best modern functionals, substantially reduce the attainable accuracy and the predictability of the method.

QMC has an additional advantage over conventional high-accuracy quantum chemistry techniques in that it may be trivially adapted to work in condensed-matter systems such as crystalline solids, rather than just in atoms and molecules. All but the most modern standard solid-state texts in fact routinely deny the possibility of solving the many-particle Schrödinger equation in any meaningful way for large crystalline systems, e.g. in Ashcroft and Mermin's well-known *Solid State Physics*<sup>43</sup> it is explicitly stated - referring to the many-electron Schrödinger equation - that, 'one has no hope of solving an equation such as this' and one must reformulate the problem in such a way as 'to make the one-electron equations least unreasonable.' This ultimately turns out to mean Hartree-Fock theory, i.e. we assume the many-electron wave function can be represented as a single determinant of one-electron orbitals; the solutions to the Hartree-Fock equations then give the optimum form of these orbitals. This presentation is slightly misleading in that the simplifying physical idea that allows one to use QMC, for example, in cases such as these is not the use of one-electron orbitals but rather the imposition of periodic boundary conditions. There can then be any form of many-electron wave function you like (including the possibility of having explicit dependence on the interparticle separations) representing a box of particles embedded in an infinite number of copies of itself. The 'particles' sampling the many-body wave function can be visualized 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 width of the exchange-correlation hole surrounding each electron must be substantially less than the lattice constant 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 this is not the case the calculation may have significant 'finite-size errors' (these must anyway be monitored and corrected for).

How accurate must the total energies be in QMC calculations in order to get chemically useful information? We note that the main goal is often, even usually, the calculation of differences in energy between two arrangements of a set of atoms, giving us the energy barrier to some process, or the energy required to create a defect, or whatever. Under most circumstances, high-quality QMC calculations will give the lowest energy (i.e. they will recover more of the 'correlation energy' missing in Hartree-Fock theory) compared to other variational techniques. This does not guarantee that the *difference* in energy of the two systems, which may be very different from each other, will be similarly accurate. This is of course, a standard problem for all electronic structure methods which in large systems must rely on a cancellation of errors in energy differences. It is well known that, for example, the ordering of binding energies of isomeric atomic clusters in DFT often come out essentially randomly depending on which exchange-correlation functional is used, meaning it is not a reliable technique for answering questions such as '*what is the smallest stable fullerene?*'<sup>44</sup>. For high-quality error cancellation it is required that the error in the energy per atom is proportional to the number of atoms otherwise, for example, a cohesive energy would not have a well-defined limit for large systems. In general, the error cancellation in QMC has been found to be reliable. It has been repeatedly demonstrated that VMC calculations lead to errors which are proportional to the number of atoms, recovering something like 70-90% of the correlation energy independent of system size. In DMC the error is also proportional to the number of atoms but the method can recover up to 100% of the correlation energy in favourable cases. For tractable QMC algorithms we also require that the number of wave function parameters must not increase too rapidly with system size (fortunately this number generally increases only linearly, or at worst quadratically, and the function can be evaluated in a time which rises as a low power of the system size). The wave function must also be easily computable at a given point in the configuration space.

Over the last decade a decent number of software packages have become available for performing QMC calculations of the kind described in this review (a reasonably up-to-date list of such codes is maintained in Ref. 45). Our own efforts in this direction - due to various members of our group in Cambridge University's Cavendish Laboratory - began in 1999 with the publication of the first version of our general-purpose QMC computer program - CASINO<sup>46,47</sup>. This code, which continues to be actively developed, is designed to perform variational and diffusion Monte Carlo calculations on essentially any conceivable system (subject to the obvious memory and time constraints). As well as the usual model systems, it can treat realistic finite objects such as atoms and molecules, or one may impose periodic boundary conditions in one, two, or three dimensions (corresponding to polymers, slabs, and solids) in order to model crystalline systems. We have coded up a large number of different types of interaction, particle and many-body wave function (the latter usually generated through interfaces to appropriate third-party codes). It can also calculate a great many properties in addition to accurate total energies. To complement the CASINO

code and provide instruction in its use, and to counter the widespread lack of knowledge of these methods, we have also for the last five years run a pedagogical programme of annual QMC summer schools in Tuscany, Italy (see Ref. 48).

## II. VARIATIONAL MONTE CARLO

VMC is a relatively straightforward stochastic numerical integration method. It is in principle capable of computing quantum-mechanical expectation values for any many-electron wave function whose value can be evaluated at arbitrary points in its configuration space. Given some trial wave function  $\Psi_T$  satisfying the appropriate boundary conditions one may, for example, simply calculate the total energy as the expectation value of the many-body Hamiltonian operator  $\hat{H}$ ,

$$\frac{\int \Psi_T^*(\mathbf{R}, \{\alpha\}) \hat{H} \Psi_T(\mathbf{R}, \{\alpha\}) d\mathbf{R}}{\int \Psi_T^*(\mathbf{R}, \{\alpha\}) \Psi_T(\mathbf{R}, \{\alpha\}) d\mathbf{R}} = E(\{\alpha\}) \geq E_0, \quad (1)$$

where  $\mathbf{R}$  is a  $3N$ -dimensional vector giving the configuration coordinates  $(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$  of the  $N$  particles in the system (we ignore spin for the moment). The numerical integration is performed by sampling the configuration space of the wave function at random points. Now for a wave function with appropriate properties the usual variational theorem tells us that by evaluating this integral we obtain an energy which is an upper bound to the exact ground-state energy, that is, energies for approximate wave functions are always higher than that of the true ground state. Such wave functions generally depend on some parameters - here collectively denoted by  $\{\alpha\}$  - and these parameters may thus be varied to minimize an objective function such as the energy; in so doing the ‘shape’ of the wave function can be optimized.

Why do we sample the wave function at *random* points? Most people are familiar with ordinary fixed-grid quadrature methods for, say, integrating a regular one-dimensional function  $f(x)$  over some interval. The simplest such method would be to define a regular grid of  $M$  points along the  $x$ -axis, evaluate the function at each of these points, and then use these numbers to calculate the mean value  $\langle f \rangle$  of the function in that interval. The integral - the area under the function - is then just that mean value times the length of the interval. One can define more complicated and efficient schemes, all amounting simply to changing the positions of the points and/or assigning weights to them. The principal difference between such schemes and a Monte Carlo algorithm then is that in the latter the sampling points are chosen at random. This allows us to assume that the well-known *central limit theorem* is valid, and then for large enough  $M$  the estimate of the mean value will be normally distributed. The statistical uncertainty in the mean value  $\langle f \rangle$  is then given by  $\sigma = \sigma_{sample}/\sqrt{M}$  with  $\sigma_{sample} = \sqrt{\langle f^2 \rangle - \langle f \rangle^2}$ , i.e. the error in the integral decreases as the square root of the number  $M$  of sampling points *irrespective of the dimensionality  $d$  of the integral*. For a standard grid method such as the trapezium rule the error decreases as  $\mathcal{O}(M^{-\frac{2}{d}})$ . Monte Carlo wins in more than four dimensions. To see the importance of this, let us take the particular example of a 100-dimensional integral. To make the estimate of this integral ten times more accurate requires one hundred times more work with Monte Carlo integration. With the trapezium rule it would require ten to the power fifty times more work. Now it is the case that when evaluating quantum-mechanical expectation values for an  $N$  particle system we must do  $3N$ -dimensional integrals; thus for high-dimensional numerical integrations such as these there is effectively *no alternative* to Monte Carlo methods. We conclude this point by noting that, though good enough for many purposes, the assumption of the validity of the central limit theorem in this context is only approximate; a good technical discussion of the subtleties involved may be found in two recent papers by Trail<sup>49,50</sup>.

Now in practice we do not wish to sample the random points from a uniform probability distribution - as effectively implied above - but, rather, to group the points in regions where the integrand is finite, and to do this in such a way as to minimize the sample variance. Such *importance sampling* requires us to generate points distributed according to some non-uniform probability distribution  $p(\mathbf{R})$ , following which the calculation of the mean proceeds as usual. This sampling can be accomplished using a random walk moved according to the rules of the *Metropolis algorithm*<sup>51</sup>. We propose random moves taken from some standard distribution (usually Gaussians of appropriate width centred on the current points), always accepting moves to points of higher probability, and occasionally rejecting moves to regions of lower probability according to a particular formula obeying a detailed balance condition. Assuming *ergodicity* - that is, any point in the configuration space can be reached in a finite number of moves - then the distribution of the moving points will converge to the desired  $p(\mathbf{R})$  after some appropriate period of *equilibration*.

To apply this procedure to evaluate an integral like Eq. 1 we need to rewrite the expression slightly so that the integrand looks like a probability distribution times some appropriate function to be evaluated at each point (that

can later be averaged). What is the best probability distribution to use, i.e. what distribution of points will minimize the sample variance? It can be shown that it is  $p_{best}(\mathbf{R}) = |f(\mathbf{R})| / \int |f(\mathbf{R}')| d\mathbf{R}'$ , i.e. we concentrate sampling points in regions where the absolute value of the integrand is large. Note that one cannot do this exactly because in general we do not know the normalization integral in the denominator; the best one can do is to make  $p(\mathbf{R})$  look as much like this as possible. It is readily seen that in our case  $p(\mathbf{R}) = |\Psi_T(\mathbf{R})|^2$  - and, although this is not the best sampling distribution to use according to the above criterion, it is close to it for an approximate eigenstate. We may therefore rewrite the expectation value of the Hamiltonian  $\hat{H}$  with respect to the trial wave function  $\Psi_T$  as

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

where the function to be evaluated -  $E_L(\mathbf{R}) = \frac{\hat{H}(\mathbf{R})\Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})}$  - is known as the *local energy*. If  $\Psi_T$  were in fact the exact ground state wave function note that, according to Schrödinger's equation, the local energy should have a constant value  $E_0$  over the whole of configuration space. For an *approximate* wave function this is no longer the case; a plot of the local energy for 1000 Metropolis-sampled points in the configuration space of an approximate trial function for a hydrogen atom (approximate since it is expanded in a finite Gaussian basis set) might look like this:

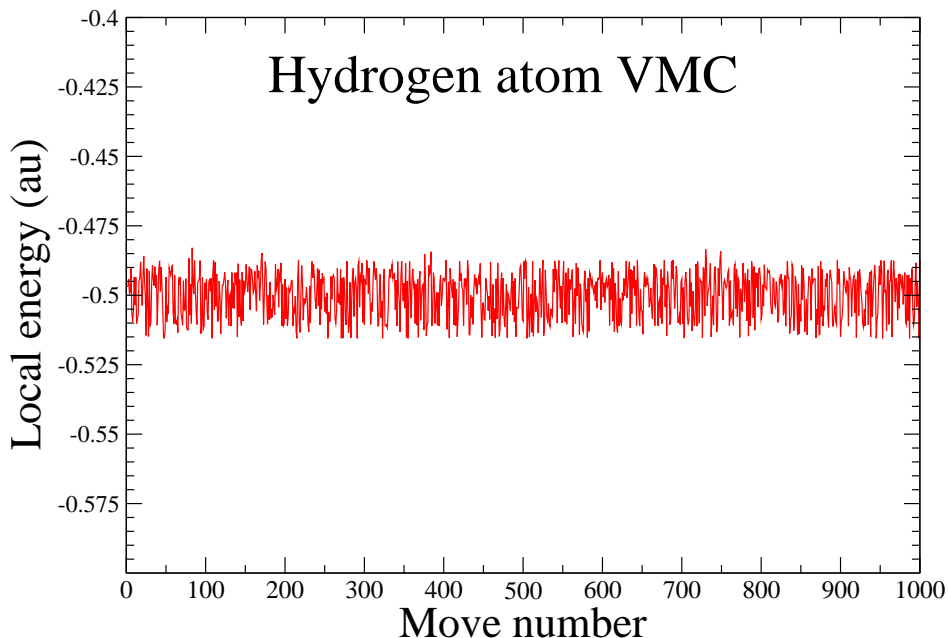


FIG. 1: Local energies of points in the random walk for a VMC run with an approximate wave function. All values are clustered around the true value of  $-0.5$  Ha.

So having used the Metropolis algorithm to generate a sequence of configurations  $\mathbf{R}$  distributed according to  $\Psi_T^2(\mathbf{R})$  we may then compute the desired expectation value by averaging the set of local energies:

$$\langle \hat{H} \rangle = \frac{1}{M} \sum_{i=1}^M E_L(\mathbf{R}_i) = \frac{1}{M} \sum_{i=1}^M \frac{\hat{H}\Psi_T(\mathbf{R}_i)}{\Psi_T(\mathbf{R}_i)}. \quad (3)$$

It should be clear from the figure that for hydrogen the energy thus obtained should (correctly) be  $-0.5$  Ha plus or minus some small error bar. The error bar may need to be refined somewhat by particular statistical techniques to account for serial correlation of the points along the walk. Clearly expectation values other than the energy could be calculated in a similar way.

### III. WAVE FUNCTIONS AND THEIR OPTIMIZATION

As previously implied, VMC is to a large extent a basis set method. You get out what you put in. It is often difficult to prepare wave functions of equivalent quality for two systems whose energy we may wish to compare,

and hence difficult to get an efficient cancellation of errors in the difference. If we had some way of writing down a formula representing the exact wave function for a given system, then great - that would be the end of the story. Unfortunately that can essentially never be done exactly, as we know, and in practice it turns out that the main use of VMC is to *prepare the input* for DMC calculations, which are much more computationally efficient if supplied with a good ‘starting guess’ for the many-body wave function. The better the guess, the more efficient the DMC, so we need now to consider how to represent wave functions and how to improve or optimize them within the limitations of VMC in order to supply that better guess.

In all optimization schemes, one needs a choice of objective function - that is, a function depending on  $\Psi_T$  whose value is to be optimized. We have seen that the energy is one possibility; the variational theorem tells us that the answer will approach the true energy from above as we use better and better wave functions. Another approach is suggested by the ‘zero variance principle’: better wave functions will ‘flatten out’ the wiggly line in Fig. 1. That is, as  $\Psi_T$  approaches an exact eigenstate the fluctuations will become smaller and the local energy  $\frac{\hat{H}\Psi}{\Psi}$  approaches a constant,  $E_0$ , everywhere in configuration space. Hence the sample 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, variance, or possibly some combination of both as the objective function to be minimized.

Now before we consider the process of optimization in more detail it is necessary to discuss how the many-electron wave function is to be represented in our calculations. It is an unfortunate fact that when learning about density functional theory, students are often presented with slightly hysterical motivation (with a generous sprinkling of exclamation marks) for avoiding playing with the wave function at all. For example, we might learn that the wave function..

“ ..is a high dimensional object, dependent on  $3N$  electron Cartesian coordinates,  $N$  electron spin coordinates, and  $3N_n$  nuclear coordinates. For a molecule such as benzene, with 12 nuclei and 42 electrons, this object exists in  $162 - 6 = 156$ -dimensional Cartesian space! The determinant has  $42! = 1.4 \times 10^{51}$  terms! We can’t even make a picture of it. If we were to store it as a numerical object, with a resolution of 0.1 au out to 10 au, we would need  $10^{312}$  numbers to store, at single precision (4 bytes/number), at 2 Gbytes/cm<sup>2</sup> (counting electrical connections access), we would need more than  $10^{293}$  km<sup>2</sup> of surface to store that information. The Earth has a surface area of less than  $10^9$  km<sup>2</sup>. The promised knowledge hidden in the Schrödinger equation is not quite easily accessible! We must make do with much much less. How much less can we do with?”<sup>55</sup>

Faced with the awful prospect of finding  $10^{284}$  planets in order to store the wave function for a single benzene molecule (and forgetting that the Kohn-Sham single determinant wave function for the noninteracting system that he’ll be forced to use in practical DFT calculations has just as much information) the horrified student gives up the game, and grabs at the density with a sigh of relief.

But of course you don’t need to store every possible value of the wave function. From a VMC perspective, you just need to be able to *sample* it. We evaluate the formula for  $\Psi_T$  at a small subset of sampling points 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 at these points? In DMC, where we don’t have a formula for the wave function, we avoid having to do that directly by appropriately *weighting* the calculated values of a trial importance-sampling function (usually just the optimized VMC wave function).

So what formula shall we use for  $\Psi_T$ ? We begin by noting that the simplest basic techniques in quantum chemistry - Hartree-Fock theory, say - give in general approximately the right density but the wrong pair-correlation function. Knowing that with Monte Carlo methods we can simply multiply in appropriate pair-correlation functions this gives us confidence that a good starting point would be the usual 1920s-style legacy of appropriate fictions. First - the fiction that each electron has its own personal wave function (an orbital). The simplest way of making a many-electron wave function out of a set of orbitals is to multiply all the orbitals together (the ‘Hartree product’) but the electrons are then all completely uncorrelated. We then note that if we use one-electron functions as building blocks we require that the many-electron function built from them satisfies the Pauli principle. The simplest way of doing this and thereby making the whole wave function correctly antisymmetric is to say that the many-electron wave function is a *single determinant* of orbitals, that is, an antisymmetrized linear combination of however many Hartree products are necessary to obtain all possible permutations of the electron labels among the orbitals.

It is important to note how we are dealing with spin. Strictly speaking, of course, the Schrödinger equation is for *spinless* particles only. We must therefore *invent* a two-valued ‘property of particles’ called spin that is tacked on to the Schrödinger theory by modifying the form of the wave function. We say there are two kinds of electron - spin-up

and spin-down - and we define two corresponding orthonormal spin-functions (formally, we define ‘spin orbitals’ by multiplying the one-electron spatial wave function by a simple ‘spin function’ with the appropriate properties). The imposed orthogonality introduces a kind of correlation (confusingly referred to as ‘exchange’) which results in same-spin electrons tending to avoid each other. At a point in the configuration space where two electrons with up-spin labels (say) are very close to each other the value of the entire many-electron wave function is correspondingly small; each electron is said to be surrounded by a ‘Fermi hole’ or ‘exchange hole’. In the normal way of doing this we declare that there is a global quantization axis with respect to which the spins are measured, thus allowing us to deal with simple spin orbitals rather than more complicated two-component spinor orbitals; the ‘spin density’ in this sense then becomes a scalar quantity rather than a vector magnetization density.

Solving the usual Hartree-Fock equations gives the optimum form of the spin orbitals - under the usually incorrect assumption that the true many-electron wave function can be accurately represented by a single determinant. To go beyond the Hartree-Fock level, the quantum chemists normally expand the difference between this single-determinant wave function and the exact one using a basis set of appropriate many-electron functions, namely, all possible excited determinants obtained by replacing orbitals occupied in the ground state determinant by unoccupied ones. The fact that literally *billions* of terms are normally required for this expansion to converge tells us that in fact the basis set looks nothing like the function we are trying to expand. This would ordinarily be understood to be a bad thing and, indeed, the approach becomes completely intractable for more than a few tens of electrons. Note also that formally speaking the method (known as ‘full configuration interaction’) only converges to the right answer if the one-electron basis set used to expand the orbitals is infinitely large. We shall adopt a different approach.

An aside: it is instructive to reflect on the nature of spin using an interpretation of quantum mechanics with a realistic ontology such as the de Broglie-Bohm pilot-wave theory<sup>52</sup>. In this interpretation - in apparent agreement with much experimental evidence, if you think about it - ‘wave-particle duality’ is taken literally by assuming that particles *and* a wave field (represented mathematically by the wave function) both have a permanent objective existence. An analysis of the particle probability current then shows us that the field apparently exerts a force on the particles, and that particles guided by a wave undergoing Schrödinger evolution eventually become distributed as the square of the wave over time, even if they were not so distributed initially. Hence the Born rule. Because the potential energy function of the wave field turns out to depend on the spin, it is then evident that spin is not a property of particles at all, but a property of the wave field; in fact it is the *polarization-dependent part of the wave field’s angular momentum*. In a relativistic theory the wave field would be a vector field with states of polarization, and this may be represented by the scalar Schrödinger field with tacked-on spinors. The effects of the Pauli exclusion principle turn out to be due to the influence of the ‘quantum force’ on the electron trajectories (the electrons have trajectories!) and their consequent inability to pass through nodal surfaces in the wave field. Interesting!

So anyway, we have now defined our starting point. How do we then get to properly correlated wave functions which depend explicitly on the distances between particles? We simply multiply what we have already by an appropriate correlation function. The resulting functional form is known as the *Slater-Jastrow wave function*, and this is the most common choice for quantum Monte Carlo calculations. It 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<sup>53,54</sup> 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.

So the basic functional form of the Slater-Jastrow function is

$$\Psi_T(\mathbf{X}) = e^{J(\mathbf{X})} \sum_n c_n D_n(\mathbf{X}), \quad (4)$$

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,

$$D(\mathbf{X}) = \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_1(\mathbf{x}_2) & \dots & \psi_1(\mathbf{x}_N) \\ \psi_2(\mathbf{x}_1) & \psi_2(\mathbf{x}_2) & \dots & \psi_2(\mathbf{x}_N) \\ \vdots & \vdots & \vdots & \vdots \\ \psi_N(\mathbf{x}_1) & \psi_N(\mathbf{x}_2) & \dots & \psi_N(\mathbf{x}_N) \end{vmatrix}. \quad (5)$$

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

$$\psi_\alpha(\mathbf{x}) = \psi_\alpha(\mathbf{r})\delta_{\sigma,\sigma_\alpha}. \quad (6)$$

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,

$$\Psi_T(\mathbf{R}) = e^{J(\mathbf{R})} \sum_n c_n D_n^\uparrow(\mathbf{r}_1, \dots, \mathbf{r}_{N_\uparrow}) D_n^\downarrow(\mathbf{r}_{N_\uparrow+1}, \dots, \mathbf{r}_N), \quad (7)$$

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  $\Psi_T(\mathbf{X})$  for any spin-independent operator.

Now what about the Jastrow factor  $e^{J(\mathbf{R})}$ ? Through its explicit dependence on interparticle distances this directly introduces correlation in such a way that the overall wave function is smaller when particles with repulsive interactions - two electrons, for example - are close together and larger when particles with attractive interactions (such as an electron and a positron) approach each other. The Jastrow is also useful for ensuring that the system obeys the appropriate Kato cusp conditions<sup>56</sup> for two-particle coalescences - a known property of the exact many-body wave function. A correctly-shaped cusp gives rise to a divergence in the local kinetic energy equal and opposite to the divergence in the interparticle Coulomb interaction as two particles move towards each other, which is helpful given that their sum - the local energy - is supposed to be constant. This is particularly important in QMC, where large local-energy divergences can lead to inaccurate statistics or, in DMC, severe numerical instabilities. Note that with regular quantum chemistry methods one does not in general worry about either electron-electron or electron-nuclear cusps, as the wave function forms generally employed do not have the correct structure to represent them. The use of Gaussian basis sets, for example, which have zero gradient at the nucleus, foregoes in principle any possibility of representing a gradient discontinuity there. One can get away with this since the usual methods are effectively *analytically integrating* the wave function, rather than sampling quantities at individual points in its configuration space as in QMC; the small differences in the ‘area under the graph’ due to having an incorrect cusp can be seen to be relatively unimportant.

The full Jastrow function that we have typically used 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 is presented in detail in Ref. 54. Essentially our Jastrow consists of separate electron-electron ( $u$ ), electron-nucleus ( $\chi$ ), and electron-electron-nucleus ( $f$ ) terms which are expanded in appropriate polynomials with optimizable coefficients and are forced to go to zero at some optimizable cutoff radii (as they must do in periodic systems). In systems with no nuclei such as the homogeneous electron gas we might use a much simpler one-parameter Jastrow function, such as

$$J(\mathbf{R}) = - \sum_{i>j}^N u_{\sigma_i,\sigma_j}(r_{ij}) \quad \text{and} \quad u_{\sigma_i,\sigma_j}(r_{ij}) = \frac{A}{r_{ij}} \left( 1 - e^{-r_{ij}/F_{\sigma_i,\sigma_j}} \right). \quad (8)$$

Here  $r_{ij}$  is the distance between the  $N$  electrons  $i$  and  $j$ , the  $\sigma$  subscript is the spin label, and the parameter  $F$  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 variance minimization or whatever. For systems with both electrons and nuclei present one can write a standard Jastrow with all three terms (ignoring the spin dependence for clarity) as follows:

$$J(\mathbf{R}, \{\mathbf{r}_I\}) = \sum_{i>j}^N u(r_{ij}) + \sum_{i=1}^N \sum_{I=1}^{N_I} \chi_I(r_{iI}) + \sum_{i>j}^N \sum_{I=1}^{N_I} f_I(r_{ij}, r_{iI}, r_{jI}) \quad (9)$$

Other terms such as an extra plane-wave expansion in the electron-electron separation for periodic systems, or an additional three-body term, are part of our standard Jastrow<sup>54</sup> and can be useful in certain circumstances but are not usually necessary. For particles with attractive interactions one finds that the usual Slater-Jastrow form is not appropriate and in order to get a better description of exciton formation one might use a determinant of ‘pairing orbitals’ instead<sup>57</sup>. A further recent advance by members of our group has been the development of a completely



general functional form for the Jastrow factor which allows the inclusion of arbitrary higher-order terms (dependent on e.g. the separation of four or more particles); this has now been implemented in our code<sup>58</sup>.

In order to convince yourself that the Slater-Jastrow function is doing what it should, consider Fig. 2. These are the results of simple VMC calculations of the spin-dependent pair-correlation function (PCF) in a silicon crystal with an electron fixed at a bond centre<sup>20</sup>. The figure on the left is for parallel spins - and corresponds to the Fermi or ‘exchange hole’. The figure on the right is for antiparallel spins, and corresponds to the ‘correlation hole’; note that the former is much wider and deeper than the latter. We have here then a pretty illustration of the different levels of theory that we use. In Hartree theory (where we use a Hartree product of all the orbitals as a wavefunction, and which thus corresponds to entirely uncorrelated electrons) then both PCFs would have a value of one everywhere. In Hartree-Fock theory, the left-hand plot would look very similar, but the antiparallel PCF on the right would be one everywhere. The energy lowering over Hartree theory caused by the fact that parallel spin electrons tend to avoid each other is essentially the ‘exchange energy’, which correctly has a negative sign. It is slightly sobering to note that the whole apparatus of quantum chemistry (an expansion in billions of determinants..) is devoted to modelling the little hole on the right and thereby evaluating the correlation energy. In QMC our quantum of solace comes from our *compact representation*; with a Slater-Jastrow function we can do the same thing in VMC using a simple polynomial expansion involving a few tens of parameters, and if this is not accurate enough we can make the necessary minor corrections to it using the DMC algorithm.

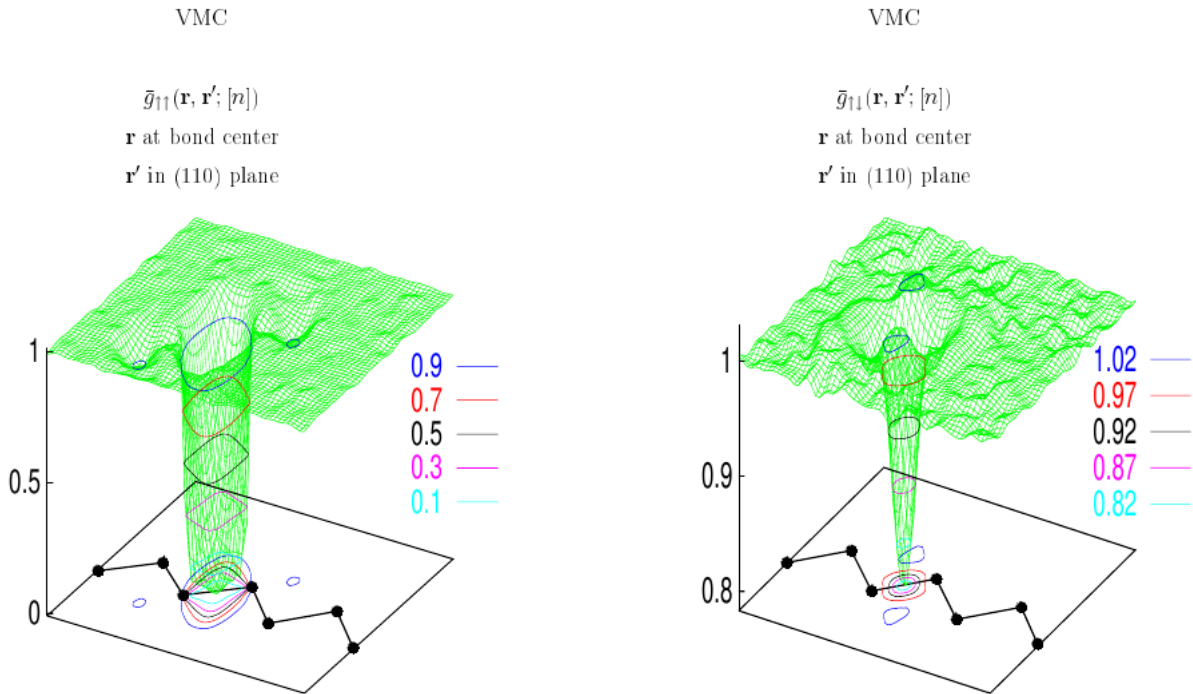


FIG. 2: VMC plots of the pair-correlation function for (on the left) parallel spins, and (on the right) antiparallel spins using a Slater-Jastrow wave function. The data is shown for crystalline silicon in the (110) plane passing through the atoms, and shows the pair-correlation functions around a single electron fixed at a bond centre. The atoms and bonds in the (110) plane are schematically represented. Figure taken from Ref. 20.

However we do not know *a priori* what the shape of the hole is, and we must therefore optimize the various parameters in the Slater-Jastrow function in order to find out. The usual procedure is to leave the Slater determinant part alone and optimize the Jastrow factor. With a full inhomogeneous Jastrow such as that of Eq. 9 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 these terms clearly has a bearing on their ease of optimization. There is of course no absolute prohibition on optimizing the Slater part and one might also envisage, for example, optimization of the coefficients of the determinants of a multi-determinant wave function, or even the orbitals in the Slater determinants themselves (though this latter is quite difficult to do in general, and often pointless). A higher-order technique called *backflow*, to be explained in a subsequent section, also involves functions with optimizable parameters. We thus turn our attention to the technicalities of the optimization procedure.

Now optimization of the wave function is clearly a critical step. It is also a numerically difficult one. It is apparent

that the parameters appear in many different contexts, they need to be optimized in the presence of noise, and there can be a great many of them. As has already been stated, there are two basic approaches. Until recently the most widely used was the optimization of the variance of the energy,

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

where  $E_V$  is the variational energy, with respect to the set of parameters  $\{\alpha\}$ . Now of course there is no reason why one may not optimize the energy directly, and because wave functions corresponding to the minimum energy turn out to have more desirable properties, this has become the preferred approach in the last few years. Historically variance minimization was much more widely used<sup>60,61</sup> - not just for trivial reasons such as the variance having a known lower bound of zero - but most importantly because of the difficulties encountered in designing a robust, numerically-stable algorithm to minimize the energy, particularly in the case of large systems.

First I shall briefly summarize how a simple variance minimization is done. Beginning with an initial set of parameters  $\alpha_0$  (generated, for example, by simply zeroing the Jastrow polynomial coefficients) we proceed to minimize  $\sigma_E^2(\alpha)$  with respect to them. A correlated-sampling approach turns out to be most efficient. First of all a set of some thousands of *configurations* distributed according to  $|\Psi_T^{\alpha_0}|^2$  is generated. Practically speaking, a configuration in this sense is just a ‘snapshot’ of the system taken at intervals during a preliminary VMC run and consists of the current particle positions and the associated interaction energies written on a line of a file. We then calculate the variance in the energies for the full sampled set of configurations. This is the objective function to be minimized. Now unfortunately every time we slightly modify the parameters the wave function changes and our configurations are no longer distributed according to the square of the current  $\Psi_T^\alpha$ , but to the square of the initial wave function  $\Psi_T^{\alpha_0}$ . In principle, therefore, we should regenerate the configurations - a relatively expensive procedure. The correlated sampling is what allows us to avoid this - we reuse the initial set of configurations by simply including appropriate weights  $w$  in the formula for the variance:

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

where

$$E_V(\alpha) = \frac{\int [\Psi_T^{\alpha_0}(\mathbf{R})]^2 w_\alpha^{\alpha_0} E_L^\alpha(\mathbf{R}) d\mathbf{R}}{\int [\Psi_T^{\alpha_0}(\mathbf{R})]^2 w_\alpha^{\alpha_0} d\mathbf{R}}, \quad (12)$$

and the weight factors  $w_\alpha^{\alpha_0}$  are given simply by

$$w_\alpha^{\alpha_0} = \frac{[\Psi_T^\alpha(\mathbf{R})]^2}{[\Psi_T^{\alpha_0}(\mathbf{R})]^2}. \quad (13)$$

The parameters  $\alpha$  are then adjusted until  $\sigma_E^2(\alpha)$  is minimized. This may be done using standard algorithms which perform an unconstrained minimization of a sum of  $m$  squares of functions which contain  $n$  variables (where  $m \geq n$ ) without requiring the derivatives of the objective function - see, for example, Ref. 59. Although in principle we do not need to regenerate the configurations at all, one finds in practice that it usually pays to recalculate them occasionally when the wave function strays very far from its initial value. Generally this needs to be done only a couple of times before we obtain complete convergence within the statistical noise.

There is a problem however. 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 initial parameters and the configuration distribution, so that the optimized parameters would not change over successive cycles. The problem arises from the fact that the weights may vary rapidly as the parameters change, especially for large systems. This can lead to severe numerical instabilities. For example, one or a few configurations acquire an exceedingly large weight, incorrectly reducing the estimate of the variance almost to zero. Somewhat surprisingly perhaps, it usually turns out that the best solution to this is to do without the weights at all, that is, we minimize the *unreweighted variance*. We can do this because the minimum value of the variance (zero) is obtained only if the local energy is constant throughout configuration space, and this is possible only for eigenfunctions of the Hamiltonian. This procedure 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. Reference 62 gives some examples of this for model systems.

It was recognized only relatively recently<sup>62</sup> that one can obtain a huge speed-up in the optimization procedure for parameters that occur linearly in the Jastrow, that is, for Jastrows expressible as  $\sum_{\alpha} \alpha_n f_n(\mathbf{R})$ . These are the most important optimizable parameters in almost all wave functions that we use. The reason this can be done is that the unweighted 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, in particular because we no longer need to generate large numbers of configurations to evaluate the variance. The main *non-linear* parameters in the Jastrow factor are the cutoff lengths where the function is constrained to go to zero. These are important variational parameters, and some attempt to optimize them should always be made. We normally recommend that a (relatively cheap) calculation using the standard variance minimization method should be carried out in order to optimize the cutoff lengths, followed by an accurate optimization of the linear parameters using the fast minimization method. For some systems, good values of the cutoff lengths can be supplied immediately (for example, in periodic systems at high density with small simulation cells, the cutoff length  $L_u$  should be set equal to the Wigner-Seitz radius of the simulation cell).

Let us now move on to outlining the theory of *energy minimization*. We know that - except in certain trivial cases - the usual trial wave function forms cannot in general provide an exact representation of energy eigenstates. The minima in the energy and variance therefore do not coincide. Energy minimization should thus produce lower VMC energies, and although it does not necessarily follow that it produces lower DMC energies, experience indicates that, more often than not, it does. It is also normally stated that the variance of the DMC energy is more or less proportional to the difference between the VMC and DMC energies<sup>63,64</sup>, so one might suppose that energy-optimized wave functions may be more efficient in DMC calculations.

For a long time, efficient energy minimization with QMC was extremely problematic. The methods that have now been developed are based on a well-known technique for finding approximations to the eigenstates of a Hamiltonian. One expands the wave function in some set of basis states,  $\Psi_T(\mathbf{R}) = \sum_{i=1}^N a_i \phi_i(\mathbf{R})$ . Following calculation of the Hamiltonian and overlap matrix elements,  $H_{ij} = \langle \phi_i | \hat{H} | \phi_j \rangle$  and  $S_{ij} = \langle \phi_i | \phi_j \rangle$ , the two-sided eigenproblem  $\sum_j H_{ij} a_j = E \sum_j S_{ij} a_j$  may be solved through standard diagonalization techniques. People have tried to do this in QMC directly<sup>65</sup> but it is apparent that the number of configurations used to evaluate the integrals converges slowly because of statistical noise in the matrix elements. As shown in Ref. 66 however, far fewer configurations are required if the diagonalization is first reformulated as a least-squares fit. Let us assume that the result of operating with  $\hat{H}$  on any basis state  $\phi_i$  is just some linear combination of all the functions  $\phi_i$  (technically speaking, the set  $\{\phi_i\}$  is then said to span an *invariant subspace* of  $\hat{H}$ ). We may thus write (for all  $i$ )

$$\hat{H}\phi_i(\mathbf{R}) = \sum_{j=1}^N A_{ij}\phi_j(\mathbf{R}). \quad (14)$$

To compute the required eigenstates and associated eigenvalues of  $\hat{H}$  we then simply diagonalize the  $A_{ij}$  matrix. Within a Monte Carlo approach we could evaluate the  $\phi_i(\mathbf{R})$  and  $\hat{H}\phi_i(\mathbf{R})$  for  $N$  uncorrelated configurations generated by a VMC calculation and solve the resulting set of linear equations for the  $A_{ij}$ . For problems of interest, however, the assumption that the set  $\{\phi_i\}$  span an invariant subspace of  $\hat{H}$  does not hold and there exists no set of  $A_{ij}$  which solves equation (14). If we took  $N$  configurations and solved the set of  $N$  linear equations, the values of  $A_{ij}$  would depend on which configurations had been chosen. To overcome this problem, a number of configurations  $M \gg N$  is sampled to obtain an overdetermined set of equations which can be solved in a least-squares sense using the singular value decomposition technique. In Ref. 66 it is recommended that equation (14) be divided by  $\Psi_T(\mathbf{R})$  so that in the limit of perfect sampling the scheme corresponds precisely to standard diagonalization.

The method of Ref. 66 is pretty good for linear parameters. How might we generalize it for non-linear parameters? The obvious way is to consider the basis of the initial trial wave function ( $\phi_0 = \Psi_T$ ) and its derivatives with respect to the variable parameters,  $\phi_i = \partial\Psi_T/\partial a_i|_{a_i^0}$ . The simplest such algorithm is in fact unstable and this turns out to be because the implied first-order approximation is often not good enough. To overcome this problem, Umrigar *et al.* introduced a stabilized method<sup>67,68</sup> which works well and is quite robust (the details need not concern us here). The VMC energies given by this method are usually slightly lower than those obtained from variance minimization.

David Ceperley once asked ‘‘How many graduate students’ lives have been lost optimizing wave functions?’’<sup>69</sup>. That was in 1996. To give a more twenty-first century feel for the time scale involved in optimizing wave functions, I can

tell you about the weekend a few years back when I added the entire G2-1 set<sup>70,71</sup> to the examples included with the CASINO distribution. This is a standard set of fifty-five molecules with various experimentally well-characterized properties intended for benchmarking of different quantum chemistry methods see e.g. Ref.72. Grossman has published the results of DMC calculations of these molecules using pseudopotentials<sup>16</sup>, and we have now done the same with all-electron calculations<sup>73,74</sup>. It took a little over three days using only a few single-processor workstations to create all fifty-five sets of example files from scratch including optimizing the Jastrow 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 were all perfectly good enough to be used as input to DMC simulations. 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 certainly 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. 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.

#### IV. 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 has the shape of a big cube 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 does, in essence, is to automatically correct the shape of the guessed square box wave function so that it looks like the correct exponentially-decaying one before calculating the expectation value. In principle it can do this even though our formula for the VMC wave function that we have spent so long justifying turns out not to have enough variational freedom to represent the true wave function. This is clearly a nice trick, particularly when - as is more usual - we have very little practical idea of what the *exact* many-electron wave function looks like.

As one might expect, the DMC algorithm is necessarily rather more involved than that for VMC. I think an approachable way of understanding it is to focus on the properties of quantum-mechanical propagators, so we shall begin by reminding ourselves about these. Let's say we wish to integrate the time-dependent Schrödinger equation,

$$i\hbar \frac{\partial \Psi(\mathbf{R}, t)}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \Psi(\mathbf{R}, t) + V(\mathbf{R}, t) \Psi(\mathbf{R}, t) = \hat{H} \Psi(\mathbf{R}, t), \quad (15)$$

where  $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$ ,  $V$  is the potential energy operator, and  $\nabla = (\nabla_1, \nabla_2, \dots, \nabla_N)$  is the  $3N$ -dimensional gradient operator. Integrating this is equivalent to wanting a formula for  $\Psi$  and, to find this, we must invert this differential equation. The result is an integral equation involving the *propagator*  $K$ :

$$\Psi(\mathbf{R}, t) = \int K(\mathbf{R}, t; \mathbf{R}', t') \Psi(\mathbf{R}', t') d\mathbf{R}'. \quad (16)$$

The propagator is interpreted as the probability amplitude for a particle to travel from one place to another (in this case, from  $\mathbf{R}'$  to  $\mathbf{R}$ ) in a given time  $t - t'$ . It is a *Green's function* for the Schrödinger equation. We see that the probability amplitude for a particle to be at  $\mathbf{R}$  sometime in the future is given by the probability amplitude of it travelling there from  $\mathbf{R}'$  - which is just  $K(\mathbf{R}, t; \mathbf{R}', t')$  - weighted by the probability amplitude of it actually starting at  $\mathbf{R}'$  in the first place - which is  $\Psi(\mathbf{R}', t')$  - summed over all possible starting points  $\mathbf{R}'$ . This is a straightforward concept.

How might we calculate the propagator? A typical way might be to use the *Feynman path-integral* method. For given start and end points  $\mathbf{R}'$  and  $\mathbf{R}$  one gets the overall amplitude by summing the contributions of the infinite number of *all* possible 'histories' or paths which include those points. It doesn't matter why for the moment (look it up!) but the amplitude contributed by a particular history is proportional to  $e^{iS_{cl}/\hbar}$  where  $S_{cl}$  is the classical *action* of that history, i.e. the time integral of the classical Lagrangian  $\frac{1}{2}mv^2 - V$  along the corresponding phase space path of the system. The full expression for the propagator in Feynman's method may then be written as

$$K^F(\mathbf{R}, t; \mathbf{R}', t') = N \sum_{\text{all paths}} \exp \left[ \frac{i}{\hbar} \int_{t'}^t L_{cl}(t'') dt'' \right]. \quad (17)$$

An alternative way to calculate the propagator is to use the de Broglie-Bohm pilot-wave interpretation of quantum mechanics<sup>52</sup>, where the electrons both objectively exist and have the obvious definite trajectories derived from a

straightforward analysis of the streamlines of the quantum-mechanical probability current. From this perspective we find we can achieve precisely the same result as the Feynman method by integrating the *quantum* Lagrangian  $L_q(t) = \frac{1}{2}mv^2 - (V + Q)$  along precisely *one* path - the path that the electron actually follows - as opposed to linearly superposing amplitudes obtained from the *classical* Lagrangian associated with the infinite number of all possible paths. Here  $Q$  is the ‘quantum potential’, which is the potential energy function of the quantum force (the force that the wave field exerts on the electrons). It is easy to show the equivalent pilot-wave propagator is:

$$K^B(\mathbf{R}, t; \mathbf{R}', t') = \frac{1}{J(t)^{\frac{1}{2}}} \exp \left[ \frac{i}{\hbar} \int_{t'}^t L_q(t'') dt'' \right]. \quad (18)$$

where  $J$  is a simple Jacobian factor. This formula should be contrasted with Eq. 17. One should also note that because de Broglie-Bohm trajectories do not cross, one need not sum over all possible starting points  $\mathbf{R}'$  to compute  $\Psi(\mathbf{R}, t)$  - one simply uses the  $\mathbf{R}'$  that the unique trajectory passes through.

What is the connection of all this with the diffusion Monte Carlo method? Well, in DMC an arbitrary starting wave function is evolved using a (Green’s function) propagator just like the ones we have been discussing. The main difference is that the propagation occurs in *imaginary time*  $\tau = it$  as opposed to real time  $t$ . For reasons that will shortly become apparent this has the effect of ‘improving’ the wave function, i.e. making it look more like the ground state as imaginary time passes. For technical reasons, it also turns out that the propagation has to take place in a sequence of very short hops in imaginary time, and so our evolution equation now looks like this:

$$\Psi(\mathbf{R}, \tau + \delta\tau) = \int K^{DMC}(\mathbf{R}, \mathbf{R}', \delta\tau) \Psi(\mathbf{R}', \tau) d\mathbf{R}'. \quad (19)$$

The evolving wave function is not represented in terms of a basis set of known analytic functions but by the distribution in space and time of randomly-diffusing electron positions over an ensemble of copies of the system (‘configurations’). So in other words, the DMC method is a ‘stochastic projector method’ whose purpose is to evolve/project out the solution to the *imaginary-time Schrödinger equation* from an arbitrary starting state. We shall write this equation - which is simply what you get by taking the regular time-dependent equation and substituting  $\tau$  for the time variable  $it$  - in atomic units as

$$-\frac{\partial \Psi^{DMC}(\mathbf{R}, \tau)}{\partial \tau} = -\frac{1}{2} \nabla^2 \Psi(\mathbf{R}, \tau) + (V(\mathbf{R}) - E_T) \Psi(\mathbf{R}, \tau). \quad (20)$$

Here the real variable  $\tau$  measures the progress in imaginary time and, for purposes to be revealed presently, I have included a constant  $E_T$  - an energy offset to the zero of the potential which only affects the wave function normalization.

How then does propagating our trial function in imaginary time ‘improve’ it? For eigenstates, the general solution to the usual time-dependent Schrödinger equation is clearly  $\phi(\mathbf{R}, t) = \phi(\mathbf{R}, 0)e^{-i(\hat{H} - E_T)t}$ . By definition, we may expand an arbitrary ‘guessed’  $\Psi(\mathbf{R}, t)$  in terms of a complete set of these eigenfunctions of the Hamiltonian  $\hat{H}$ :

$$\Psi(\mathbf{R}, t) = \sum_{n=0}^{\infty} c_n \phi_n(\mathbf{R}) e^{-i(E_n - E_T)t}. \quad (21)$$

On substituting  $it$  with imaginary time  $\tau$  the oscillatory time-dependence of the complex exponential phase factors becomes an exponential decay:

$$\Psi(\mathbf{R}, \tau) = \sum_{n=0}^{\infty} c_n \phi_n(\mathbf{R}) e^{-(E_n - E_T)\tau} \quad (22)$$

Let us assume our initial guess for the wave function is not orthogonal to the ground state (i.e.  $c_0 \neq 0$ ). Then if we magically choose the constant  $E_T$  to be the ground state eigenvalue  $E_0$  (or, in practice, keep very tight control of it through some kind of feedback procedure) then it is clear we should eventually get *imaginary-time independence* of the probability distribution, in the sense that as  $\tau \rightarrow \infty$ , our initial  $\Psi(\mathbf{R}, 0)$  comes to look more and more like the stationary ground state  $\phi_0(\mathbf{R})$  as the contribution of the excited-state eigenfunctions dies away:

$$\Psi(\mathbf{R}, \tau) = c_0 \phi_0 + \sum_{n=1}^{\infty} c_n \phi_n(\mathbf{R}) e^{-(E_n - E_0)\tau}. \quad (23)$$

So now we know *why* we do this propagation, how in practice do we find an expression for the propagator  $K$ ? Consider now the imaginary-time Schrödinger equation in two parts:

$$\frac{\partial \Psi(\mathbf{R}, \tau)}{\partial \tau} = \frac{1}{2} \nabla^2 \Psi(\mathbf{R}, \tau) \quad (24)$$

$$\frac{\partial \Psi(\mathbf{R}, \tau)}{\partial \tau} = -(V(\mathbf{R}) - E_T) \Psi(\mathbf{R}, \tau). \quad (25)$$

These two formulae respectively have the form of the usual diffusion equation and of a rate equation with a position-dependent rate constant. The appropriate propagator for the diffusion equation is well-known; it is a  $3N$ -dimensional Gaussian with variance  $\delta\tau$  in each dimension. The propagator for the rate equation is also known - it gives a so-called ‘branching factor’ which can be interpreted as a position-dependent weight/stochastic survival probability for a member of an ensemble. Multiplying the two together to get the following propagator for the imaginary-time Schrödinger equation is an approximation - the ‘*short time approximation*’ - valid only in the limit of small  $\delta\tau$  (which is why we need to do the evolution as a sequence of short hops):

$$K^{DMC}(\mathbf{R}, \mathbf{R}', \delta\tau) = \frac{1}{(2\pi\delta\tau)^{\frac{3N}{2}}} \exp\left(-\frac{|\mathbf{R} - \mathbf{R}'|^2}{2\delta\tau}\right) \exp\left[-\delta\tau \left(\frac{V(\mathbf{R}) + V(\mathbf{R}') - 2E_T}{2}\right)\right]. \quad (26)$$

Let us then summarize with a simple example how the DMC algorithm works. If we interpret  $\Psi$  as a *probability density*, then the diffusion equation  $\frac{\partial \Psi}{\partial \tau} = \frac{1}{2} \nabla^2 \Psi$  represents the movement of  $N$  diffusing particles. If we turn this around we may decide to *represent*  $\Psi(\mathbf{x}, \tau)$  by an *ensemble* of such sets of particles. Each member of such an ensemble will be called a ‘configuration’. We interpret the full propagator  $K^{DMC}(\mathbf{R}, \mathbf{R}', \delta\tau)$  as the probability of a configuration moving from  $\mathbf{R}'$  to  $\mathbf{R}$  in a time  $\delta\tau$ . The branching factor in the propagator will generally be interpreted as a stochastic survival probability for a given configuration rather than as a simple weight, as the latter is prone to numerical instabilities. This means that the configuration population becomes dynamically variable; configurations that stray into regions of high  $V$  have a good chance of being killed (removed from the calculation); in low  $V$  regions configurations have a high probability of multiplying (i.e. they create copies of themselves which then propagate independently). It is solely this branching or reweighting that ‘changes the shape of the wave function’ as it evolves. So, as we have seen, after a sufficiently long period of imaginary-time evolution all the excited states will decay away leaving only the ground-state wave function, at which point the propagation may be continued in order to accumulate averages of interesting observables.

As a simple example, consider Fig 3. Here we make a deliberately bad guess that the ground-state wave function for a single electron in a harmonic potential well is a constant in the vicinity of the well and zero everywhere else. We begin with seven copies of the system or configurations in our ensemble; the electrons in this ensemble are initially randomly distributed according to the uniform probability distribution in the region where the trial function is finite. The particle distribution is then evolved in imaginary time according to the scheme developed above. The electrons are subsequently seen to become distributed according to the proper Gaussian shape of the exact ground-state wave function. It is evident from the figure that the change in shape is produced by the branching factor occasionally eliminating configurations in high  $V$  regions and duplicating them in low  $V$  regions..

This ‘pure DMC’ algorithm works very well in a single-particle system with a nicely-behaved potential, as in the example. Unfortunately it suffers from two very serious drawbacks which become evident in multi-particle systems with divergent Coulomb potentials.

The first problem arises due to our assumption that  $\Psi$  is a probability distribution - necessarily positive everywhere - even though the antisymmetric nature of multi-particle fermionic wave functions means that it must have both positive and negative parts separated by a ‘nodal surface’, that is, a  $3N - 1$ -dimensional hypersurface on which it has the value zero. One might think that two separate populations of configurations with attached positive and negative weights might get round this problem (essentially the well-known ‘fermion sign problem’) but in practice there is a severe signal-to-noise issue. It is possible to construct formally-exact algorithms of this nature which overcome some of the worst practical problems<sup>75</sup> but to date all seem highly inefficient with poor system size scaling.

The second problem is less fundamental but in practice very severe. The required rate of removing or duplicating configurations diverges when the potential energy diverges (which occurs whenever two particles are coincident) due to the presence of  $V$  in the branching factor of Eq. 26. This leads to stability problems and poor statistical behaviour.

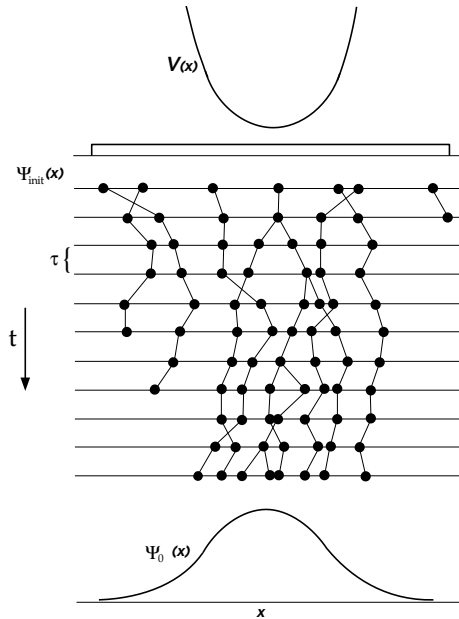


FIG. 3: Schematic illustration of the DMC algorithm for a single electron in a harmonic potential well, showing the evolution of the shape of the wave function due to propagation in imaginary time. Figure taken from Ref. 5.

These problems may be dealt with at the cost of introducing the most important approximation in the DMC algorithm: the *fixed-node approximation*<sup>76</sup>. We say, in effect, that particles may not cross the nodal surface of the *trial* wave function  $\Psi_T$ , that is, there is an infinite repulsive potential barrier on the nodes. This forces the DMC wave function  $\Psi$  to be zero on that hypersurface. If the nodes of the trial function coincide with the exact ones, then such an algorithm will give the exact ground-state energy (it is of course well-known that the exact de Broglie-Bohm particle trajectories cannot pass through the nodal surface). If the trial function nodes do not coincide with the exact ones then the DMC energy will be higher than the ground-state energy (but less than or equal to the VMC energy). The variational principle thus applies.

To make such an algorithm efficient we must introduce importance sampling, and this is done in the following way. We require that the imaginary-time evolution produces the *mixed distribution*  $f = \Psi_T \Psi$ , rather than the pure distribution. Substituting this into the imaginary time Schrödinger equation Eq. 20 we obtain

$$-\frac{\partial f(\mathbf{R}, \tau)}{\partial \tau} = -\frac{1}{2} \nabla^2 f(\mathbf{R}, \tau) + \nabla \cdot [\mathbf{v}_D(\mathbf{R}) f(\mathbf{R}, \tau)] + (E_L(\mathbf{R}) - E_T) f(\mathbf{R}, \tau), \quad (27)$$

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

$$\mathbf{v}_D(\mathbf{R}) = \nabla \ln |\Psi_T(\mathbf{R})| = \frac{\nabla \Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})}, \quad (28)$$

and

$$E_L(\mathbf{R}) = \Psi_T^{-1} \left( -\frac{1}{2} \nabla^2 + V(\mathbf{R}) \right) \Psi_T, \quad (29)$$

is the usual local energy. The propagator from  $\mathbf{R}'$  to  $\mathbf{R}$  for the importance sampled algorithm now looks like this:

$$K^{DMC}(\mathbf{R}, \mathbf{R}', \delta\tau) = \frac{1}{(2\pi\delta\tau)^{\frac{3N}{2}}} \exp \left[ -\frac{(\mathbf{R} - \mathbf{R}' - \delta\tau \mathbf{F}(\mathbf{R}'))^2}{2\delta\tau} \right] \exp \left[ -\frac{\delta\tau}{2} (E_L(\mathbf{R}) + E_L(\mathbf{R}') - 2E_T) \right]. \quad (30)$$

Because the nodal surface of  $\Psi$  is constrained to be that of  $\Psi_T$  then their product  $f$  is positive everywhere and can now be properly interpreted as a probability distribution. The time evolution generates the distribution  $f = \Psi_T \Psi$ ,

where  $\Psi$  is now the lowest energy wave function with the same nodes as  $\Psi_T$ . This solves the first of our two problems. The second problem of the poor statistical behaviour due to the divergences in the potential energy is also solved because the term  $(V(\mathbf{R}) - E_T)$  in Eq. 20 has been replaced by  $(E_L(\mathbf{R}) - E_T)$  in Eq. 27 which is much smoother. Indeed, if  $\Psi_T$  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  $\Psi_T$  it is possible to eliminate the local energy divergences due to coincident particles by choosing a trial function which has the correct cusp-like behaviour at the relevant points in the configuration space<sup>56</sup>. Note that this is all reflected in the branching factor of the new propagator of Eq. 30.

The nodal surface partitions the configuration space into regions that we call ‘nodal pockets’. The fixed-node approximation implies that we are restricted to sampling only those nodal pockets that are occupied by the initial set of configurations, and this appears to introduce some kind of ergodicity concern since at first sight it seems that we ought to sample every nodal pocket. This would be an impossible task in large systems. However, the *tiling theorem* for exact fermion ground states<sup>77,78</sup> asserts that all nodal pockets are in fact equivalent and related by permutation symmetry; one need therefore only sample one of them. This theorem is intimately connected with the existence of a variational principle for the DMC ground state energy<sup>78</sup>. Other interesting investigations of properties of nodal surfaces have been published.<sup>79–81</sup>

A practical importance-sampled 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 then evolved according to the short-time approximation to the Green function of the importance-sampled imaginary-time Schrödinger equation (Eq. 27), which involves repeated steps of biased diffusion followed by the deletion and/or duplication of configurations. The bias in the diffusion is caused by the drift vector arising out of the importance sampling which directs the sampling towards parts of configuration space where  $|\Psi_T|$  is large (i.e. it plays the role of an Einsteinian *osmotic velocity*). This drift step is always directed away from the node, and  $\nabla\Psi_T$  is in fact a normal vector of the nodal hypersurface. 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. Note that throughout this process the reference energy  $E_T$  is varied to keep the configuration population under control through a specific feedback mechanism. The initial stages of a DMC simulation - for solid antiferromagnetic NiO crystal with 128 atoms per cell using unrestricted Hartree-Fock trial functions of the kind discussed in Ref. 82 and 83 - are shown in Fig. 4.

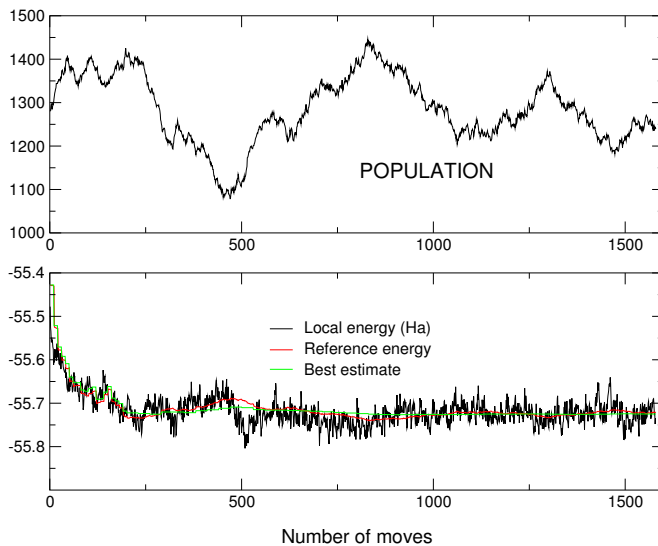


FIG. 4: DMC simulation of solid antiferromagnetic NiO. In the lower panel, the noisy black line is the local energy after each move, the smoother green (possibly grey) line is the current best estimate of the DMC energy, and the red (slightly darker grey) line is  $E_T$  in Eqn. 27 which is varied to 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 which is 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.



The DMC energy is given by

$$E_{DMC} = \frac{\int f(\mathbf{R})E_L(\mathbf{R}) d\mathbf{R}}{\int f(\mathbf{R}) d\mathbf{R}} \approx \sum_i E_L(\mathbf{R}_i). \quad (31)$$

This energy expression would be exact if the nodal surface of  $\Psi_T$  were exact, and the fixed-node error is second order in the error in the  $\Psi_T$  nodal surface (when a variational theorem exists<sup>78</sup>). The accuracy of the fixed-node approximation can be tested on small systems and normally leads to very satisfactory results. The trial wave function thus limits the final accuracy that can be obtained 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. For other expectation values of operators that do not commute with the Hamiltonian then the DMC mixed estimator is biased and other techniques are required in order to sample the pure distribution<sup>84-86</sup>.

A final point: the necessity of using the fixed-node approximation suggests that the best way of optimizing wave functions would be to do it in DMC directly. The nodal surface could then in principle be optimized to the shape which minimizes the DMC energy. The backflow technique discussed in Sec. V A has some bearing on the problem, but the usual procedure involving optimization of the energy or variance in VMC will not usually lead to the optimal nodes in the sense that the fixed-node DMC energy is minimal. The large number of parameters - up to a few hundred - in your typical Slater-Jastrow(-backflow) wave function means that direct variation of the parameters in DMC is too expensive (though this has been done, see e.g. Refs. 87,88). Furthermore, we note that optimizing the energy in DMC is tricky for the nodal surface as the contribution of the region near the nodes to the energy is small. More exotic ways of optimizing the nodes are still being actively developed<sup>89,90</sup>.

## V. BITS AND PIECES

### A. More about wave functions, orbitals and basis sets

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? One might naturally consider optimizing the orbitals themselves. This has been done, for example, with the atomic orbitals of a neon atom by Drummond *et al.*<sup>91</sup> optimizing a parametrized function that is added to the self-consistent orbitals. 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 technique which has been tried<sup>92,93</sup> is to optimize the potential that generates the orbitals rather than the orbitals themselves. It was also suggested by Grossman and Mitas<sup>94</sup> that another way to improve the orbitals over the Hartree-Fock form is to use a determinant of the *natural orbitals* which diagonalize the one-electron density matrix. While the motivation here is that the convergence of configuration interaction expansions is improved by using natural orbitals instead of Hartree-Fock orbitals, it is not clear why this would work in QMC. The calculation of reasonably accurate natural orbitals costs a lot and such an approach is therefore 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. 62.

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 molecules and even in periodic calculations such as the study of the neutral vacancy in diamond due to Hood *et al.*<sup>27</sup>. 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<sup>91</sup>. Of course there is very little point in using methods that use expansions of large numbers of determinants to generate QMC trial functions, not only because using methods which scale so badly as a preliminary calculation completely defeats the whole point of QMC, but because the medium and short-range correlation which these expansions describe<sup>95,96</sup> is dealt with directly and vastly more efficiently by the Jastrow factor.

By far the most useful method for going beyond the Slater-Jastrow form is the *backflow* technique that we have already alluded to. Backflow correlations were originally derived from a current conservation argument by Feynman<sup>97</sup>, and Feynman and Cohen<sup>98</sup> to provide a picture of the excitations in liquid <sup>4</sup>He and the effective mass of a <sup>3</sup>He impurity in <sup>4</sup>He. In a modern context they can also be derived from an imaginary-time evolution argument<sup>99,100</sup>. In the simplest form of backflow trial function the electron coordinates  $\mathbf{r}_i$  appearing in the Slater determinants of Eq. (7) are replaced by *quasiparticle coordinates*,

$$\bar{\mathbf{r}}_i = \mathbf{r}_i + \sum_{j \neq i}^N \eta(r_{ij})(\mathbf{r}_i - \mathbf{r}_j), \quad (32)$$

where  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ . This is supposed to represent the characteristic ‘flow pattern’ where the quantum fluid is ‘pushed out of the way’ in front of a moving particle and fills in the space behind it. 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<sup>99,101</sup> 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<sup>102</sup>. For real polyatomic systems a much more complicated inhomogeneous backflow function is required; the one developed in our group and implemented in the CASINO program by López Ríos<sup>103</sup> has the following functional form:

$$\Psi_{\text{BF}}(\mathbf{R}) = e^{J(\mathbf{R})} \det \left[ \psi_i(\mathbf{r}_i^\uparrow + \xi_i(\mathbf{R})) \right] \det \left[ \psi_i(\mathbf{r}_j^\downarrow + \xi_j(\mathbf{R})) \right], \quad (33)$$

with the backflow displacement for electron  $i$  in a system of  $N$  electrons and  $N_n$  nuclei given by

$$\xi_i = \sum_{j \neq i}^N \eta_{ij} \mathbf{r}_{ij} + \sum_I^{N_{\text{ion}}} \mu_{iI} \mathbf{r}_{iI} + \sum_{j \neq i}^N \sum_I^{N_{\text{ion}}} (\Phi_i^{jI} \mathbf{r}_{ij} + \Theta_i^{jI} \mathbf{r}_{iI}). \quad (34)$$

Here  $\eta_{ij} = \eta(r_{ij})$  is a function of electron-electron separation,  $\mu_{iI} = \mu(r_{iI})$  is a function of electron-ion separation, and  $\Phi_i^{jI} = \Phi(r_{iI}, r_{jI}, r_{ij})$  and  $\Theta_i^{jI} = \Theta(r_{iI}, r_{jI}, r_{ij})$ . The functions  $\eta$ ,  $\mu$ ,  $\Phi$  and  $\Theta$  are parametrized using power expansions with optimizable coefficients<sup>103</sup>.

Now of course the use of backflow wave functions can significantly increase the cost of a QMC calculation. 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 (assuming localized orbitals and basis set) is thus  $N^3$  rather than  $N^2$ . 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 in the sense that the number of moves required to obtain a fixed error in the energy is smaller. In our Ne atom calculations<sup>91</sup>, 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 calculation to fixed error bar increased only by a factor between two and three. 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 (though the problem with obtaining accurate energy *differences* in VMC presumably remains). Finally, it should be

noted that backflow is expected to improve the QMC estimates of all expectation values, not just the energy. We like it.

We now move on to consider the issue of basis sets. 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 (such as CASTEP<sup>104</sup>, ABINIT<sup>105</sup>, and PWSCF<sup>106</sup>) are useless for generating trial wave functions for CASINO; a post-processing utility can be used to re-expand a function expanded in plane-waves in another localized basis before the wave function is read into CASINO. The usual thing here is to use some form of localized spline functions on a grid such as ‘blip’ functions<sup>107,108</sup>.

Another reasonable way to do this is to expand the orbitals in a basis of Gaussian-type functions. These are localized, relatively 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 probably the various versions of the GAUSSIAN software<sup>3</sup>. In addition to the regular single-determinant methods, these codes implement various techniques involving multi-determinant correlated wave functions and are flexible tools for developing accurate molecular trial wave functions. For systems with periodic boundary conditions, the Gaussian basis set program CRYSTAL<sup>109</sup> turns out to be very useful; it can perform all-electron or pseudopotential Hartree-Fock and DFT calculations both for molecules and for systems periodic in one, two, or three dimensions. For some systems, Slater basis sets may be useful in QMC (since they provide a more compact representation than Gaussians, and hence more rapidly calculable orbitals)<sup>74</sup>. To this end, we have implemented an interface to the program ADF<sup>110</sup>.

There is one more issue we must consider that is relevant to all basis sets but particular to the case of Gaussian-type functions. This has to do with *cusplike conditions*. 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 will be greatly reduced. It is relatively easy to produce an accurate representation of this cusp when using a grid-based numerical representation of the orbitals. However, as we have already remarked, such representations cannot really be used for large polyatomic systems because of the excessive storage requirements and we would prefer to use a Gaussian basis set. But 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; these need to be applied at each nucleus for every orbital which is larger than a given tolerance at that nucleus. The scheme we developed to correct for this is outlined in Ref. 73. Generalizations of this method have been developed for other basis set types.

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 Figure 5) 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 (or

grey) 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 looking at the bottom two panels in Figure 5, which show the local energy as a function of move number for a carbon monoxide molecule with and without cusp corrections.

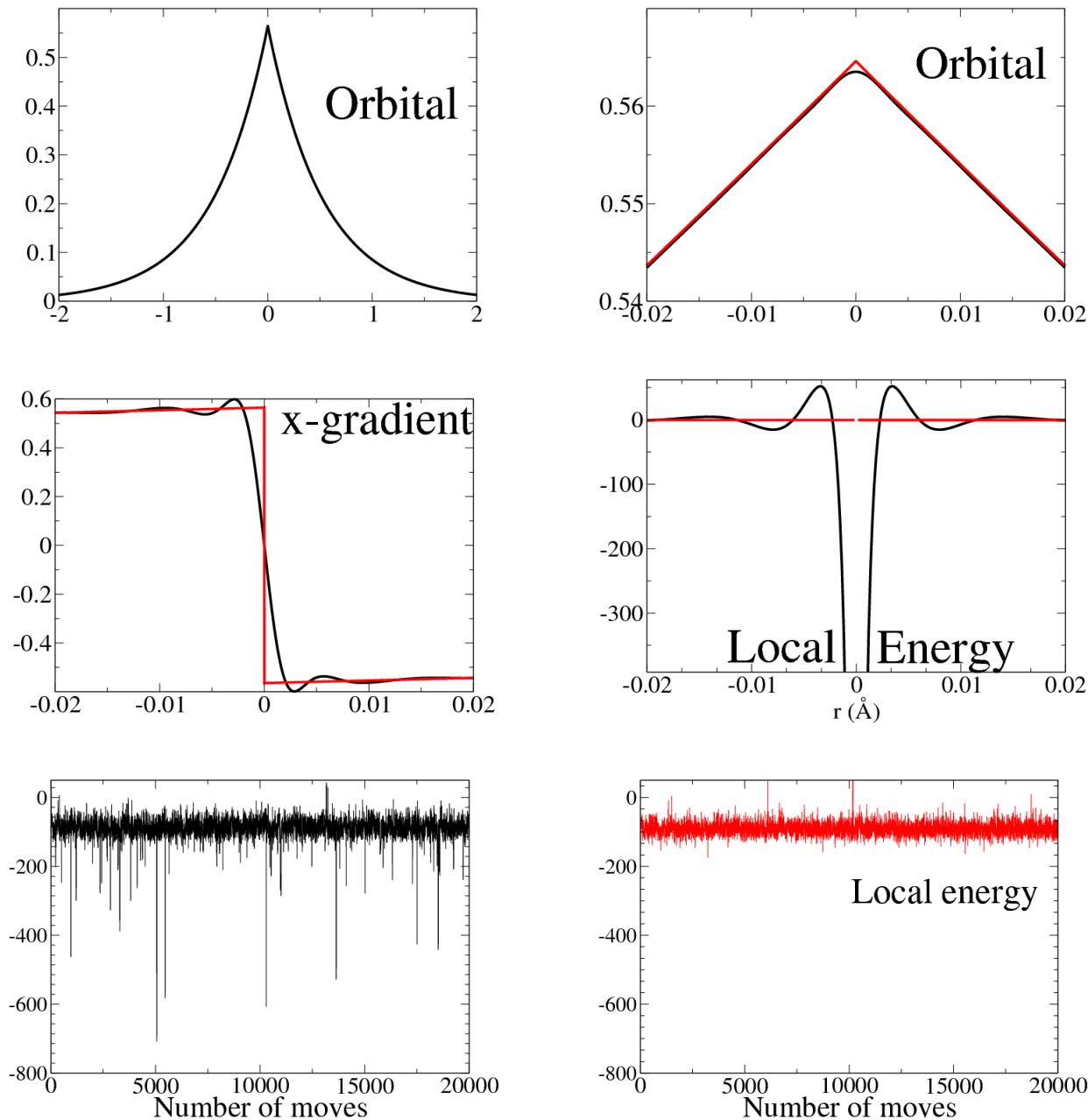


FIG. 5: The top two rows show the effect of Gaussian basis set cusp corrections in the hydrogen atom (red/grey lines corrected; black lines not). The bottom row shows local energy as a function of move number in a VMC calculation for a carbon monoxide molecule with a standard reasonably good Gaussian 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.

The problem with electron-nucleus cusps is clearly more significant for atoms of higher atomic number. In order to understand how this helps 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<sup>64</sup>. By ensuring that the electron-nucleus cusps were accurately represented it proved perfectly possible to produce converged DMC energies with acceptably small error bars for atoms up to xenon ( $Z=54$ ).

## B. Pseudopotentials

Well *perfectly possible* I said. Possible, maybe, but definitely somewhat tiresome. On trying to do all-electron calculations for heavier atoms than xenon we were quickly forced to stop when smoke was observed coming out of the side of the computer<sup>111</sup>. Might it therefore be better to do heavy atoms using pseudopotentials, as is commonly done with other methods such as DFT? In electronic structure calculations pseudopotentials or effective core potentials are used to remove the inert core electrons from the problem and to improve the computational efficiency. Although QMC scales very favourably with system size in general it has been estimated<sup>63</sup> that the scaling of all-electron calculations with the atomic number  $Z$  is approximately  $Z^{5.5}$  which in the relatively recent past was generally considered to rule out applications to atoms with  $Z$  greater than about ten. Our paper<sup>64</sup> pushing all-electron QMC calculations to  $Z = 54$  was therefore a significant step. The use of a pseudopotential then serves to reduce the effective value of  $Z$  and to improve the scaling to  $Z^{3.5}$ . Although errors are inevitably introduced, the gain in computational efficiency is easily sufficient to make pseudopotentials preferable in heavier atoms. They also offer a simple way to incorporate approximate relativistic corrections.

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. In VMC at least this problem can in principle be somewhat reduced by the use of acceleration schemes<sup>112,113</sup>. 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

$$V_{\text{loc}} + \hat{V}_{\text{nl}} = \sum_i V_{\text{loc}}^{\text{ps}}(r_i) + \sum_i \hat{V}_{\text{nl},i}^{\text{ps}}, \quad (35)$$

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

$$\hat{V}_{\text{nl},i}^{\text{ps}} g(\mathbf{r}_i) = \sum_l V_{\text{nl},l}^{\text{ps}}(r_i) \sum_{m=-l}^l Y_{lm}(\Omega_{\mathbf{r}_i}) \int Y_{lm}^*(\Omega_{\mathbf{r}'_i}) g(\mathbf{r}'_i) d\Omega'_i, \quad (36)$$

where the angular integration is over the sphere passing through the  $\mathbf{r}_i$ . 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

$$\hat{V}_{\text{nl},i}^{\text{ps}} g(\mathbf{r}_i) = \sum_l V_{\text{nl},l}^{\text{ps}}(r_i) \frac{2l+1}{4\pi} \int P_l[\cos(\theta'_i)] g(\mathbf{r}'_i) d\Omega'_i, \quad (37)$$

where  $P_l$  denotes a Legendre polynomial.

While the use of non-local pseudopotentials is relatively straightforward in a VMC calculation<sup>115,116</sup>, there is an issue with DMC. The fixed-node boundary condition turns out not to be compatible with the non-locality. This forces us to introduce an additional approximation (the ‘locality approximation’<sup>117</sup>) whereby the non-local pseudopotential operator  $\hat{V}_{\text{nl}}$  acts on the trial function rather than the DMC wave function, that is, we replace  $\hat{V}_{\text{nl}}$  by  $\Psi_T^{-1} \hat{V}_{\text{nl}} \Psi_T$ . The leading-order error term is proportional to  $(\Psi_T - \Psi_0)^2$  where  $\Psi_0$  is the exact fixed-node ground state wave function<sup>117</sup>. Unfortunately this error may be positive or negative and so the method is no longer strictly variational. An alternative to this approximation is the ‘semi-localization’ scheme for DMC non-local pseudopotentials introduced by Casula *et al.* in 2005<sup>118,119</sup>; as well as restoring the variational property this method appears to have better numerical stability than the older scheme.

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<sup>114</sup>. 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<sup>120</sup>, 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. 121).

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 and it well-known that this can lead to significant time step errors in DMC calculations<sup>120</sup>. It was thus apparent a few years ago that none of the available sets were ideal for QMC calculations and it was decided that it would be helpful if we generated an on-line periodic table of smooth non-divergent Hartree-Fock pseudopotentials (with relativistic corrections) developed specifically for QMC. This project has now been completed and it is described in detail by Trail and Needs in Refs. 122 and 123. The resulting pseudopotentials are available online at the website specified in Ref. 124; the repository includes both Dirac-Fock and Hartree-Fock potentials, and a choice of small or large core potentials (the latter being more amenable to plane-wave calculations). Burkatzki *et al.* have since developed another set of pseudopotentials also intended for use in QMC calculations<sup>125</sup>. Although data is limited, tests<sup>126,127</sup> appear to show that the Trail-Needs pseudopotentials give essentially the same results as the Burkatzki ones, though the smaller core radii of the former appears to lead to a slight increase in efficiency.

### C. Periodic systems

As with other methods, QMC calculations for extended systems may be performed using finite clusters or infinitely-large crystals with periodic boundary conditions. The latter are generally preferred because they approximate the desired large-size limit (i.e. the infinite system size *without* periodic boundary conditions) more closely. One can also use the standard supercell approach for *aperiodic* systems such as point defects. For such cases, cells containing a point defect and a small part of the host crystal are repeated periodically throughout space; the supercell must clearly be made large enough so the interactions between defects in different cells are negligible.

In periodic DFT calculations the charge density and potentials are taken to have the periodicity of a suitably-chosen lattice. The single-particle orbitals can then be made to obey Bloch’s theorem and the results for the infinite system are obtained by summing quantities obtained from the different Bloch wave vectors within the first Brillouin zone. The situation with many-particle wave functions is rather different, since it is not possible to reduce the problem to solving within a primitive unit cell. Such a reduction is allowed in single-particle methods because the Hamiltonian is invariant under the translation of a single electronic coordinate by a translation vector of the primitive lattice, but this is not a symmetry of the many-body Hamiltonian<sup>128,129</sup>. Consequently QMC calculations must be performed at a single  $\mathbf{k}$ -point. This normally gives a poor approximation to the result for the infinite system, unless one chooses a pretty large non-primitive simulation cell. One may also average over the results of QMC calculations done at different single  $\mathbf{k}$ -points<sup>130</sup>.

There are also a number of problems associated with the long-range Coulomb interaction in many-body techniques such as QMC. It is well-known that simply summing the  $1/r$  interaction out over cells on the surface of an ever-expanding cluster never settles down because of the contribution from shape-dependent arrangements of surface charge. The usual solution to this problem is to employ the Ewald method<sup>131</sup>. The Ewald interaction contains an effective ‘depolarization field’ intended to cancel the field produced by the surface charges (and is thus equivalent to what you get if you put the large cluster in a medium of infinite dielectric constant). Long-ranged interactions also induce long-ranged exchange-correlation interactions, and if the simulation cell is not large enough these effects are described incorrectly. Such effects are absent in local DFT calculations because the interaction energy is written in terms of the electronic charge density, but Hartree-Fock calculations show very strong effects of this kind and various ways to accelerate the convergence have been developed. The finite-size effects arising from the long-ranged interaction can be divided into potential and kinetic energy contributions<sup>132,133</sup>. The potential-energy component can be removed from

the calculations by replacing the Ewald interaction by the so-called model periodic Coulomb (MPC) interaction<sup>134–136</sup>. Recent work has added substantially to our understanding of finite-size effects, and theoretical expressions have been derived for them<sup>132,133</sup>, but at the moment it seems that they cannot entirely replace extrapolation procedures. An alternative approach to estimating finite-size errors in QMC calculations has been developed recently<sup>137</sup>. DMC results for the 3D homogeneous electron gas are used to obtain a system-size-dependent local density approximation functional. The correction to the total energy is given by the difference between the DFT energies for the finite-sized and infinite systems. This approach is interesting, although it does rely on the LDA giving a reasonable description of the system. As will be shown later, DMC calculations using periodic boundary conditions with thousands of atoms per cell have now been done, and the technology is clearly approaching maturity.

#### D. Differences, derivatives and forces

Calculations in computational electronic structure theory almost always involve the evaluation of *differences* in energy, and all methods that work in complex systems rely for their accuracy on the cancellation of errors in such energy differences. Apart from the statistical ones, all known errors in DMC have the same sign and partially cancel out in the subtraction because the method is variational. That said, incomplete cancellation of nodal errors is the most important source of error in DMC results, even though DMC often retrieves 95% or more of the correlation energy. *Correlated sampling*<sup>138</sup> is one way of improving the computation of the energy difference between two similar systems with a smaller statistical error than those obtained for the individual energies. This is relatively straightforward in VMC, and a version of it was briefly described in the earlier section on variance minimization.

Now as well as simple differences, we would quite often like to calculate *derivatives*. Many quantities of physical interest can be formulated as an energy derivative and thus an ability to calculate them accurately in QMC considerably enhances the scope of the method. Normally, of course, this sort of thing would be encountered in the calculation of forces on atoms but if we expand the energy in a Taylor series in a perturbation such as the strength of an applied electric field, for example, then the coefficients of the first- and second-order terms respectively give the *dipole moment* and the various elements of the *dipole polarizability tensor*:

$$E(F_i) = E(0) + \underbrace{F_i \left[ \frac{\partial E}{\partial F_i} \right]_{F_i=0}}_{\text{dipole moment}} + \frac{1}{2} \sum_{j=1}^3 F_i F_j \underbrace{\left[ \frac{\partial^2 E}{\partial F_i \partial F_j} \right]_{F_i=0, F_j=0}}_{\text{dipole polarizability tensor}} + \dots \quad (38)$$

One may also calculate the dipole moment (no surprise) by evaluating the expectation value of the dipole-moment operator. However, since the operator doesn't commute with the Hamiltonian then there will be a significant error using the mixed distribution in DMC - you need to use the pure distribution using future walking<sup>84,85</sup> or whatever. This is a significant extra complication, and by formulating the thing as a derivative you avoid having to do that. As well as the electric field, the perturbation could be the displacement of nuclear positions (giving forces, etc.) or a combination of both (for example, the intensity of peaks in IR spectra depends on changes in the dipole moment corresponding to changes in geometry). Such energy derivatives can of course be computed numerically (by finite differencing) or analytically (by differentiating the appropriate energy expressions), the latter being clearly preferable in this case.

First of all we shall focus on atomic forces. These are generally used in three main areas of computational electronic structure theory: structural optimization, the computation of vibrational properties, and in explicit 'molecular dynamics' simulations of atomic behaviour<sup>139</sup>. Unfortunately, methods for calculating accurate forces in QMC in a reasonable amount of computer time have proved elusive - at least until relatively recently - due to the lack of readily calculable expressions with reasonable statistical properties.

As is usual, we begin with a discussion of the Hellmann-Feynman theorem (HFT), which in this context is the statement that the force is the expectation value of the gradient of the Hamiltonian  $\hat{H}$ :

$$F = -\nabla E = - \frac{\int \Psi \nabla \hat{H} \Psi \, d\mathbf{R}}{\int \Psi \Psi \, d\mathbf{R}}, \quad (39)$$

The other terms in the expression for the gradient of the expectation value of the energy (the ones involving derivatives of the wave function itself) have disappeared only because we are assuming that the wave function is an exact eigenstate. Inevitably then, the use of the HFT is an approximation in QMC because we only have an inexact

trial function. The correct QMC expressions for the forces must contain additional (‘Pulay’) terms which depend on wave function derivatives. There is also an additional term which accounts for the action of the gradient operator on parameters which only indirectly couple with the nuclear positions (e.g. orbital coefficients) but this can be greatly reduced by optimizing the wave function through minimization of the *energy*, rather than the variance. There is another kind of Pulay term which arises in DMC. The HFT is expected to be valid for the exact DMC algorithm since it solves for the ground state of the fixed-node Hamiltonian exactly. However this Hamiltonian differs from the physical one due to the presence of the infinite potential barrier on the trial nodal surface which constrains the DMC wave function  $\phi_0$  to go to zero there. As we vary the nuclear position(s) the nodal surface moves, and hence the infinite potential barrier moves, giving a contribution to  $\nabla\hat{H}$  that depends on both  $\Psi_T$  and its first derivative<sup>140–142</sup>.

In order to calculate the Pulay terms arising from the derivative of the mixed estimator of Eq. (31) we need in principle to calculate a derivative of the DMC wave function  $\phi_0$ . Because we don’t have any kind of formula for  $\phi_0$ , this derivative cannot be readily evaluated, and what has been done in the past is to use the expression for the derivative of the trial function  $\Psi_T$  in its place<sup>142–150</sup>. The resulting errors are of first order in  $(\Psi_T - \phi_0)$  and  $(\Psi'_T - \phi'_0)$ ; therefore its accuracy depends sensitively on the quality of the trial function and its derivative. In practice the results obtained from this procedure are not generally accurate enough.

Instead of using the usual mixed DMC energy expression one may calculate forces from the ‘pure DMC’ energy given by  $E_D = \int \phi_0 \hat{H} \phi_0 d\mathbf{R} / \int \phi_0 \phi_0 d\mathbf{R}$  which, by construction, is equal to the mixed DMC energy. It’s more expensive to do things this way, but the benefits are now clear. Despite the fact that the derivative  $E'_D$  contains the derivative of the DMC wave function,  $\phi'_0$ , Badinski *et al.*<sup>142</sup> were able to show that  $\phi'_0$  can be eliminated from the pure DMC formula to give the following exact expression (where  $d\mathbf{S}$  is a nodal surface element):

$$E'_D = \frac{\int \phi_0 \phi_0 \phi_0^{-1} \hat{H}' \phi_0 d\mathbf{R}}{\int \phi_0 \phi_0 d\mathbf{R}} - \frac{1}{2} \frac{\int \phi_0 \phi_0 \Psi_T^{-2} |\nabla_{\mathbf{R}} \Psi_T|^2 d\mathbf{S}}{\int \phi_0 \phi_0 d\mathbf{R}}, \quad (40)$$

Of course it’s not easy to compute integrals over the nodal surface, and luckily the expression can be converted into a regular volume integral with no  $\phi'_0$ . The error in the required approximation is then of order  $(\Psi_T - \phi_0)^2$ , giving

$$E'_D = \frac{\int \phi_0 \phi_0 \left[ \phi_0^{-1} \hat{H}' \phi_0 + \Psi_T^{-1} (\hat{H} - E_D) \Psi'_T \right] d\mathbf{R}}{\int \phi_0 \phi_0 d\mathbf{R}} + \frac{\int \Psi_T \Psi_T (E_L - E_D) \Psi_T^{-1} \Psi'_T d\mathbf{R}}{\int \Psi_T \Psi_T d\mathbf{R}} + \mathcal{O}[(\Psi_T - \phi_0)^2]. \quad (41)$$

One may readily evaluate this expression by generating configurations distributed according to the pure ( $\phi_0^2$ ) and variational ( $\Psi_T^2$ ) distributions. The approximation is in the Pulay terms, which are smaller in pure than in mixed DMC and, in addition, the approximation in equation (41) is second-order in contrast to the first-order error obtained by simply substituting  $\Psi'_T$  for  $\phi'_0$ . This equation satisfies the zero variance condition; if  $\Psi_T$  and  $\Psi'_T$  are exact the variance of the force obtained from this formula is zero (the variance of the Hellman-Feynman estimator is, strictly speaking, infinite!). While it remains true that not many papers have been published with actual applications of these methods (some calculations of very accurate forces in small molecules can be found, for example, in Refs. 150 and 151) one can certainly say that reasonable approximations for the difficult expressions have been found, and the outlook for QMC forces is very promising.

## VI. APPLICATIONS

Time and space preclude me from presenting a long list of applications. Here are two. First, a somewhat unfair comparison of the worst DFT functional with VMC and DMC for some cohesive energies of tetrahedrally-bonded semiconductors. Second, the equations of state of diamond and iron. Many other applications can be found, for example, in Ref. 5.

### A. Cohesive energies

A number of VMC and DMC studies have been performed of the cohesive energies of solids. This quantity is given by the difference between the summed energies of the appropriate isolated atoms and the energies of the same atoms



in the bulk crystal. This is generally reckoned to be a severe test of QMC methods because the trial wave functions used in the two cases must be closely matched in quality in order to maximize the effective cancellation of errors. Data for Si, Ge, C and BN have been collected in Table 1. The Local Spin Density Approximation (LSDA) density functional theory data shows the standard overestimation of the cohesive energy while the QMC data is in good agreement with experiment. Studies such as these have been important in establishing DMC as an accurate method for calculating the energies of crystalline solids.

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

TABLE I: 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. References : *a.* Farid and Needs<sup>152</sup>, and references therein. *b.* Rajagopal *et al.*<sup>128</sup>, *d.* Fahy, Wang, and Louie<sup>115</sup>. 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<sup>153</sup>, *f.* Hood *et al.*<sup>27</sup>, *g.* Leung *et al.*<sup>26</sup>, *h.* Estimated by Knittle *et al.*<sup>154</sup> from experimental results on hexagonal BN.

### B. Equations of state of diamond and iron

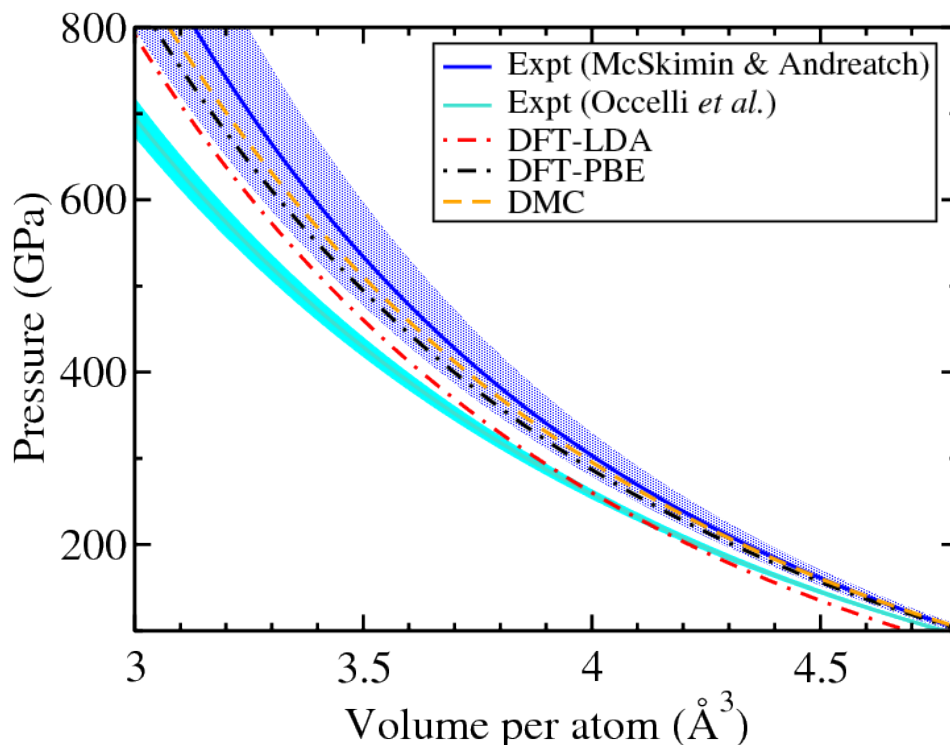


FIG. 6: Equation of state of diamond at high pressures from measurements by McSkimin and Andreatch<sup>155</sup> and Ocelli *et al.*<sup>156</sup>, and as calculated using DFT with two different functionals and DMC<sup>12</sup>. The shaded areas indicate the uncertainty in the experimental equations of state. The zero-point phonon pressure calculated using DFT with the PBE functional is included in the theoretical curves.

The equation of state is the equilibrium relationship between the pressure, volume, and temperature. Computed equations of state are of particular interest in regions where experimental data are difficult to obtain. Diamond anvil

cells are widely used in high-pressure research, and one of the important problems is the measurement of the pressure inside the cell. The most common approach is to place a small grain of ruby in the sample chamber and measure the frequency of a strong laser-stimulated fluorescence line. The resolution is, however, poor at pressures above  $\sim 100$  GPa and alternative methods are being investigated. One possibility is to measure the Raman frequency of diamond itself, assuming that the highest frequency derives from the diamond faces adjacent to the sample chamber. Calibrating such a scale requires an accurate equation of state and the corresponding pressure-dependence of the Raman frequency.

Maezono *et al.* performed VMC, DMC and DFT calculations of the equation of state of diamond<sup>12</sup>. The DMC and DFT data are shown in Fig. 6, along with equations of state derived from experimental data<sup>155,156</sup>. The experimentally-derived equations of state differ significantly at high pressures. It is now believed that the pressure calibration in the more modern experiment of Occelli *et al.*<sup>156</sup> is inaccurate and our DMC data support this view. As can be seen in Fig. 6, the equations of state calculated within DFT depend on the choice of exchange-correlation functional, undermining confidence in the DFT method. A recent QMC study of the equation of state and Raman frequency of cubic boron nitride has produced data which could be used to calibrate pressure measurements in diamond anvil cells<sup>157</sup>.

Another example of a DMC equation of state was produced by Sola, Brodholt and Alfè<sup>158</sup>, who calculated the equation of state of hexagonal close-packed (hcp) iron under Earth's core conditions. With up to 150 atoms or 2400 electrons per cell these represent some of the largest systems studied with DMC to date, and demonstrate the ability of QMC to treat heavier transition metal atoms. Fig. 7 shows the calculated equation of state which agrees closely with experiments and with previous DFT calculations. (DFT is expected to work well in this system and the DMC calculations appear to confirm this). Notice the discontinuity due to the hcp-bcc phase transition in the experimental values reported by Dewaele *et al.* in Ref. 159. At low pressures, the calculations and experiments differ because of the magnetism, which is not taken into account in these particular calculations (though it could be in principle).

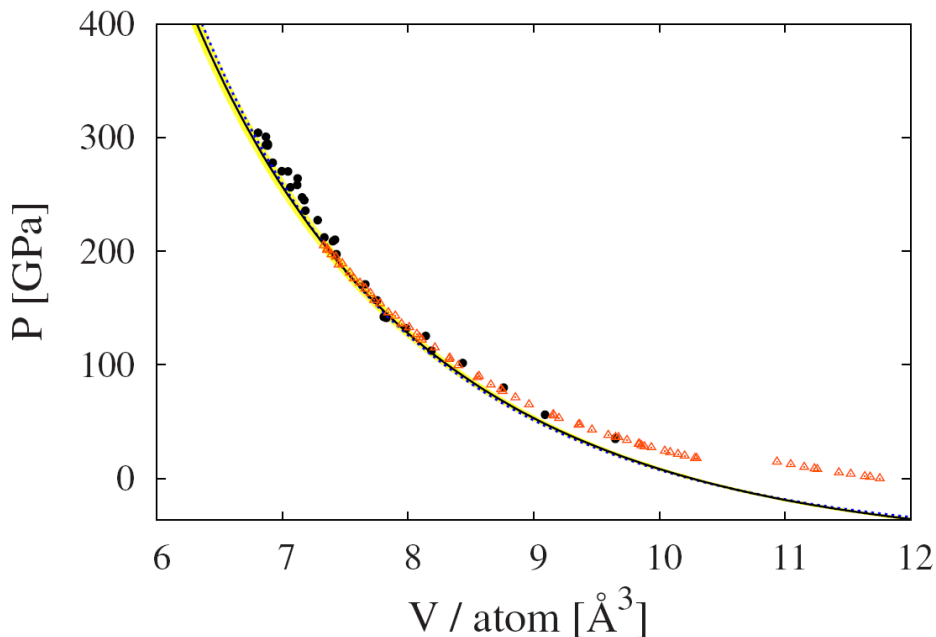


FIG. 7: Pressure versus volume curve in iron obtained from DMC calculations (solid line)<sup>158</sup>. The yellow error band in the DMC curve is due to the errors in the parameters of a fit to the Birch-Murnaghan equation of state. DFT-PW91 results (dotted line, Ref. 160) and experimental data (circles, Ref. 161, and open triangles, Ref. 159) are also reported for comparison.

## VII. CONCLUSIONS

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. Our group has developed the software package CASINO<sup>46–48</sup> which 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 crystals. Young people also seem to appreciate that QMC is way cooler than boring old density functional theory. So that's all right then.

## VIII. ACKNOWLEDGEMENT

MDT would like to thank the Royal Society for the award of a long-term university research fellowship. He also wishes to acknowledge the many contributions of R.J. Needs, N.D. Drummond and P. López Ríos to the work described in this article, along with all the other members of the Cavendish Laboratory's TCM Group plus our many collaborators around the world. Computing facilities were largely provided by the Cambridge High Performance Computing Service.

- 
- <sup>1</sup> C.J. Cramer, *Essentials of Computational Chemistry* (Wiley, 2002), pp. 191-232.  
<sup>2</sup> R.G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules* (Oxford University Press, 1994).  
<sup>3</sup> Gaussian 09, M.J. Frisch *et al.*, Gaussian, Inc., Wallingford CT (2009).  
<sup>4</sup> B.L. Hammond, W.A. Lester, Jr., and P.J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry* (World Scientific, Singapore, 1994).  
<sup>5</sup> W.M.C. Foulkes, L. Mitas, R.J. Needs and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).  
<sup>6</sup> D.M. Ceperley and B.J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).  
<sup>7</sup> S.H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.* **58**, 1200 (1980).  
<sup>8</sup> J.P. Perdew and A. Zunger, *Phys. Rev. B* **23**, 5048 (1981).  
<sup>9</sup> Y.S.M. Wu, A. Kuppermann and J.B. Anderson, *Phys. Chem. Chem. Phys.* **1**, 929 (1999).  
<sup>10</sup> V. Natoli, R.M. Martin and D.M. Ceperley, *Phys. Rev. Lett.* **70**, 1952 (1993).  
<sup>11</sup> K.T. Delaney, C. Pierleoni and D.M. Ceperley, *Phys. Rev. Lett.* **97**, 235702 (2006).  
<sup>12</sup> R. Maezono, A. Ma, M.D. Towler and R.J. Needs, *Phys. Rev. Lett.* **98**, 025701 (2007).  
<sup>13</sup> M. Pozzo and D. Alfè, *Phys. Rev. B* **77**, 104103 (2008).  
<sup>14</sup> D. Alfè, M. Alfredsson, J. Brodholt, M.J. Gillan, M.D. Towler and R.J. Needs, *Phys. Rev. B* **72**, 014114 (2005).  
<sup>15</sup> S. Manten and A. Lüchow, *J. Chem. Phys.* **115**, 5362 (2001).  
<sup>16</sup> J.C. Grossman, *J. Chem. Phys.* **117**, 1434 (2002).  
<sup>17</sup> A. Aspuru-Guzik, O. El Akramine, J.C. Grossman and W.A. Lester Jr., *J. Chem. Phys.* **120**, 3049 (2004).  
<sup>18</sup> I.G. Gurtubay, N.D. Drummond, M.D. Towler and R.J. Needs, *J. Chem. Phys.* **124**, 024318 (2006).  
<sup>19</sup> I.G. Gurtubay and R.J. Needs, *J. Chem. Phys.* **127**, 124306 (2007).  
<sup>20</sup> R.Q. Hood, M.-Y. Chou, A.J. Williamson, G. Rajagopal, R.J. Needs and W.M.C. Foulkes, *Phys. Rev. Lett.* **78**, 3350 (1997).  
<sup>21</sup> R.Q. Hood, M.-Y. Chou, A.J. Williamson, G. Rajagopal and R.J. Needs, *Phys. Rev. B* **57**, 8972 (1998).  
<sup>22</sup> M. Nekovee, W.M.C. Foulkes and R.J. Needs, *Phys. Rev. Lett.* **87**, 036401 (2001).  
<sup>23</sup> M. Nekovee, W.M.C. Foulkes and R.J. Needs, *Phys. Rev. B* **68**, 235108 (2003).  
<sup>24</sup> A.J. Williamson, J.C. Grossman, R.Q. Hood, A. Puzder and G. Galli, *Phys. Rev. Lett.* **89**, 196803 (2002).  
<sup>25</sup> N.D. Drummond, A.J. Williamson, R.J. Needs and G. Galli, *Phys. Rev. Lett.* **95**, 096801 (2005).  
<sup>26</sup> W.-K. Leung, R.J. Needs, G. Rajagopal, S. Itoh and S. Ihara, *Phys. Rev. Lett.* **83**, 2351 (1999).  
<sup>27</sup> R.Q. Hood, P.R.C. Kent, R.J. Needs and P.R. Briddon, *Phys. Rev. Lett.* **91**, 076403 (2003).  
<sup>28</sup> D. Alfè and M.J. Gillan, *Phys. Rev. B* **71**, 220101 (2005).  
<sup>29</sup> M.D. Towler and R.J. Needs, *Int. J. Mod. Phys. B* **17**, 5425 (2003).  
<sup>30</sup> L.K. Wagner and L. Mitas, *Chem. Phys. Lett* **370**, 412 (2003).  
<sup>31</sup> L.K. Wagner and L. Mitas, *J. Chem. Phys.* **126**, 034105 (2007).  
<sup>32</sup> L. Mitas and R.M. Martin, *Phys. Rev. Lett* **72**, 2438 (1994).  
<sup>33</sup> A.J. Williamson, R.Q. Hood, R.J. Needs and G. Rajagopal, *Phys. Rev. B* **57**, 12140 (1998).  
<sup>34</sup> M.D. Towler, R.Q. Hood and R.J. Needs, *Phys. Rev. B* **62**, 2330 (2000).  
<sup>35</sup> A. Ghosal, A.D. Guclu, C.J. Umrigar, D. Ullmo and H. Baranger, *Nature Physics* **2**, 336 (2006).  
<sup>36</sup> S.B. Healy, C. Filippi, P. Kratzer, E. Penev and M. Scheffler, *Phys. Rev. Lett* **87**, 016105 (2001).  
<sup>37</sup> C. Filippi, S.B. Healy, P. Kratzer, E. Pehlke and M. Scheffler, *Phys. Rev. Lett* **89**, 166102 (2002).  
<sup>38</sup> Y.-H. Kim, Y. Zhao, A. Williamson, M.J. Heben and S. Zhang, *Phys. Rev. Lett.* **96**, 016102 (2006).  
<sup>39</sup> J. Carlson, S.-Y. Chang, V.R. Pandharipande and K.E. Schmidt, *Phys. Rev. Lett* **91**, 050401 (2003).  
<sup>40</sup> G.E. Astrakharchik, J. Boronat, J. Casulleras and S. Giorgini, *Phys. Rev. Lett.* **93**, 200404 (2004).  
<sup>41</sup> J. Carlson and S. Reddy, *Phys. Rev. Lett* **100**, 150403 (2008).  
<sup>42</sup> E. Schrödinger, *Ann. Phys.* **79**, 361 (1926).  
<sup>43</sup> N.W. Ashcroft and N.D. Mermin, *Solid State Physics* (Saunders, 1976), p.330.  
<sup>44</sup> P.R.C. Kent, M.D. Towler, R.J. Needs and G. Rajagopal, *Phys. Rev. B* **62**, 15394 (2000).  
<sup>45</sup> [www.qmcwiki.org/index.php/Research\\_resources](http://www.qmcwiki.org/index.php/Research_resources)

- <sup>46</sup> R.J. Needs, M.D. Towler, N.D. Drummond and P. López Ríos, CASINO version 2.5 User Manual, University of Cambridge, Cambridge (2009).
- <sup>47</sup> CASINO web site : [www.tcm.phy.cam.ac.uk/~mdt26/casino2.html](http://www.tcm.phy.cam.ac.uk/~mdt26/casino2.html)
- <sup>48</sup> [www.vallico.net/tti/tti.html](http://www.vallico.net/tti/tti.html) - click 'PUBLIC EVENTS'.
- <sup>49</sup> J.R. Trail, *Phys. Rev. E* **77**, 016703 (2008).
- <sup>50</sup> J.R. Trail, *Phys. Rev. E* **77**, 016704 (2008).
- <sup>51</sup> N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.M. Teller and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- <sup>52</sup> M.D. Towler, *Pilot waves, Bohmian metaphysics and the foundations of quantum mechanics*, Graduate lecture course available at [www.tcm.phy.cam.ac.uk/~mdt26/pilot\\_waves.html](http://www.tcm.phy.cam.ac.uk/~mdt26/pilot_waves.html) (2009).
- <sup>53</sup> R.J. Jastrow, *Phys. Rev.* **98**, 1479 (1955).
- <sup>54</sup> N.D. Drummond, M.D. Towler and R.J. Needs, *Phys. Rev. B* **70**, 235119 (2004).
- <sup>55</sup> S. Aragon, *Density functional theory: a primer*, San Francisco State University teaching material, available online.
- <sup>56</sup> T. Kato, *Commun. Pure Appl. Math.* **10**, 151 (1957).
- <sup>57</sup> S. de Palo, F. Rapisarda and G. Senatore, *Phys. Rev. Lett.* **88**, 206401 (2002).
- <sup>58</sup> P. López Ríos and R.J. Needs, unpublished.
- <sup>59</sup> J.E. Dennis, D.M. Gay and R.E. Welsch, *Algorithm 573: NL2SOL—An Adaptive Nonlinear Least-Squares Algorithm*, ACM Transactions on Mathematical Software **7**, 369 (1981).
- <sup>60</sup> C.J. Umrigar, K.G. Wilson and J.W. Wilkins, *Phys. Rev. Lett* **60**, 1719 (1988).
- <sup>61</sup> P.R.C. Kent, R.J. Needs and G. Rajagopal, *Phys. Rev. B* **59**, 12344 (1999).
- <sup>62</sup> N.D. Drummond and R.J. Needs, *Phys. Rev. B* **72**, 085124 (2005).
- <sup>63</sup> D.M. Ceperley, *J. Stat. Phys.* **43**, 815 (1986).
- <sup>64</sup> A. Ma, N.D. Drummond, M.D. Towler and R.J. Needs, *Phys. Rev. E* **71**, 066704 (2005).
- <sup>65</sup> K.E. Riley and J.B. Anderson, *Mol. Phys.* **101**, 3129 (2003).
- <sup>66</sup> M.P. Nightingale and V. Melik-Alaverdian, *Phys. Rev. Lett.* **87**, 043401 (2001).
- <sup>67</sup> C.J. Umrigar, J. Toulouse, C. Filippi, S. Sorella and R.G. Hennig, *Phys. Rev. Lett.* **98**, 110201 (2007).
- <sup>68</sup> J. Toulouse and C.J. Umrigar, *J. Chem. Phys.* **126**, 084102 (2007).
- <sup>69</sup> D.M. Ceperley, *Top-ten reasons why no-one uses quantum Monte Carlo*, Ceperley group website (1996). Since removed.
- <sup>70</sup> J.A. Pople, M. Head-Gordon, D.J. Fox, K. Raghavachari and L.A. Curtiss, *J. Chem. Phys.* **90**, 5622 (1989).
- <sup>71</sup> L.A. Curtiss, C. Jones, G.W. Trucks, K. Raghavachari and J.A. Pople, *J. Chem. Phys.* **93**, 2537 (1990).
- <sup>72</sup> L.A. Curtiss, K. Raghavachari, P.C. Redfern and J.A. Pople, *J. Chem. Phys.* **106**, 1063 (1997).
- <sup>73</sup> A. Ma, N.D. Drummond, M.D. Towler and R.J. Needs, *J. Chem. Phys.* **122**, 224322 (2005).
- <sup>74</sup> N. Nemeč, M.D. Towler and R.J. Needs, *J. Chem. Phys.* **22**, 074202 (2010)
- <sup>75</sup> M.H. Kalos, L. Colletti and F. Pederiva, *J. Low Temp. Phys.* **138**, 747 (2005).
- <sup>76</sup> J.B. Anderson, *J. Chem. Phys.* **63**, 1499 (1975); *Ibid.* **65**, 4121 (1976).
- <sup>77</sup> D.M. Ceperley, *J. Stat. Phys.* **63**, 1237 (1991).
- <sup>78</sup> W.M.C. Foulkes, R.Q. Hood and R.J. Needs, *Phys. Rev. B* **60**, 4558 (1999).
- <sup>79</sup> W. Glauser, W. Brown, W. Lester, D. Bressanini and B. Hammond, *J. Chem. Phys.* **97**, 9200 (1992).
- <sup>80</sup> D. Bressanini and P.J. Reynolds, *Phys. Rev. Lett.* **95**, 110201 (2005).
- <sup>81</sup> M. Bajdich, L. Mitás, G. Drobný and L.K. Wagner, *Phys. Rev. B* **60**, 4558 (1999).
- <sup>82</sup> M.D. Towler, N.L. Allan, N.M. Harrison, V.R. Saunders, W.C. Mackrodt and E. Aprà, *Phys. Rev. B* **50**, 5041 (1994).
- <sup>83</sup> R.J. Needs and M.D. Towler, *Int. J. Mod. Phys. B* **17**, 5425 (2003).
- <sup>84</sup> S.K. Liu, M.H. Kalos and G.V. Chester, *Phys. Rev. A* **10**, 303 (1974).
- <sup>85</sup> R.N. Barnett, P.J. Reynolds and W.A. Lester Jr., *J. Comput. Phys.* **96**, 258 (1991).
- <sup>86</sup> S. Baroni and S. Moroni, *Phys. Rev. Lett* **82**, 4745 (1999).
- <sup>87</sup> N.D. Drummond, Z. Radnai, J.R. Trail, M.D. Towler and R.J. Needs, *Phys. Rev. B* **69**, 085116 (2004).
- <sup>88</sup> N.D. Drummond and R.J. Needs, *Phys. Rev. Lett* **102**, 126402 (2009).
- <sup>89</sup> A. Lüchow, R. Petz and T.C. Scott, *J. Chem. Phys.* **126**, 144110 (2007).
- <sup>90</sup> F.A. Reboredo, R.Q. Hood and P.R.C. Kent, *Phys. Rev. B* **79**, 195117 (2009).
- <sup>91</sup> N.D. Drummond, P. López Ríos, A. Ma, J.R. Trail, G. Spink, M.D. Towler and R.J. Needs, *J. Chem. Phys.* **124**, 224104 (2006).
- <sup>92</sup> S. Fahy, in *Quantum Monte Carlo Methods in Physics and Chemistry*, Nato Science Series C: Mathematical and Physical Sciences, Vol. 525, ed. M.P. Nightingale and C.J. Umrigar (Kluwer Academic, Dordrecht, 1999), p. 101.
- <sup>93</sup> C. Filippi and S. Fahy, *J. Chem. Phys.* **112**, 3523 (2000).
- <sup>94</sup> J.C. Grossman and L. Mitás, *Phys. Rev. Lett.* **74**, 1323 (1995).
- <sup>95</sup> W. Kutzelnigg and J.D. Morgan III, *J. Phys. Chem.* **96**, 4484 (1992).
- <sup>96</sup> D. Prendergast, M. Nolan, C. Filippi, S. Fahy and J.C. Greer, *J. Chem. Phys.* **115**, 1626 (2001).
- <sup>97</sup> R.P. Feynman, *Phys. Rev.* **94**, 262 (1954).
- <sup>98</sup> R.P. Feynman and M. Cohen, *Phys. Rev.* **102**, 1189 (1956).
- <sup>99</sup> Y. Kwon, D.M. Ceperley and R.M. Martin, *Phys. Rev. B* **48**, 12037 (1993).
- <sup>100</sup> M. Holzmann, D.M. Ceperley, C. Pierleoni and K. Esler, *Phys. Rev. E* **68**, 046707 (2003).
- <sup>101</sup> Y. Kwon, D.M. Ceperley and R.M. Martin, *Phys. Rev. B* **58**, 6800 (1998).
- <sup>102</sup> C. Pierleoni, D.M. Ceperley and M. Holzmann, *Phys. Rev. Lett* **93**, 146402 (2004).
- <sup>103</sup> P. López Ríos, A. Ma, N.D. Drummond, M.D. Towler and R.J. Needs, *Phys. Rev. E* **74**, 066701 (2006).
- <sup>104</sup> 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.: Cond. Mat.* **14**, 2717 (2002).

- <sup>105</sup> *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, *Computational Materials Science* **25**, 478 (2002).
- <sup>106</sup> S. Baroni, A. Dal Corso, S. de Gironcoli and P. Giannozzi, [www.pwscf.org](http://www.pwscf.org).
- <sup>107</sup> E. Hernandez, M.J. Gillan and C.M. Goringe, *Phys. Rev. B* **55**, 13485 (1997).
- <sup>108</sup> D. Alfè and M.J. Gillan, *Phys. Rev. B* **70**, 161101 (2004).
- <sup>109</sup> R. Dovesi, V.R. Saunders, C. Roetti, R. Orlando, C.M. Zicovich-Wilson, F. Pascale, B. Civalleri, K. Doll, N.M. Harrison, I.J. Bush, Ph. D'Arco and M. Llunell, CRYSTAL06 User's Manual (University of Torino, Torino, 2006).
- <sup>110</sup> G. te Velde, F.M. Bickelhaupt, E.J. Baerends, C.F. Guerra, S.J.A. van Gisbergen, J.G. Snijders, and T. Ziegler, *J. Comput. Chem.* **22**, 931 (2001).
- <sup>111</sup> A practice which has recently been outlawed in our department by new University anti-smoking legislation. My thanks to an anonymous referee for supplying me with this joke.
- <sup>112</sup> C.J. Umrigar, *Phys. Rev. Lett.* **71**, 408 (1993).
- <sup>113</sup> M.L. Stedman, W.M.C. Foulkes and M. Nekovee, *J. Chem. Phys.* **109**, 2630 (1998).
- <sup>114</sup> P.H. Acioli and D.M. Ceperley, *J. Chem. Phys.* **100**, 8169 (1994).
- <sup>115</sup> S. Fahy, X.W. Wang and S.G. Louie, *Phys. Rev. B* **42**, 3503 (1990).
- <sup>116</sup> S. Fahy, X.W. Wang and S.G. Louie, *Phys. Rev. Lett.* **61**, 1631 (1998).
- <sup>117</sup> L. Mitas, E.L. Shirley and D.M. Ceperley, *J. Chem. Phys.* **95**, 3467 (1991).
- <sup>118</sup> M. Casula, C. Filippi and S. Sorella, *Phys. Rev. Lett* **95**, 100201 (2005).
- <sup>119</sup> M. Casula, *Phys. Rev. B* **74**, 161102 (2006).
- <sup>120</sup> C.W. Greeff and W.A. Lester Jr., *J. Chem. Phys.* **109**, 1607 (1998).
- <sup>121</sup> E.L. Shirley and R.M. Martin, *Phys. Rev. B* **47**, 15413 (1993).
- <sup>122</sup> J.R. Trail and R.J. Needs, *J. Chem. Phys.* **122**, 174109 (2005).
- <sup>123</sup> J.R. Trail and R.J. Needs, *J. Chem. Phys.* **122**, 014112 (2005).
- <sup>124</sup> [www.tcm.phy.cam.ac.uk/~mdt26/casino2\\_pseudopotentials.html](http://www.tcm.phy.cam.ac.uk/~mdt26/casino2_pseudopotentials.html)
- <sup>125</sup> M. Burkatzki, C. Filippi and M. Dolg, *J. Chem. Phys.* **126**, 234105 (2007); *Ibid.* **129**, 164115 (2008).
- <sup>126</sup> J.R. Trail and R.J. Needs, *J. Chem. Phys.* **128**, 204103 (2008).
- <sup>127</sup> B. Santra, A. Michaelides, M. Fuchs, A. Tkatchenko, C. Filippi and M. Scheffler, *J. Chem. Phys.* **129**, 194111 (2008).
- <sup>128</sup> G. Rajagopal, R.J. Needs, A.J. James, S.D. Kenny and W.M.C. Foulkes, *Phys. Rev. B* **51**, 10591 (1995).
- <sup>129</sup> G. Rajagopal, R.J. Needs, S.D. Kenny, W.M.C. Foulkes and A.J. James, *Phys. Rev. Lett.* **73**, 1959 (1994).
- <sup>130</sup> C. Lin, F.H. Zong and D.M. Ceperley, *Phys. Rev. E* **64**, 016702 (2001).
- <sup>131</sup> P.P. Ewald, *Ann. Phys.* **64**, 25 (1921).
- <sup>132</sup> S. Chiesa, D.M. Ceperley, R.M. Martin and M. Holzmann, *Phys. Rev. Lett.* **97**, 076404 (2006).
- <sup>133</sup> N.D. Drummond, R.J. Needs, A. Sorouri and W.M.C. Foulkes, *Phys. Rev. B* **78**, 125106 (2008).
- <sup>134</sup> 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).
- <sup>135</sup> 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).
- <sup>136</sup> P.R.C. Kent, R.Q. Hood, A.J. Williamson, R.J. Needs, W.M.C. Foulkes and G. Rajagopal, *Phys. Rev. B* **59**, 1917 (1999).
- <sup>137</sup> H. Kwee, S. Zhang and H. Krakauer, *Phys. Rev. Lett.* **100**, 126404 (2008).
- <sup>138</sup> M. Dewing and D.M. Ceperley, *Methods for Coupled Electronic-Ionic Monte Carlo in Recent Advances in Quantum Monte Carlo Methods, Part II*, ed. by W.A. Lester, S.M. Rothstein and S. Tanaka (World Scientific, Singapore, 2002).
- <sup>139</sup> J.C. Grossman and L. Mitas, *Phys. Rev. Lett* **94**, 056403 (2005).
- <sup>140</sup> K.C. Huang, R.J. Needs and G. Rajagopal, *J. Chem. Phys.* **112**, 4419 (2000).
- <sup>141</sup> F. Schautz and H.-J. Flad, *J. Chem. Phys.* **112**, 4421 (2000).
- <sup>142</sup> A. Badinski, P.D. Haynes and R.J. Needs, *Phys. Rev. B* **77**, 085111 (2008).
- <sup>143</sup> P.J. Reynolds, R.N. Barnett, B.L. Hammond, R.M. Grimes and W.A. Lester Jr., *Int. J. Quant. Chem.* **29**, 589 (1986).
- <sup>144</sup> R. Assaraf and M. Caffarel, *Phys. Rev. Lett.* **83**, 4682 (1999).
- <sup>145</sup> M. Casalegno, M. Mella and A.M. Rappe, *J. Chem. Phys.* **118**, 7193 (2003).
- <sup>146</sup> R. Assaraf and M. Caffarel, *J. Chem. Phys.* **119**, 10536 (2003).
- <sup>147</sup> M.W. Lee, M. Mella and A.M. Rappe, *J. Chem. Phys.* **122**, 244103 (2005).
- <sup>148</sup> A. Badinski and R.J. Needs, *Phys. Rev. E* **76**, 036707 (2007).
- <sup>149</sup> A. Badinski and R.J. Needs, *Phys. Rev. B* **78**, 035134 (2008).
- <sup>150</sup> A. Badinski, J.R. Trail and R.J. Needs, *J. Chem. Phys.* **129**, 224101 (2008).
- <sup>151</sup> A. Badinski, P.D. Haynes, J.R. Trail and R.J. Needs, *J. Phys.: Condensed Matter* **22**, 074202 (2010).
- <sup>152</sup> B. Farid and R. J. Needs, *Phys. Rev. B* **45**, 1067 (1992).
- <sup>153</sup> A. Malatesta, S. Fahy and G.B. Bachelet, *Phys. Rev. B* **56**, 12201 (1997).
- <sup>154</sup> E. Knittle, R. Wentzcovitch, R. Jeanloz and M.L. Cohen, *Nature* **337**, 349 (1989).
- <sup>155</sup> H.J. McSkimin and P. Andreatch, *J. Appl. Phys.* **43**, 2944 (1972).
- <sup>156</sup> F. Occelli, P. Loubeyre and R. LeToullec, *Nat. Mater.* **2**, 151 (2003).
- <sup>157</sup> K.P. Esler, R.E. Cohen, B. Militzer, J. Kim, R.J. Needs and M.D. Towler, unpublished (2009).
- <sup>158</sup> E. Sola, J.P. Brodholt and D. Alfè, *Phys. Rev. B* **79**, 024107 (2009).
- <sup>159</sup> A. Dewaele, P. Loubeyre, F. Occelli, M. Mezouar, P.I. Dorogokupets and M. Torrent, *Phys. Rev. Lett.* **97**, 215504 (2006).
- <sup>160</sup> P. Söderlind, J.A. Moriarty and J.M. Wills, *Phys. Rev. B* **53**, 14063 (1996).
- <sup>161</sup> K. Mao, Y. Wu, L.C. Chen and J.F. Shu, *J. Geophys. Res.* **95**, 21737 (1990).