

Quantum Monte Carlo simulations of solids

W. M. C. Foulkes

CMTH Group, Department of Physics, Imperial College of Science, Technology and Medicine, Prince Consort Road, London SW7 2BZ, England

L. Mitas

National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL 61801, USA

R. J. Needs and G. Rajagopal

TCM Group, Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK

This article describes the variational and fixed-node diffusion quantum Monte Carlo methods and how they may be used to calculate the properties of many-electron systems. These stochastic wave-function-based approaches provide a very direct treatment of quantum many-body effects and serve as benchmarks against which other techniques may be compared. They complement the less demanding density-functional approach by providing more accurate results and a deeper understanding of the physics of electronic correlation in real materials. The algorithms are intrinsically parallel and currently available high performance computers allow applications to systems containing a thousand or more electrons. With these tools one can study complicated problems such as the properties of surfaces and defects, while including electron correlation effects with high precision. We provide a pedagogical overview of the techniques and describe a selection of applications to ground and excited states of solids and clusters.

CONTENTS

| | | | |
|--|----|---|----|
| I. INTRODUCTION | 2 | B. Phases of the electron gas | 26 |
| II. INTERACTING ELECTRONS IN SOLIDS | 3 | C. Static response of the electron gas | 27 |
| A. The many-electron Schrödinger equation | 3 | D. The relativistic electron gas | 27 |
| B. Hartree-Fock theory | 4 | E. Exchange and correlation energies | 28 |
| C. Post Hartree-Fock methods | 5 | F. Compton scattering in Si and Li | 29 |
| D. Density-functional theory | 5 | G. Solid Hydrogen | 30 |
| E. Quantum Monte Carlo methods | 6 | H. Clusters | 31 |
| III. MONTE CARLO METHODS | 7 | I. Formation energies of silicon self-interstitials | 32 |
| A. Statistical foundations | 7 | J. Jellium surfaces | 33 |
| B. The Metropolis algorithm | 8 | VI. EXCITED STATES | 34 |
| C. Variational Monte Carlo | 9 | A. Introduction | 34 |
| D. Diffusion Monte Carlo | 10 | B. VMC and DMC calculations of excitations in solids | 34 |
| 1. The imaginary-time Schrödinger equation | 10 | C. Other QMC methods for excited states | 35 |
| 2. The fixed-node approximation | 12 | VII. WAVE FUNCTION OPTIMIZATION | 35 |
| a. One-electron example | 12 | A. Introduction | 35 |
| b. Many-electron version | 13 | B. The cost function | 36 |
| c. The fixed-node variational principle | 14 | C. Numerical stability of variance minimization | 36 |
| d. The tiling theorem | 15 | D. Minimization procedures | 37 |
| 3. Importance sampling | 15 | VIII. PSEUDOPOTENTIALS | 37 |
| IV. TRIAL WAVE FUNCTIONS | 18 | A. The need for pseudopotentials | 37 |
| A. Introduction | 18 | B. Non-local pseudopotentials | 38 |
| B. Slater-Jastrow wave functions | 18 | C. Core polarization potentials | 38 |
| C. The Slater determinant | 20 | D. Pseudopotentials in variational Monte Carlo | 39 |
| D. The Jastrow factor | 21 | E. Pseudopotentials in diffusion Monte Carlo | 39 |
| E. Spin | 23 | F. Alternatives to non-local pseudopotentials | 40 |
| F. The cusp conditions | 24 | IX. PERIODIC BOUNDARY CONDITIONS AND FINITE-SIZE ERRORS | 41 |
| V. SELECTED APPLICATIONS OF QUANTUM MONTE CARLO TO GROUND STATES | 25 | A. Introduction | 41 |
| A. Cohesive energies of solids | 25 | B. Bloch's theorem for many-body systems | 41 |

| | |
|--|----|
| C. Periodic boundary conditions and Coulomb interactions | 42 |
| D. Coping with finite-size errors | 43 |
| 1. Introduction | 43 |
| 2. Finite-size correction and extrapolation formulae | 43 |
| 3. Choosing the simulation-cell wave vector | 44 |
| 4. Interaction energy finite-size effects | 44 |
| X. COMPUTATIONAL ISSUES | 45 |
| A. Representation of the single-particle orbitals | 45 |
| B. Evaluation of the trial wave function | 46 |
| C. Evaluation of the local energy | 46 |
| D. Efficient treatment of the Coulomb interaction | 47 |
| E. Scaling with the number of particles | 47 |
| F. QMC on Parallel Computers | 47 |
| XI. CONCLUSIONS | 48 |
| ACKNOWLEDGMENTS | 50 |
| References | 50 |

I. INTRODUCTION

The past few decades have seen dramatic improvements in our ability to simulate complicated physical systems using computers. This has led to the concept of computer simulation as a “third way” of doing science, closer to experiment than theory but complementary to both. Like experiments, simulations produce data rather than theories and should be judged on the quality of those data. We aim to show that the quantum Monte Carlo (QMC) methods described in this review are accurate and reliable, and that the data and insights they provide are often difficult or impossible to obtain in any other way.

In studying the quantum mechanical properties of solids we are fortunate that the underlying physical laws have been known for over 70 years. The difficulty arises from the complexity of the equations which, as Dirac (1929) famously wrote, are “much too complicated to be soluble”. It hardly seems possible that one could solve the $3N$ -dimensional Schrödinger equation that describes the N interacting electrons in a solid, but this is exactly what QMC methods allow us to do. With current parallel computers we can simulate periodic systems containing up to a thousand or so electrons, which is enough to model many properties of solids with impressive accuracy. QMC is now recognized as a valuable tool for calculating the physical properties of real materials.

The term “quantum Monte Carlo” covers several different techniques based on random sampling. The simplest of these, variational Monte Carlo (VMC), uses a stochastic integration method to evaluate expectation values for a chosen trial wave function. In a system of 1000 electrons the required integrals are 3000 dimensional, and for such problems Monte Carlo integration is much more efficient than conventional quadrature methods such as Simpson’s rule. The main drawback of VMC is that the accuracy of the result depends entirely on the accuracy of

the trial wave function. The other method we consider, diffusion Monte Carlo (DMC), overcomes this limitation by using a projection technique to enhance the ground-state component of a starting trial wave function. During the last ten years it has become clear that VMC and DMC can produce very accurate ground-state expectation values for weakly correlated solids. They can also provide some information about specific excited states, although not yet full excitation spectra.

Although VMC and DMC are the only methods currently used to study realistic continuum models of electrons in solids, other methods are useful in other systems. Auxiliary-field Monte Carlo (for a recent review see Senatore and March, 1994) is often applied to model Hamiltonians such as the Hubbard model, and path-integral Monte Carlo (reviewed by Ceperley, 1995a, 1995b) is the main QMC method used to study bosonic systems such as liquid helium. Both auxiliary-field and path-integral QMC are in principle capable of simulating interacting electrons at finite temperature and may one day be used to study real solids. In this review, however, we focus on the two techniques that have already proved their worth in solids: the zero-temperature VMC and DMC methods.

Because VMC and DMC simulations are very expensive, most calculations of the ground-state properties of solids use less accurate methods based on Hohenberg-Kohn-Sham density-functional theory (DFT) or Hartree-Fock (HF) theory (see, e.g., Parr and Yang, 1989). The efficiency of these techniques arises from the replacement of the electron-electron Coulomb interactions by an effective one-electron potential; this makes the many-electron Schrödinger equation separable and hence much easier to solve. DFT, in particular, has been remarkably successful and is now used in fields ranging from solid state physics and quantum chemistry to geophysics and molecular biology.

Although DFT is exact in principle, it relies in practice on approximations to the unknown exchange-correlation energy functional, which accounts for the complicated correlated motion of the electrons. Most DFT calculations use either the local density approximation or the generalized gradient approximation, both of which work surprisingly well under most circumstances. Obvious failures are uncommon, but include the underestimation of s to d transfer energies in transition metal atoms and difficulties in describing van der Waals bonding, both of which have been dealt with successfully using QMC techniques. Problems also occur in covalent sp -bonded materials, as demonstrated by the QMC studies of carbon clusters described in Sec. V. The unreliability of the DFT results for these clusters provides a good demonstration of the limited predictive power of existing approximate exchange-correlation energy functionals.

The strong electronic correlations in solids containing d and f electrons produce a wide range of subtle effects,

including the anomalous normal-state transport properties of high-temperature superconductors and the very large low temperature specific heat coefficients of heavy fermion materials. Correlation effects such as these are not well described by DFT or HF and are normally studied using simplified Hamiltonians such as the Hubbard model. Given a set of atomic positions, it is often possible to devise a model Hamiltonian that gives an excellent account of the low-lying electronic excitations on an energy scale that may be as low as a few degrees Kelvin. The simplicity is not achieved without cost, however, since most model Hamiltonians drastically approximate the “independent-electron” energy given so accurately by HF and DFT calculations. The independent-electron contribution is normally more than 99% of the total energy, and so model Hamiltonians rarely give accurate ground-state energies as functions of the atomic positions. This review is concerned with studies of interatomic forces and chemical bonding, for which the relevant energy scales range from a few meV to several eV, and for which continuum approaches based on the full many-electron Schrödinger equation are more appropriate.

Continuum QMC simulations require much more computer time than DFT or HF calculations, but can already achieve chemical accuracy (usually defined as 1 kcal per mole, which is about 0.04 eV per molecule) in small systems and may soon achieve it in solids. Chemical accuracy is sufficient to address most issues involving interatomic forces and chemical reactions, but is still an order of magnitude worse than the accuracy required to study phenomena such as superconductivity. The computational cost of a QMC calculation increases as the cube of the number of electrons, which makes applications to large systems feasible, and there is no fundamental reason why the accuracy should fall off for large systems. QMC algorithms are also quite simple to program and very well suited to massively parallel computation.

QMC methods have already proved themselves important in a very broad sense — they can provide much more than just accurate numbers. A good example is provided by Ceperley and Alder’s (1980) simulations of the homogeneous electron gas, which underlie the accurate parameterized local density approximations used in most density-functional calculations. The rapid acceptance of these QMC-based approximations helped spark the twenty years of growth and development that made density-functional methods so popular and powerful.

Our intention is that most of what follows should be accessible to a wide audience, but no doubt some readers will find it helpful to consult other sources. Kalos and Whitlock (1986) give a straightforward introduction to Monte Carlo methods in general, and Hammond, Lester, and Reynolds (1994) cover the use of QMC in quantum chemistry. Other useful recent sources are the review by Anderson (1999) and the books edited by Lester (1997)

and Nightingale and Umrigar (1999).

In Sec. II we give an overview of techniques used to study the quantum mechanics of many-electron systems. Section III discusses the statistical foundations of the Monte Carlo approach and describes the theory behind the VMC and DMC methods. The successes of VMC and DMC rest on the surprising accuracy of relatively simple trial wave functions, which are considered in Sec. IV. In Sec. V we describe a selection of QMC simulations of the ground-state properties of solids and clusters, while in Sec. VI we highlight studies of excited states. In Sec. VII we describe the methods used to optimize trial wave functions, and in Sec. VIII we discuss pseudopotentials and why they are needed in QMC simulations. Section IX looks at the issue of periodic boundary conditions and the proper treatment of finite-size errors. In Sec. X we consider issues of computational efficiency. Section XI concludes this review.

II. INTERACTING ELECTRONS IN SOLIDS

A. The many-electron Schrödinger equation

One of the great challenges of condensed matter physics is to obtain accurate approximate solutions of the many-electron Schrödinger equation. Because the mass of an electron is so small compared to that of a nucleus ($m_e/M \approx 10^{-3} - 10^{-5}$), the dynamics of electrons and nuclei in solids can, to a good approximation, be decoupled. Within the Born-Oppenheimer approximation, the many-electron wave function and energy may then be obtained by solving a time-independent Schrödinger equation in which the nuclear positions are considered fixed. The non-relativistic Born-Oppenheimer Hamiltonian takes the form,

$$\hat{H} = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_i \sum_\alpha \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{d}_\alpha|} + \frac{1}{2} \sum_i \sum_{j \neq i} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad (2.1)$$

where \mathbf{r}_i are the electron positions, \mathbf{d}_α are the nuclear positions, and Z_α are the nuclear charges.*

The first quantum mechanical studies of chemically bonded systems appeared soon after the birth of quantum mechanics. Heitler and London (1927) calculated the binding energy and internuclear separation of H_2 by approximating the wave function as an antisymmetrized product of atomic $1s$ functions centered on the two nuclei. A different approach was taken by Hartree (1928),

*We use Hartree atomic units, $e = m_e = \hbar = 4\pi\epsilon_0 = 1$, for all equations.

Fock (1930), and Slater (1930), who argued that it is reasonable to replace the intractable system of interacting electrons by one involving independent electrons moving in a self-consistent field. Coulson and Fischer (1949) showed that the Heitler-London and Hartree-Fock (HF) wave functions were exact in the limiting cases of strong electron correlation and weak electron correlation, respectively, and that the truth lay somewhere in between. It turns out that many chemically bonded systems are weakly correlated, so that the HF approach is normally a good starting point. This point of view underlies the molecular orbital theory developed by Hund (1928), Mulliken (1928), Hückel (1931a, 1931b, 1932), and others.

Calculating the electronic structure of solids poses a formidable challenge because of the large number of particles involved. The first steps were taken by Sommerfeld and others, who made the bold proposal that in metals the forces on the electrons due to the ions approximately cancel the forces due to the other electrons (Sommerfeld and Bethe, 1933), leading to the free-electron model. The work of Bloch (1928) added the important concepts of Bloch waves and energy bands. It was soon realized, however, that HF theory fails to give a proper description of metals, and hence that correlation effects have to be included from the outset.

The most important and widely used electronic structure method currently applied to solids is based on the many-electron density-functional theory (DFT) developed by Hohenberg and Kohn (1964) and Kohn and Sham (1965). This approach reduces the complicated many-electron problem to an independent electron problem with an effective one-electron potential that depends on the electron density. Although at first glance it appears that DFT must be an approximate method, it is in principle capable of describing electron correlation effects exactly. DFT has been the main tool used to study the electronic structure of solids for more than two decades (Jones and Gunnarsson, 1989; Parr and Yang, 1989; Dreizler and Gross, 1990), and over the last decade has also become popular in quantum chemistry.

The HF and DFT methods are very relevant to our discussion of QMC. The QMC techniques we focus on here rely on the availability of reasonable approximations to the many-electron wave function, which are often constructed using one-electron orbitals obtained from HF or DFT calculations. A brief introduction to these methods and to correlated wave function methods is given in the following subsections. Many of the concepts and equations will be referred to later in the article.

B. Hartree-Fock theory

The HF approximation (Fock, 1930; Slater, 1930) is the simplest theory that correctly incorporates the permutation symmetry of the wave function. The many-electron

wave function must be antisymmetric under particle exchange, so that

$$\Psi(\dots, \mathbf{x}_i, \dots, \mathbf{x}_j, \dots) = -\Psi(\dots, \mathbf{x}_j, \dots, \mathbf{x}_i, \dots) , \quad (2.2)$$

where $\mathbf{x}_i = \{\mathbf{r}_i, \sigma_i\}$ represents the space and spin coordinates of an electron. The antisymmetry ensures that no two electrons can have the same set of quantum numbers and that the Pauli exclusion principle is satisfied. The simplest wave function with the required antisymmetry is a Slater determinant,

$$D(\mathbf{X}) = \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_1(\mathbf{x}_2) & \dots & \psi_1(\mathbf{x}_N) \\ \psi_2(\mathbf{x}_1) & \psi_2(\mathbf{x}_2) & \dots & \psi_2(\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_N(\mathbf{x}_1) & \psi_N(\mathbf{x}_2) & \dots & \psi_N(\mathbf{x}_N) \end{vmatrix} , \quad (2.3)$$

where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ is shorthand for the list of all the electron coordinates. In most cases the single-particle orbitals are assumed to be products of spatial and spin factors,

$$\psi_i(\mathbf{x}_j) = \psi_i(\mathbf{r}_j) \delta_{\sigma_i, \sigma_j} , \quad (2.4)$$

where $\delta_{\sigma_i, \sigma_j} = 1$ if $\sigma_j = \sigma_i$ and zero otherwise. If the determinant contains N_\uparrow orbitals with $\sigma_i = \uparrow$ and $N_\downarrow = N - N_\uparrow$ with $\sigma_i = \downarrow$, it is an eigenfunction of \hat{S}_z with eigenvalue $(N_\uparrow - N_\downarrow)/2$.

The exact ground state wave function of an interacting system cannot be represented as a single Slater determinant, but a Slater determinant can nevertheless be used as a variational trial function. If this trial function is optimized by minimizing the expectation value of \hat{H} with respect to the orbitals $\psi_i(\mathbf{r})$, one obtains the self-consistent HF equations,

$$\begin{aligned} \epsilon_i \psi_i(\mathbf{r}) = & \left(-\frac{1}{2} \nabla^2 + V_{\text{ion}}(\mathbf{r}) \right) \psi_i(\mathbf{r}) \\ & + \sum_j \int d\mathbf{r}' \frac{|\psi_j(\mathbf{r}')|^2}{|\mathbf{r} - \mathbf{r}'|} \psi_i(\mathbf{r}) \\ & - \sum_j \delta_{\sigma_i, \sigma_j} \int d\mathbf{r}' \frac{\psi_j^*(\mathbf{r}') \psi_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \psi_j(\mathbf{r}) , \end{aligned} \quad (2.5)$$

where the Lagrange multipliers ϵ_i arise from the orthonormality constraints on the single-particle orbitals. The ionic potential is

$$V_{\text{ion}}(\mathbf{r}) = - \sum_{\alpha} \frac{Z_{\alpha}}{|\mathbf{r} - \mathbf{d}_{\alpha}|} , \quad (2.6)$$

as in Eq. (2.1), while the other two terms, known as the Hartree and exchange terms, describe the electron-electron interactions. The effect of the exchange term is to keep electrons of like spin apart and, as a result, each electron has around it a Fermi or exchange hole containing unit positive charge.

C. Post Hartree-Fock methods

The single-determinant HF theory includes the exchange effects arising from the antisymmetry of the many-electron wave function, but neglects the electronic correlations caused by the electron-electron Coulomb repulsion. Correlation energies are a small fraction of the total energy, but they can be very important. For example, the correlation energy of a nitrogen molecule amounts to only 0.5% of the total electronic energy, but accounts for nearly 50% of the molecular binding energy.

Correlation can be included by using a linear combination of determinants in a post Hartree-Fock method. However, the central problem with such expansions is that very large numbers of determinants are needed to describe many-electron wave functions accurately. There are two reasons for this poor convergence. The first, which applies equally in small and large systems, is that many determinants are needed to describe the cusp-like gradient discontinuities that occur whenever two electrons have the same position (see Sec. IV). The second is that the required number of determinants increases very rapidly with the system size.

In a full configuration interaction calculation one includes all the determinants that can be formed from the molecular orbitals calculated with a particular basis set. Consider, for example, a system of N electrons occupying states chosen from a set of $2N$ molecular orbitals. The total number of N -electron Slater determinants is

$$\frac{(2N)!}{N!N!} \approx \frac{e^{2N \ln 2N}}{e^{2N \ln N}} = e^{2N \ln 2}, \quad (2.7)$$

for large N . The number of determinants therefore rises exponentially with N . The full configuration interaction approach in conjunction with a good basis set can only be applied to small systems and the current practical limit for highly accurate calculations is reached for small molecules such as H_2O .

One way of reducing the computational cost is to include only the most important determinants. This is normally done by considering low energy excitations from a reference determinant, which is normally the HF ground state. For example, if single and double excitations are included we have the configuration interaction singles and doubles method, for which the computational cost scales as N^6 . Unfortunately, such truncated configuration interaction methods are not size consistent; that is, the energy does not scale linearly with the number of electrons. For example, in a configuration interaction singles and doubles calculation for N widely separated hydrogen molecules the correlation energy increases only as \sqrt{N} . Such a method is clearly unsuitable for applications to solids. The size consistency problem can be overcome via coupled cluster (CC) expansions (Čížek, 1969). These implicitly include all excitations from the reference determinant, although the coefficients in the expansion are

approximated and the method is non-variational. CC methods are capable of yielding highly accurate results and are direct competitors of QMC methods for molecular calculations, but they are very expensive in large systems. For example, the computational cost of a CCSD calculation (coupled cluster with single and double excitations) scales as N^6 . HF and post HF methods are described in a straightforward manner in the book by Szabo and Ostlund (1989).

D. Density-functional theory

Computational methods based on DFT are now the most popular approach for calculating the electronic properties of solids and large molecules. DFT is a formally exact theory based on a theorem of Hohenberg and Kohn (1964), which states that the ground state properties of a many-electron system may be obtained by minimizing a functional $E[n(\mathbf{r})]$ of the electron number density $n(\mathbf{r})$. The minimum value of the functional is the exact ground state energy, and the minimum is attained when $n(\mathbf{r})$ is the exact ground state density.

Kohn and Sham (1965) introduced the idea of an auxiliary non-interacting system with the same electron density as the real system. This enabled them to express the electron density of the interacting system in terms of the one-electron wave functions of the non-interacting system,

$$n(\mathbf{r}) = \sum_{i=1}^N |\psi_i(\mathbf{r})|^2, \quad (2.8)$$

and to write the Hohenberg-Kohn energy functional in the form,

$$E[n(\mathbf{r})] = -\frac{1}{2} \sum_{i=1}^N \int \psi_i^*(\mathbf{r}) \nabla^2 \psi_i(\mathbf{r}) d\mathbf{r} + \int n(\mathbf{r}) V_{\text{ion}}(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \iint \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{xc}[n(\mathbf{r})], \quad (2.9)$$

where the terms on the right hand side are the kinetic energy of the non-interacting system with electron density $n(\mathbf{r})$, the energy of interaction with the ionic potential, the Hartree energy, and the exchange-correlation (XC) energy. Equation (2.9) may be taken as the definition of the XC energy functional $E_{xc}[n(\mathbf{r})]$, which is not known exactly and has to be approximated.

By minimizing the total energy functional of Eq. (2.9), Kohn and Sham showed that one could calculate the ground state ψ_i , and hence the ground state number density and energy, by solving the self-consistent set of equations,

$$\left(-\frac{1}{2} \nabla^2 + V_{\text{ion}}(\mathbf{r}) + V_H(\mathbf{r}) + V_{xc}(\mathbf{r}) \right) \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}), \quad (2.10)$$

where the Hartree potential is

$$V_H(\mathbf{r}) = \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad (2.11)$$

and the XC potential is

$$V_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[n(\mathbf{r})]}{\delta n(\mathbf{r})}. \quad (2.12)$$

The best known approximation for $E_{xc}[n(\mathbf{r})]$ is the local density approximation (LDA),

$$E_{xc}^{\text{LDA}}[n(\mathbf{r})] = \int \epsilon_{xc}^{\text{hom}}(n(\mathbf{r})) n(\mathbf{r}) d\mathbf{r}, \quad (2.13)$$

where $\epsilon_{xc}^{\text{hom}}(n)$ is the XC energy per electron in a uniform electron gas of density n . The non-uniform electron gas at \mathbf{r} is therefore treated as if it was part of a uniform electron gas of constant density $n = n(\mathbf{r})$. This approximation is obviously accurate when the electron density is almost uniform, but also works surprisingly well when the distribution of electrons is strongly inhomogeneous, such as at surfaces and in molecules. The analogous approximation for spin-polarized systems, which is known as the local spin-density approximation (LSDA), has also proved very successful. The LDA and LSDA were used throughout the development of DFT and are still widely used in studies of solids.

In cases when the LDA is not accurate enough it seems sensible to try approximating $E_{xc}[n]$ in terms of the local density and its gradient. It turns out that simple gradient expansions are rather badly behaved, but that better approximations can be devised by expressing ϵ_{xc} as a carefully chosen non-linear function of $n(\mathbf{r})$ and $|\nabla n(\mathbf{r})|$. This is the idea behind the generalized gradient approximation (GGA) (Langreth and Mehl, 1983; Becke, 1988; Perdew *et al.*, 1992), which is now used in most quantum chemical applications of DFT. Several non-local functionals have also been proposed, of which the best known are the average density approximation and weighted density approximation of Gunnarsson, Jonson, and Lundqvist (1979).

DFT results obtained using the best available approximate XC energy functionals are typically an order of magnitude less accurate than good QMC results, but since DFT is much less computationally demanding than QMC it has a much wider variety of interesting applications. Computational experiments based on DFT are already used in physics, chemistry, biochemistry, materials science, and geophysics, and are beginning to be used in industrially important fields such as drug design.

E. Quantum Monte Carlo methods

There are many different QMC methods, but this review concentrates on only two: variational quantum

Monte Carlo (VMC) and fixed-node diffusion quantum Monte Carlo (DMC). Like all QMC methods, these are closely related to Monte Carlo methods used in classical statistical mechanics (Nightingale, 1999). In the VMC method expectation values are calculated via Monte Carlo integration over the $3N$ -dimensional space of electron coordinates. The more sophisticated DMC method is a projector approach in which a stochastic imaginary-time evolution is used to improve a starting trial wave function.

Other Monte Carlo methods, such as auxiliary-field QMC (reviewed by Senatore and March, 1994) and path-integral QMC (reviewed by Ceperley, 1995a, 1995b), may also be used to study interacting many-electron systems. Recent progress along these lines includes the development of the shifted-contour auxiliary field method (Baer, Head-Gordon and Neuhauser, 1998; Baer and Neuhauser, 2000), which has already been applied to small molecules, and the introduction of the reptation QMC method (Baroni and Moroni, 1999), which is an interesting hybrid of path-integral and diffusion QMC. The more established restricted path-integral fermion Monte Carlo method (Ceperley, 1995b) has been used to investigate the formation of a gas of molecular hydrogen from a neutral system of electrons and protons at high temperature (Magro *et al.*, 1996), and to calculate the forces between protons in an electron gas (Zong and Ceperley, 1998). Unlike VMC and DMC, which are zero temperature methods in which one considers a single wave function, path-integral and auxiliary-field QMC may both be used to compute expectation values at finite temperature. These methods have not yet been applied to real solids, however, and we shall not discuss them further in this review. We also restrict ourselves to continuum models and choose not to consider lattice models such as the Hubbard model. Although the methods described in this article may be applied to lattice problems, other QMC methods such as the auxiliary field approach are also useful in such systems (Blankenbecler, Scalapino, and Sugar, 1981).

McMillan (1965) was the first to use the VMC method in his study of liquid ^4He . One of the first applications to a many-fermion system was by Ceperley, Chester, and Kalos (1977). A very early calculation by Donsker and Kac (1950) employed a type of projector Monte Carlo method, but their algorithm was inefficient and fails for unbounded potentials such as the Coulomb interaction. Kalos (1962, 1967) developed the projector Green Function Monte Carlo method and enormously improved its efficiency by introducing the concept of importance sampling, in which a trial or guiding wave function is used to steer the Monte Carlo moves (Kalos, Levesque and Verlet, 1974). The DMC algorithm (Grimm and Storer, 1971; Anderson, 1975, 1976) discussed here may be viewed as an accurate and convenient approximation to the full Green Function Monte Carlo algorithm, and

was developed later. Importance sampling was introduced into the DMC algorithm by Grimm and Storer (1971).

Because of the fermionic nature of the many-electron wave function it must have positive and negative regions. This simple fact underlies the infamous fermion sign problem, which plagues all projector QMC methods. The search for exact solutions to the sign problem is still active, but to date no entirely satisfactory exact solution exists. DMC simulations of large systems therefore use the fixed-node approximation of Anderson (1975). A many-electron trial wave function is chosen and used to define a trial nodal surface. (The nodal surface of a wave function $\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ is the $3N - 1$ dimensional surface on which $\Psi = 0$.) The fixed-node DMC algorithm then projects out the many-electron wave function with the lowest possible energy expectation value consistent with that fixed nodal surface. The results are not exact unless the trial nodal surface is exact, but the fixed-node method is computationally stable and the calculated energy is variational and often very accurate.

The importance-sampled fixed-node fermion DMC algorithm was first applied to the electron gas by Ceperley and Alder (1980). The first application to solid hydrogen was in 1987 (Ceperley and Alder, 1987), and the first VMC and DMC calculations of solids containing heavier atoms followed soon after (Fahy, Wang, and Louie, 1988, 1990a; Li, Ceperley and Martin, 1991). Important recent technical developments have included the introduction of variance minimization techniques to optimize trial wave functions (Umrigar, Wilson and Wilkins, 1988), and the use of non-local pseudopotentials in VMC and DMC calculations (Hurley and Christiansen, 1987; Hammond, Reynolds, and Lester, Jr., 1987; Fahy, Wang and Louie, 1988; Mitas, Shirley, and Ceperley, 1991). These improvements have stimulated applications to a much wider range of systems.

The computer time required to calculate the energy of a system to some given accuracy using the fermion VMC and DMC methods effectively scales as N^3 (see Sec. X.E). Of course QMC does not offer a free lunch — the price to be paid for this advantageous scaling with N is that the answer is obtained with a statistical error bar that decays only as the inverse of the square root of the computer time. The total amount of time required for accurate QMC calculations is therefore quite large. However, QMC algorithms are intrinsically parallel, which is becoming a significant computational advantage as the availability of powerful and relatively inexpensive parallel machines improves.

For small systems the DMC method is capable of reaching chemical accuracy, which is usually defined as ~ 1 kcal per mole (≈ 0.04 eV per molecule). The advantageous scaling with system size means that the attainable accuracy does not fall off rapidly as the number

of electrons N increases. Accurate trial wave functions are necessary to achieve high accuracy at a reasonable computational cost, but since the DMC wave function is generated stochastically the results are largely free of the errors caused by the limited basis set used in most other techniques. By contrast, the quality of results obtained using the less accurate VMC method is entirely determined by the quality of the trial wave function.

The VMC and DMC methods are best suited to calculating energies because these have the very advantageous zero-variance property; as the trial wave function approaches the exact ground state (or any other exact energy eigenstate) the statistical fluctuations in the energy reduce to zero. Other ground state expectation values, such as static correlation functions, can also be calculated, but the absence of a zero-variance property makes this more problematic. The VMC and DMC methods are less well adapted to studying excited states, but have nevertheless been used successfully to calculate a wide range of excited state properties of atoms, molecules, and solids.

III. MONTE CARLO METHODS

A. Statistical foundations

The archetypal example of a MC simulation is the evaluation of a multi-dimensional integral by sampling the integrand statistically and averaging the sampled values. In conventional quadrature methods the accuracy depends on the fineness of the integration mesh. If one uses a d -dimensional cubic mesh to evaluate a d -dimensional integral using Simpson's rule, for example, the error scales as $M^{-4/d}$, where M is the total number of mesh points. Therefore, as the dimension d increases (we have d as large as a few thousand in some of our applications), the error falls off increasingly slowly with M . Using MC methods, on the other hand, the statistical error in the value of the integral decreases as the square root of the number of sampling points used, regardless of dimensionality. This is a consequence of the *central limit theorem*, one of the cornerstones of the mathematical theory of statistics.

Suppose we define a $3N$ -dimensional vector \mathbf{R} by

$$\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) , \quad (3.1)$$

where \mathbf{r}_i is the position of the i 'th electron. A particular value of \mathbf{R} is sometimes called a *walker*, a *configuration*, or a *psip* in the QMC literature. (Note that here and throughout the rest of Sec. III we consider a system of spinless fermions for simplicity; Sec. IV discusses the minor extensions required to treat the spin properly.) The probability density of finding the electrons in the configuration \mathbf{R} will be denoted by $\mathcal{P}(\mathbf{R})$, where

$$\mathcal{P}(\mathbf{R}) \geq 0, \quad (3.2)$$

$$\int d\mathbf{R} \mathcal{P}(\mathbf{R}) = 1. \quad (3.3)$$

Let $\{\mathbf{R}_m : m = 1, M\}$ be a set of mutually independent (uncorrelated) configurations distributed according to the probability distribution $\mathcal{P}(\mathbf{R})$. We define a new random variable \mathbf{Z}_f by

$$\mathbf{Z}_f = \frac{f(\mathbf{R}_1) + f(\mathbf{R}_2) + \dots + f(\mathbf{R}_M)}{M}, \quad (3.4)$$

where $f(\mathbf{R})$ is any reasonable function with mean μ_f and variance σ_f^2 given by

$$\mu_f = \int d\mathbf{R} f(\mathbf{R}) \mathcal{P}(\mathbf{R}), \quad (3.5)$$

$$\sigma_f^2 = \int d\mathbf{R} [f(\mathbf{R}) - \mu_f]^2 \mathcal{P}(\mathbf{R}). \quad (3.6)$$

Then, under rather general conditions (Feller, 1968), the central limit theorem tells us that, for large enough M , \mathbf{Z}_f is normally distributed with mean μ_f and standard deviation σ_f/\sqrt{M} . This implies that, regardless of $\mathcal{P}(\mathbf{R})$, the mean value of a large number of measurements of some function of \mathbf{R} will be a good estimator of the mean of that function with respect to $\mathcal{P}(\mathbf{R})$. Moreover, the standard deviation of the mean of the measurements will decrease as $1/\sqrt{M}$, irrespective of the dimension of the integral.

These ideas can be applied to the evaluation of integrals such as

$$I = \int d\mathbf{R} g(\mathbf{R}). \quad (3.7)$$

First we introduce an “importance function” $\mathcal{P}(\mathbf{R})$ and rewrite the integral in the form

$$I = \int d\mathbf{R} f(\mathbf{R}) \mathcal{P}(\mathbf{R}), \quad (3.8)$$

where $f(\mathbf{R}) \equiv g(\mathbf{R})/\mathcal{P}(\mathbf{R})$. The importance function is chosen such that it obeys Eqs. (3.2) and (3.3) and hence may be interpreted as a probability density. In principle, the value of I may now be obtained by drawing an infinite set of random vectors from the distribution $\mathcal{P}(\mathbf{R})$ and computing the sample average:

$$I = \lim_{M \rightarrow \infty} \left\{ \frac{1}{M} \sum_{m=1}^M f(\mathbf{R}_m) \right\}. \quad (3.9)$$

A MC estimate of I may be obtained by averaging over a large but finite sample of vectors drawn from $\mathcal{P}(\mathbf{R})$:

$$I \approx \frac{1}{M} \sum_{m=1}^M f(\mathbf{R}_m). \quad (3.10)$$

The variance of the estimate of the integral is then σ_f^2/M , which itself can be estimated as

$$\frac{\sigma_f^2}{M} \approx \frac{1}{M(M-1)} \sum_{m=1}^M \left[f(\mathbf{R}_m) - \frac{1}{M} \sum_{n=1}^M f(\mathbf{R}_n) \right]^2, \quad (3.11)$$

and an estimate of the size of the error bar on the computed value of I is $\pm \sigma_f/\sqrt{M}$.

A judicious choice of the importance function $\mathcal{P}(\mathbf{R})$ significantly reduces the variance for a fixed sample size. It can be readily shown that the importance function that gives the smallest variance is $\mathcal{P}(\mathbf{R}) = |g(\mathbf{R})/I|$. However, since I is not known, the integrand $f(\mathbf{R}) = g(\mathbf{R})/\mathcal{P}(\mathbf{R})$ corresponding to this importance function is also unknown. In general, the best that one can do is make $\mathcal{P}(\mathbf{R})$ as similar to $|g(\mathbf{R})/I|$ as possible.

B. The Metropolis algorithm

When using the QMC method described in Sec. III.A to evaluate multi-dimensional integrals, it is necessary to sample complicated probability distributions in high-dimensional spaces. In general, the normalizations of these distributions are unknown and they are so complicated that they cannot be sampled directly. The Metropolis rejection algorithm (Metropolis *et al.*, 1953) has the great advantage that it allows an arbitrarily complex distribution to be sampled in a straightforward way without knowledge of its normalization.

The Metropolis algorithm generates the sequence of sampling points \mathbf{R}_m by moving a single walker according to the following rules:

1. Start the walker at a random position \mathbf{R} .
2. Make a trial move to a new position \mathbf{R}' chosen from some probability density function $T(\mathbf{R}' \leftarrow \mathbf{R})$. After the trial move the probability that the walker initially at \mathbf{R} is now in the volume element $d\mathbf{R}'$ is $d\mathbf{R}' \times T(\mathbf{R}' \leftarrow \mathbf{R})$.
3. Accept the trial move to \mathbf{R}' with probability

$$A(\mathbf{R}' \leftarrow \mathbf{R}) = \text{Min} \left(1, \frac{T(\mathbf{R} \leftarrow \mathbf{R}') \mathcal{P}(\mathbf{R}')}{T(\mathbf{R}' \leftarrow \mathbf{R}) \mathcal{P}(\mathbf{R})} \right). \quad (3.12)$$

If the trial move is accepted the point \mathbf{R}' becomes the next point on the walk; if the trial move is rejected the point \mathbf{R} becomes the next point on the walk. If $\mathcal{P}(\mathbf{R})$ is high, most trial moves away from \mathbf{R} will be rejected and the point \mathbf{R} may occur many times in the set of points making up the random walk.

4. Return to step 2 and repeat.

The initial points generated by this algorithm depend on the starting point and should be discarded. Eventually, however, the simulation settles down and sets of points snipped out of the random walk are distributed according to $\mathcal{P}(\mathbf{R})$.

To understand how this algorithm works, imagine an enormous number of walkers all executing random walks according to the above rules. After an equilibration period, we assume that the distribution of walkers settles down to a unique steady equilibrium state in which the average number of walkers in the volume element $d\mathbf{R}$ is denoted by $n(\mathbf{R})d\mathbf{R}$. Once the equilibrium has been established, we also assume that the average number of walkers moving from $d\mathbf{R}$ to $d\mathbf{R}'$ in one time step is the same as the number moving from $d\mathbf{R}'$ to $d\mathbf{R}$. Our everyday experience of diffusion and other random walk processes suggests that these assumptions are probably correct, but readers who wish to see a clear derivation of the conditions under which they hold are referred to Feller (1968).

Since the probability that the next move of a walker at \mathbf{R} takes it to $d\mathbf{R}'$ is $d\mathbf{R}'A(\mathbf{R}' \leftarrow \mathbf{R})T(\mathbf{R}' \leftarrow \mathbf{R})$, the average number moving from $d\mathbf{R}$ to $d\mathbf{R}'$ during a single move will be

$$d\mathbf{R}'A(\mathbf{R}' \leftarrow \mathbf{R})T(\mathbf{R}' \leftarrow \mathbf{R}) \times n(\mathbf{R})d\mathbf{R}. \quad (3.13)$$

This must be balanced by the number moving from $d\mathbf{R}'$ to $d\mathbf{R}$,

$$\begin{aligned} A(\mathbf{R}' \leftarrow \mathbf{R})T(\mathbf{R}' \leftarrow \mathbf{R})n(\mathbf{R})d\mathbf{R}d\mathbf{R}' \\ = A(\mathbf{R} \leftarrow \mathbf{R}')T(\mathbf{R} \leftarrow \mathbf{R}')n(\mathbf{R}')d\mathbf{R}'d\mathbf{R}, \end{aligned} \quad (3.14)$$

and hence the equilibrium distribution $n(\mathbf{R})$ satisfies

$$\frac{n(\mathbf{R})}{n(\mathbf{R}')} = \frac{A(\mathbf{R} \leftarrow \mathbf{R}')T(\mathbf{R} \leftarrow \mathbf{R}')}{A(\mathbf{R}' \leftarrow \mathbf{R})T(\mathbf{R}' \leftarrow \mathbf{R})}. \quad (3.15)$$

Since the ratio of Metropolis algorithm acceptance probabilities is

$$\frac{A(\mathbf{R} \leftarrow \mathbf{R}')}{A(\mathbf{R}' \leftarrow \mathbf{R})} = \frac{T(\mathbf{R}' \leftarrow \mathbf{R})\mathcal{P}(\mathbf{R})}{T(\mathbf{R} \leftarrow \mathbf{R}')\mathcal{P}(\mathbf{R}')}, \quad (3.16)$$

it follows that

$$\frac{n(\mathbf{R})}{n(\mathbf{R}')} = \frac{\mathcal{P}(\mathbf{R})}{\mathcal{P}(\mathbf{R}')}. \quad (3.17)$$

The equilibrium walker density $n(\mathbf{R})$ is therefore proportional to $\mathcal{P}(\mathbf{R})$, and the probability of finding any given walker in $d\mathbf{R}$ is $\mathcal{P}(\mathbf{R})d\mathbf{R}$ as required. A rigorous derivation of this result is given by Feller (1968).

C. Variational Monte Carlo

The variational Monte Carlo (VMC) method is the simpler of the two quantum Monte Carlo methods discussed in this review. It is based on a combination of the variational principle and the Monte Carlo evaluation of integrals discussed in Sec. III.A. The VMC method relies on the availability of a trial wave function Ψ_T that is a reasonably good approximation to the true ground state wave function. The subject of how to produce good trial wave functions is dealt with in Sec. IV. The trial wave function must satisfy some basic conditions. Both Ψ_T and $\nabla\Psi_T$ must be continuous wherever the potential is finite, and the integrals $\int \Psi_T^* \Psi_T$ and $\int \Psi_T^* \hat{H} \Psi_T$ must exist. To keep the variance of the energy finite we also require $\int \Psi_T^* \hat{H}^2 \Psi_T$ to exist.

The expectation value of \hat{H} evaluated with a trial wave function Ψ_T provides a rigorous upper bound on the exact ground state energy E_0 :

$$E_V = \frac{\int \Psi_T^*(\mathbf{R}) \hat{H} \Psi_T(\mathbf{R}) d\mathbf{R}}{\int \Psi_T^*(\mathbf{R}) \Psi_T(\mathbf{R}) d\mathbf{R}} \geq E_0. \quad (3.18)$$

In a VMC simulation this bound is calculated using the Metropolis Monte Carlo method. Equation (3.18) is rewritten in the form

$$E_V = \frac{\int |\Psi_T(\mathbf{R})|^2 \left(\Psi_T(\mathbf{R})^{-1} \hat{H} \Psi_T(\mathbf{R}) \right) d\mathbf{R}}{\int |\Psi_T(\mathbf{R})|^2 d\mathbf{R}}, \quad (3.19)$$

and the Metropolis algorithm is used to sample a set of points $\{\mathbf{R}_m : m = 1, M\}$ from the configuration-space probability density $\mathcal{P}(\mathbf{R}) = |\Psi_T(\mathbf{R})|^2 / \int |\Psi_T(\mathbf{R})|^2 d\mathbf{R}$. At each of these points the “local energy” $E_L(\mathbf{R}) = \Psi_T(\mathbf{R})^{-1} \hat{H} \Psi_T(\mathbf{R})$ is evaluated and the average energy accumulated:

$$E_V \approx \frac{1}{M} \sum_{m=1}^M E_L(\mathbf{R}_m). \quad (3.20)$$

The trial moves are usually sampled from a Gaussian centered on the current position of the walker, the variance of the Gaussian being chosen such that average acceptance probability is roughly 50% or such that the diffusion constant of the random walk is maximized. The flow chart shown in Fig. 1 illustrates how a typical VMC simulation works.

Expectation values of operators other than the Hamiltonian may also be expressed as $3N$ -dimensional integrals and evaluated in an analogous way. The variance of the QMC result for any given observable depends on the details of the integrand, suggesting that it may be possible to reduce the variance by modifying the estimator in such a way that its expectation value is unaffected. Very recently, Assaraf and Caffarel (1999) have demonstrated that a significant variance reduction may be obtained in

both VMC and DMC using this technique, although their approach has only been tested in small systems so far.

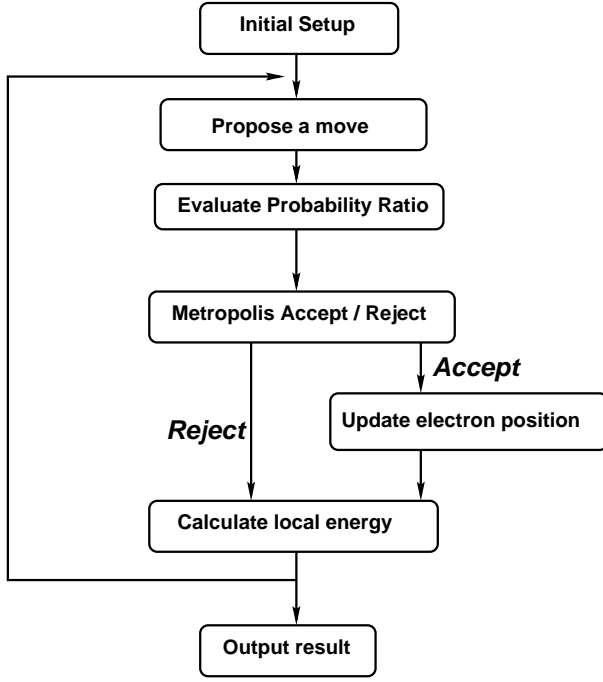


FIG. 1. Flow chart illustrating the VMC algorithm.

D. Diffusion Monte Carlo

1. The imaginary-time Schrödinger equation

Diffusion Monte Carlo (DMC) is a stochastic projector method for solving the imaginary-time many-body Schrödinger equation,

$$-\partial_t \Phi(\mathbf{R}, t) = (\hat{H} - E_T) \Phi(\mathbf{R}, t), \quad (3.21)$$

where t is a real variable measuring the progress in imaginary time, $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ is a $3N$ -dimensional vector specifying the coordinates of all N electrons, and E_T is an energy offset. Eq. (3.21) may be rewritten in the integral form

$$\Phi(\mathbf{R}, t + \tau) = \int G(\mathbf{R} \leftarrow \mathbf{R}', \tau) \Phi(\mathbf{R}', t) d\mathbf{R}', \quad (3.22)$$

where

$$G(\mathbf{R} \leftarrow \mathbf{R}', \tau) = \langle \mathbf{R} | \exp[-\tau(\hat{H} - E_T)] | \mathbf{R}' \rangle \quad (3.23)$$

is a Green function that obeys the same equation as the wave function,

$$-\partial_t G(\mathbf{R} \leftarrow \mathbf{R}', t) = (\hat{H}(\mathbf{R}) - E_T) G(\mathbf{R} \leftarrow \mathbf{R}', t), \quad (3.24)$$

with the initial condition $G(\mathbf{R} \leftarrow \mathbf{R}', 0) = \delta(\mathbf{R} - \mathbf{R}')$. Using the spectral expansion

$$\exp(-\tau \hat{H}) = \sum_i |\Psi_i\rangle \exp(-\tau E_i) \langle \Psi_i|, \quad (3.25)$$

the Green function can be expressed as

$$G(\mathbf{R} \leftarrow \mathbf{R}', \tau) = \sum_i \Psi_i(\mathbf{R}) e^{-\tau(E_i - E_T)} \Psi_i^*(\mathbf{R}'), \quad (3.26)$$

where $\{\Psi_i\}$ and $\{E_i\}$ denote the complete sets of eigenfunctions and eigenvalues of \hat{H} , respectively. Eqs. (3.21) and (3.22) are equivalent as can be readily demonstrated by differentiating Eq. (3.22) with respect to τ and setting $\tau = 0$.

It is straightforward to show that as $\tau \rightarrow \infty$ the operator $\exp[-\tau(\hat{H} - E_T)]$ projects out the lowest eigenstate $|\Psi_0\rangle$ that has non-zero overlap with the chosen $|\Phi(t=0)\rangle = |\Phi_{\text{init}}\rangle$:

$$\begin{aligned} & \lim_{\tau \rightarrow \infty} \langle \mathbf{R} | \exp[-\tau(\hat{H} - E_T)] | \Phi_{\text{init}} \rangle \\ &= \lim_{\tau \rightarrow \infty} \int G(\mathbf{R} \leftarrow \mathbf{R}', \tau) \Phi_{\text{init}}(\mathbf{R}') d\mathbf{R}' \\ &= \lim_{\tau \rightarrow \infty} \sum_i \Psi_i(\mathbf{R}) \exp[-\tau(E_i - E_T)] \langle \Psi_i | \Phi_{\text{init}} \rangle \\ &= \lim_{\tau \rightarrow \infty} \Psi_0(\mathbf{R}) \exp[-\tau(E_0 - E_T)] \langle \Psi_0 | \Phi_{\text{init}} \rangle. \end{aligned} \quad (3.27)$$

By adjusting E_T to equal E_0 , the exponential factor in the last line of Eq. (3.27) can be made constant, while the higher states in the previous line are all exponentially damped because their energies are higher than E_0 . This fundamental property of the projector $\exp[-\tau(\hat{H} - E_T)]$ is the basis of the diffusion Monte Carlo method and similar projector-based approaches.

Let us for a moment neglect the potential terms in the Hamiltonian so that Eq. (3.21) simplifies to

$$\partial_t \Phi(\mathbf{R}, t) = \frac{1}{2} \sum_{i=1}^N \nabla_i^2 \Phi(\mathbf{R}, t). \quad (3.28)$$

This is a *diffusion equation in a $3N$ -dimensional space*, the Green function for which is a $3N$ -dimensional Gaussian with variance τ in each dimension,

$$G(\mathbf{R} \leftarrow \mathbf{R}', \tau) = (2\pi\tau)^{-3N/2} \exp\left[-\frac{|\mathbf{R} - \mathbf{R}'|^2}{2\tau}\right]. \quad (3.29)$$

It is easy to verify that this Gaussian form tends to $\delta(\mathbf{R} - \mathbf{R}')$ as $\tau \rightarrow 0$ and that it satisfies Eq. (3.28).

In the theory of stochastic processes Eq. (3.28) is called the *master equation of a diffusion stochastic process* (Karlin and Taylor, 1981) and its solution $\Phi(\mathbf{R}, t)$ describes the distribution of the diffusing Brownian particles in space and time. The correspondence *Brownian particles* \longleftrightarrow *distribution function* works both ways and we can represent the distribution $\Phi(\mathbf{R}, t)$ by a set of discrete Brownian sampling points or random walkers:

$$\Phi(\mathbf{R}, t) \xrightarrow{\text{discrete}} \sum_k \delta(\mathbf{R} - \mathbf{R}_k) . \quad (3.30)$$

The Green function of Eq. (3.29) can be interpreted as a transition probability density for the evolution of the walkers. By substituting Eq. (3.30) into Eq. (3.22) we obtain

$$\Phi(\mathbf{R}, t + \tau) \xrightarrow{\text{discrete}} \sum_k G(\mathbf{R} \leftarrow \mathbf{R}_k, \tau) . \quad (3.31)$$

The result is a set of Gaussians each with a variance of $3N\tau$. In order to restore the original discrete representation each Gaussian is *sampled* by a new delta function that defines the evolved position of the walker. The procedure of propagation/resampling is then repeated until convergence is reached. The result of this process is a set of *samples* distributed according to the solution of Eq. (3.28).

Now consider the full Hamiltonian, which includes both kinetic and potential energy terms. Except in a few special cases, once the particles are interacting we do not know explicit expressions for the exact Green function and have to use approximations. An approximation to the Green function can be obtained using the Trotter-Suzuki formula for the exponential of a sum of operators \hat{A} and \hat{B} :

$$e^{-\tau(\hat{A}+\hat{B})} = e^{-\tau\hat{B}/2} e^{-\tau\hat{A}} e^{-\tau\hat{B}/2} + \mathcal{O}[\tau^3] . \quad (3.32)$$

If we write $\hat{H} = \hat{T} + \hat{V}$, where \hat{T} is the full N -electron kinetic energy operator and \hat{V} is the total potential energy operator (including both electron-ion and electron-electron terms), application of the Trotter-Suzuki formula with $\hat{A} = \hat{T}$ and $\hat{B} = \hat{V} - E_T$ gives:

$$G(\mathbf{R} \leftarrow \mathbf{R}', \tau) = \langle \mathbf{R} | e^{-\tau(\hat{T}+\hat{V}-E_T)} | \mathbf{R}' \rangle \\ \approx e^{-\tau[V(\mathbf{R})-E_T]/2} \langle \mathbf{R} | e^{-\tau\hat{T}} | \mathbf{R}' \rangle e^{-\tau[V(\mathbf{R}')-E_T]/2} . \quad (3.33)$$

The approximate Green function for small τ is therefore given by (Reynolds *et al.*, 1982)

$$G(\mathbf{R} \leftarrow \mathbf{R}', \tau) \approx (2\pi\tau)^{-3N/2} \exp \left[-\frac{(\mathbf{R} - \mathbf{R}')^2}{2\tau} \right] \\ \times \exp [-\tau(V(\mathbf{R}) + V(\mathbf{R}') - 2E_T)/2] , \quad (3.34)$$

where the error is proportional to τ^3 .

The factor

$$P = \exp [-\tau(V(\mathbf{R}) + V(\mathbf{R}') - 2E_T)/2] \quad (3.35)$$

acts as a time-dependent renormalization (reweighting) of the diffusion Green function. This change of normalization can be incorporated into the process of walker evolution in various ways. One possibility is to assign each walker a weight and accumulate the product of weights during the propagation. However, this option

does not give the most efficient sampling since the weights of the walkers rapidly become very different and, in the long time limit, one dominates exponentially over the rest. A better alternative is to use the *branching* or birth/death algorithm, in which P determines the number of walkers that survive to the next step (Reynolds *et al.*, 1982). In its simplest form the procedure is as follows:

1. If $P < 1$ the walker continues its evolution with probability P .
2. If $P \geq 1$ the walker continues; in addition, at the same position, a new walker is created with probability $P - 1$.

Both possibilities can be conveniently coded as a single command,

$$M_{\text{new}} = \text{INT}(P + \eta) , \quad (3.36)$$

where M_{new} is the number of walkers evolving to the next step at a given position, INT denotes the integer part of a real number, and η is a random number drawn from a uniform distribution on the interval $[0, 1]$. From the expression for P it is clear that in regions of high potential energy the walkers disappear while in regions of low potential energy they proliferate. The branching algorithm therefore transforms the weight accumulation in the low energy regions into an increase in the density of walkers there.

The energy offset E_T , which determines the overall asymptotic renormalization [see Eq. (3.27)], is used to control the total population of walkers. During the propagation E_T is occasionally adjusted so that the overall number of walkers fluctuates around a desired mean value. Typical mean values of the number of walkers used in simulations are between 10^2 and 10^3 . Figure 2 illustrates this DMC algorithm for the case of a single particle moving in a one-dimensional potential well.

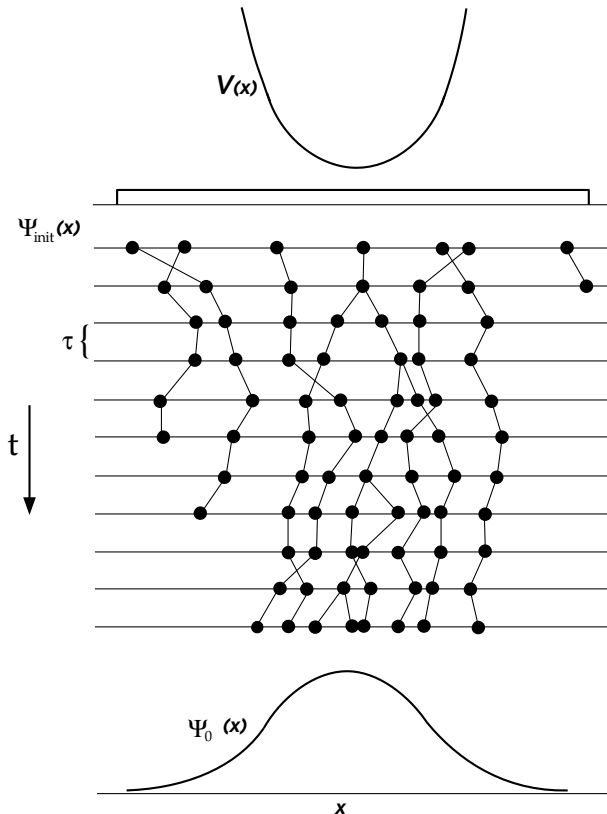


FIG. 2. Illustration of the walker evolution in the DMC method. The example shows a one-dimensional problem in which a single particle is confined by a potential well $V(x)$. The initial walker distribution samples a uniform Ψ_{init} . As the imaginary-time propagation proceeds the distribution converges towards the ground state Ψ_0 . Note the occasional disappearance of walkers in the region of high potential energy and the proliferation around the potential minimum.

So far we have assumed that the wave function is positive everywhere. However, the fermion antisymmetry implies that many-fermion wave functions cannot be positive everywhere and must take both positive and negative values. Unfortunately, probabilistic methods such as DMC can handle only positive distributions. A straightforward generalization of the DMC algorithm, such as an assignment of sign variables to walkers, while formally correct, leads to the fermion sign problem and an exponentially decaying signal to noise ratio. More elaborate algorithms with walker signs have been successfully applied to small molecules, but practical applications are restricted to few-electron systems (Diedrich and Anderson, 1992). Nevertheless, for small enough systems, very impressive results have been obtained. For example, the energy surface for the reaction $\text{H}_2 + \text{H}$ has been evaluated at a few thousand points with an astonishing accuracy of 0.01 kcal/mol (Wu, Kuppermann, and Anderson, 1999), improving existing results by an order of magnitude. The investigation of approximation-free algorithms for fermions is a very active area of current research but

deeper analysis is out of the scope of this review; we refer the reader to the literature (Hammond, Lester, and Reynolds, 1994; Kalos and Pederiva, 1999).

2. The fixed-node approximation

Fixed-node DMC (Anderson, 1975, 1976; Reynolds *et al.*, 1982; Moskowitz *et al.*, 1982) is an alternative method for dealing with the fermion antisymmetry. Although not exact, it gives ground-state energies that satisfy a variational principle and are usually very accurate. Furthermore, unlike the exact methods described above, the fixed-node algorithm is stable in large systems. The fixed-node approximation is therefore used in almost all current large-scale applications of DMC. The version described here assumes that the ground-state wave function is real and hence only works in systems with time-reversal symmetry (i.e., a real Hamiltonian). There is, however, a successful generalization known as the fixed-phase approximation (Ortiz, Ceperley, and Martin, 1993) for use in systems without time-reversal symmetry. This is particularly useful for studying interacting electrons in an applied magnetic field or states with non-zero angular momentum.

The basic idea behind fixed-node DMC is very simple. A trial many-electron wave function is chosen and used to define a trial many-electron nodal surface. In a three-dimensional system containing N electrons, the many-electron trial state is a function of $3N$ variables (the x , y , and z coordinates of each electron) and the trial nodal surface is the $3N - 1$ dimensional surface on which that function is zero and across which it changes sign. The fixed-node DMC algorithm then produces the lowest energy many-electron state with the given nodal surface. Fixed-node DMC may therefore be regarded as a variational method that gives exact results if the trial nodal surface is exact. It differs from VMC, however, in that the state is sampled stochastically without making any assumptions about its functional form between the nodes.

a. One-electron example

The fixed-node approximation is best introduced by means of a simple example. Figure 3 illustrates the standard DMC algorithm for the case of a particle in a box. Walkers initially distributed according to the starting state (the broken line) diffuse randomly until they cross one of the box walls, at which point they are removed from the simulation. This boundary condition follows from Eq. (3.35), which shows that the reweighting factor P is zero where the potential is infinite.

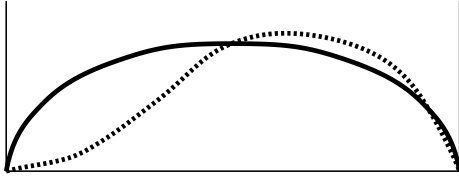


FIG. 3. A possible initial walker density distribution (broken line) and the final walker distribution (solid line) for a DMC simulation of a particle in a box.

The absorption of walkers at the walls guarantees that the average walker density tends to zero at the edges of the box, and hence that the eigenstate obtained satisfies the correct boundary conditions. The energy offset E_T is chosen such that P is slightly greater than unity within the box; this ensures that the diffusing walkers multiply steadily and counteracts the steady loss of walkers at the walls. Irrespective of the shape of the initial distribution, the walker density settles down to the final state shown, which is proportional to the ground-state wave function.

How might this algorithm be adapted to give the first excited state, which has odd parity and a node at the box center? One approach is to start with a trial function of odd parity, such as that shown in Fig. 4(a), and write it as the difference between the two positive distributions shown in Fig. 4(b):

$$\Phi(x, t = 0) = \Phi_+(x, t = 0) - \Phi_-(x, t = 0), \quad (3.37)$$

where

$$\Phi_+(x, t = 0) = \frac{1}{2} \left(|\Phi(x, t = 0)| + \Phi(x, t = 0) \right), \quad (3.38)$$

$$\Phi_-(x, t = 0) = \frac{1}{2} \left(|\Phi(x, t = 0)| - \Phi(x, t = 0) \right). \quad (3.39)$$

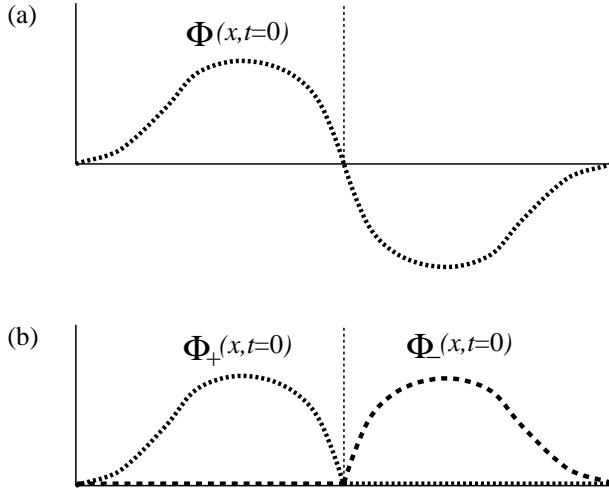


FIG. 4. Decomposition of an antisymmetric starting state into non-negative components. Panel (a) shows a possible starting state $\Phi(x, t = 0)$ for a DMC simulation of the lowest state with odd parity of a particle in a box; (b) shows the two non-negative distributions $\Phi_+(x, t = 0)$ and $\Phi_-(x, t = 0)$ into which $\Phi(x, t = 0)$ is decomposed in Eq. (3.37).

The Hamiltonian commutes with the parity operator and so preserves the parity of the starting state $\Phi(x, t = 0)$. Furthermore, since the imaginary-time Schrödinger equation is linear, the two positive distributions $\Phi_+(x, t)$ and $\Phi_-(x, t)$ may be obtained using two separate DMC simulations. The odd function $\Phi(x, t)$ may then be calculated by subtracting $\Phi_-(x, t)$ from $\Phi_+(x, t)$. This is the simplest of the exact solutions to the sign problem mentioned above, but it is clear that both Φ_+ and Φ_- tend to the even-parity ground state as $t \rightarrow \infty$. The odd-parity components of interest decay like $\exp(-\Delta E t)$ relative to the even components, where ΔE is the difference between the odd- and even-parity ground-state energies. In practice, therefore, the statistical noise in Φ_+ and Φ_- soon swamps the exponentially decaying odd-parity signal and the subtraction gives nonsense.

The fixed-node solution to this problem works as follows. Imagine introducing an extra absorbing barrier at the center of the box, dividing the interior into two separate simulation regions. The DMC walkers are initially scattered throughout both regions, but the simulations on the left and right then progress independently. In the large t limit, the walker densities in the two regions are proportional to the lowest energy eigenfunctions satisfying zero boundary conditions at both the box walls and the absorbing barrier. The odd-parity eigenfunction of interest satisfies these same Dirichlet boundary conditions, which determine the eigenfunctions within each region. The $t \rightarrow \infty$ walker densities within each region must therefore be proportional to the odd-parity eigenfunction in that region.

b. Many-electron version

Although the nodes of the many-electron ground state of a solid are very complicated, the fixed-node approximation works in the same way. If the nodes are known exactly, absorbing barriers may be placed everywhere on the nodal surface, dividing up the configuration space into a set of disjoint nodal pockets. Parallel DMC simulations are then carried out in all the nodal pockets, and the solution within each pocket tends to \pm the exact ground-state wave function in that pocket.

The wave function $\Psi(x_1, x_2, \dots, x_N)$ of N spinless electrons in one dimension must be zero whenever any two electrons coincide. These coincidence conditions define $(N - 1)$ -dimensional *coincidence planes* passing through the N -dimensional configuration space. Given that the nodal surface is also $(N - 1)$ -dimensional, it is possible that the coincidence planes exhaust the nodal surface and hence determine it exactly. Ceperley (1991) has shown that this is indeed the case for the ground state of a one-dimensional system. In d -dimensional systems, however, the coincidence planes are $dN - d$ dimensional whereas

the nodal surface is $dN - 1$ dimensional. For $d > 1$ the co-incidence planes are therefore no more than a framework through which the nodal surface must pass and cannot determine it completely.

The fixed-node approximation (Anderson, 1975, 1976) circumvents this difficulty by using an approximate nodal surface, which is normally obtained from a variational trial wave function. We show below that fixed-node DMC energies are variational, i.e., the fixed-node DMC energy is always greater than or equal to the exact ground-state energy, with the equality holding only when the trial nodal surface is exact. Since the DMC energy is normally a smooth function of the trial nodal surface, the errors in the energy are normally second order in the errors in the trial nodal surface. In practice, the errors in fixed-node DMC energies are typically about 5% of the correlation energy. Lest you imagine that this high accuracy arises because nodal surfaces are in some sense simple or easy to guess, look at Fig. 5 (which is taken from Ceperley, 1991). The little that is understood about real nodal surfaces is discussed in the papers by Caffarel and Claverie (1988), Ceperley (1991), Glauser *et al.* (1992), and Foulkes, Hood, and Needs (1999).

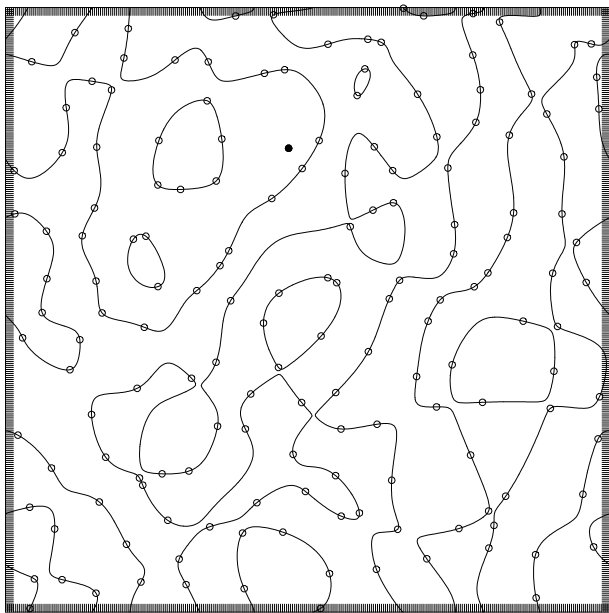


FIG. 5. A two-dimensional slice through the 321-dimensional nodal surface of a two-dimensional electron gas containing 161 spin-up electrons subject to periodic boundary conditions (Ceperley, 1991). The slice was defined by freezing 160 of the electrons at the random positions shown by the open circles and leaving the one remaining electron, represented by the filled circle, free to move around.

The implementation of the fixed-node DMC algorithm is straightforward. One scatters DMC walkers throughout the configuration space and moves them in the usual way. The only new ingredient is that after every DMC

move the sign of the trial wave function is checked and the walker deleted if it has crossed the trial nodal surface. (N.B. In the importance-sampled algorithm discussed in Sec. III.D.3 this deletion step is replaced by a rejection step.) Within each pocket, the fixed-node DMC algorithm projects out the lowest-energy nodeless wave function satisfying zero boundary conditions on the enclosing nodal surface.

The only communication between the simulations in different pockets is via the energy offset E_T , which is gradually adjusted to keep the total walker population roughly constant. If some pockets are more favorable than others (that is, if they have a greater average value of the reweighting factor P), the walker population becomes more and more concentrated in these pockets. The calculated energy therefore tends to the eigenvalue of the lowest-energy nodeless wave function within the most favorable pockets. This selection of favorable pockets does indeed happen in some excited-state DMC simulations, but Ceperley (1991) proved that the nodal pockets of the ground state of any Hamiltonian with a reasonable local potential are all equivalent by symmetry. As long as the trial nodal surface is that of the ground state of some local Hamiltonian (perhaps the LDA one), this *tiling theorem* ensures that all nodal pockets are equally favorable. A derivation of the tiling theorem is given below.

c. The fixed-node variational principle

Consider the nodal pockets of an antisymmetric trial function $\Psi_T(\mathbf{R})$. We can group these pockets into classes equivalent by permutation symmetry as follows. Pick a nodal pocket at random and color it blue. Now pick a point \mathbf{R} within the blue pocket and apply a permutation \mathcal{P} to it. This maps \mathbf{R} to $\mathcal{P}\mathbf{R}$, which may or may not be in the same nodal pocket as \mathbf{R} . If the new point $\mathcal{P}\mathbf{R}$ is in a new nodal pocket, color that pocket blue as well. Repeat this process for every permutation \mathcal{P} until all nodal pockets equivalent to the first one by permutation symmetry have been found and colored blue.

There are now two possibilities: either the blue pockets fill up the configuration space, in which case all nodal pockets are equivalent by permutation symmetry, or they do not. If they do not, pick one of the remaining pockets at random, color it red, and repeat the same procedure, applying permutation operators until all equivalent pockets have been found and colored red. If there are still pockets left, pick new colors and repeat the process until the configuration space is filled by pockets of different colors. The nodal pockets are thus divided into equivalence classes, with all the members of a class related by permutation symmetry.

If $\Psi_T(\mathbf{R})$ is used to define the nodal surface for a fixed-node DMC simulation, the DMC walkers will eventually

concentrate in the nodal pockets of the lowest-energy class. Within any such pocket v^α , the walker distribution will tend to the *pocket ground state* $\Psi_0^\alpha(\mathbf{R})$, which is the lowest-energy real normalized wave function that is zero outside v^α and satisfies the fixed-node boundary conditions on the surface of v^α . This function satisfies the equation

$$\hat{H}\Psi_0^\alpha(\mathbf{R}) = E_0^\alpha\Psi_0^\alpha(\mathbf{R}) + \delta^\alpha. \quad (3.40)$$

Within v^α the pocket ground state is a nodeless eigenfunction of the Schrödinger equation with eigenvalue E_0^α ; outside v^α it is zero. Since $\Psi_0^\alpha(\mathbf{R})$ normally approaches the surface of v^α with non-zero slope, its gradient changes discontinuously as the surface is crossed; the action of the kinetic energy operator on these gradient discontinuities produces the delta functions denoted δ^α . The DMC energy for pocket v^α (or any symmetry-equivalent pocket) is equal to the pocket eigenvalue E_0^α .

Moskowitz *et al.* (1982) and Reynolds *et al.* (1982) proved that E_0^α is greater than or equal to the exact ground-state energy E_0 . Here we follow Reynolds' proof, which starts from a single pocket ground state $\Psi_0^\alpha(\mathbf{R})$ and uses the permutations \mathcal{P} to construct a real antisymmetric wave function,

$$\tilde{\Psi}_0^\alpha(\mathbf{R}) = \frac{1}{N_{\mathcal{P}}} \sum_{\mathcal{P}} (-1)^{\zeta_{\mathcal{P}}} \Psi_0^\alpha(\mathcal{P}\mathbf{R}) \equiv \hat{A}\Psi_0^\alpha(\mathbf{R}), \quad (3.41)$$

where $N_{\mathcal{P}}$ is the total number of permutations and $\zeta_{\mathcal{P}}$ is the parity of permutation \mathcal{P} . The pocket ground state $\Psi_0^\alpha(\mathbf{R})$ has a non-zero overlap with the antisymmetric trial function $\Psi_T(\mathbf{R})$ and must therefore have a non-zero antisymmetric component. The projection operator \hat{A} picks out this component and so $\tilde{\Psi}_0^\alpha(\mathbf{R})$ cannot be zero. Since $\Psi_0^\alpha(\mathbf{R})$ is zero everywhere on the nodal surface of $\Psi_T(\mathbf{R})$, and since that nodal surface is invariant under all permutations \mathcal{P} , the function $\tilde{\Psi}_0^\alpha(\mathbf{R})$ is also zero everywhere on the nodal surface of $\Psi_T(\mathbf{R})$.

The real antisymmetric function $\tilde{\Psi}_0^\alpha(\mathbf{R})$ is now substituted into the standard quantum mechanical variational principle to give,

$$\begin{aligned} E_0 &\leq \frac{\int \tilde{\Psi}_0^\alpha \hat{H} \tilde{\Psi}_0^\alpha d\mathbf{R}}{\int \tilde{\Psi}_0^\alpha \tilde{\Psi}_0^\alpha d\mathbf{R}} = \frac{\int \tilde{\Psi}_0^\alpha \hat{H} \hat{A} \Psi_0^\alpha d\mathbf{R}}{\int \tilde{\Psi}_0^\alpha \hat{A} \Psi_0^\alpha d\mathbf{R}} \\ &= \frac{\int \tilde{\Psi}_0^\alpha \hat{H} \Psi_0^\alpha d\mathbf{R}}{\int \tilde{\Psi}_0^\alpha \Psi_0^\alpha d\mathbf{R}} = E_0^\alpha, \end{aligned} \quad (3.42)$$

where we have used the fact that \hat{A} commutes with \hat{H} , that it is self-adjoint, and that it is idempotent (so $\hat{A}\tilde{\Psi}_0^\alpha(\mathbf{R}) = \tilde{\Psi}_0^\alpha(\mathbf{R})$). The delta function terms appearing in $\hat{H}\Psi_0^\alpha$ do not contribute to the energy expectation value because they occur on the fixed nodal surface where $\tilde{\Psi}_0^\alpha(\mathbf{R}) = 0$. If the nodal surface of $\Psi_T(\mathbf{R})$ is the same as the nodal surface of the exact ground state the equality holds and the fixed-node DMC energy E_0^α is equal to

E_0 ; but if $\Psi_T(\mathbf{R})$ does not have the correct nodal surface then $E_0^\alpha > E_0$. This is the fixed-node variational principle. Since the calculated ground-state energy E_0^α is minimized and equal to the exact ground-state energy when the trial nodal surface is exact, the error in the calculated energy is in general second order in the error in the nodal surface.

d. The tiling theorem

Given any many-electron Hamiltonian with a reasonable local potential, the tiling theorem (Ceperley, 1991) states that all the ground-state nodal pockets belong to the same class. (The definition of a class was discussed in Sec. III.D.2.c above.) This is a many-fermion generalization of the statement that the ground state of a one-particle Hamiltonian with a local potential has no nodes. The tiling theorem and the fixed-node variational principle are very closely related, as the following derivation demonstrates.

Assume, for the sake of argument, that $\Psi_0(\mathbf{R})$ is the ground state of a Hamiltonian \hat{H} with a reasonable local potential, but that the nodal pockets of $\Psi_0(\mathbf{R})$ fall into two or more different classes. Color the nodal pockets belonging to one of these classes blue and consider the function $\tilde{\Psi}_0(\mathbf{R})$, defined to equal $\Psi_0(\mathbf{R})$ in the blue pockets and zero elsewhere. This function is properly antisymmetric but has gradient discontinuities on the nodal surface separating the blue pockets from the rest of configuration space. The action of the kinetic energy operator turns these gradient discontinuities into delta functions, but since they occur where $\tilde{\Psi}_0(\mathbf{R}) = 0$ they do not affect the energy expectation value,

$$E_0 = \frac{\langle \tilde{\Psi}_0 | \hat{H} | \tilde{\Psi}_0 \rangle}{\langle \tilde{\Psi}_0 | \tilde{\Psi}_0 \rangle}. \quad (3.43)$$

We have therefore reached a contradiction: only the exact ground state (or a linear combination of degenerate ground states) could give an energy expectation value equal to E_0 ; but $\tilde{\Psi}_0(\mathbf{R})$ cannot be an exact eigenstate unless the potential diverges almost everywhere on the nodal surface separating the blue pockets from the rest of configuration space. If there is a region of this surface on which the potential is finite, any exact eigenstate within the blue region is bound to leak out into the surrounding regions. This implies that the exact ground state cannot have more than one class of nodal pocket.

3. Importance sampling

The simple version of the DMC algorithm outlined in the previous section is usually spectacularly inefficient, mainly because the renormalization factor P from

Eq. (3.35) may fluctuate wildly from step to step. In some circumstances, such as when two charged particles coincide, P can even diverge, making the renormalization process ill-defined.

One can overcome these difficulties by carrying out an *importance-sampling* transformation using a “trial” or “guiding” wave function $\Psi_T(\mathbf{R})$ (Grimm and Storer, 1971; Ceperley and Kalos, 1979; Reynolds *et al.*, 1982). Let us multiply Eq. (3.21) by $\Psi_T(\mathbf{R})$ and introduce a new function $f(\mathbf{R}, t) = \Phi(\mathbf{R}, t)\Psi_T(\mathbf{R})$. After rearranging terms we obtain:

$$-\partial_t f(\mathbf{R}, t) = -\frac{1}{2}\nabla^2 f(\mathbf{R}, t) + \nabla \cdot [\mathbf{v}_D(\mathbf{R})f(\mathbf{R}, t)] + (E_L(\mathbf{R}) - E_T)f(\mathbf{R}, t), \quad (3.44)$$

where $\nabla = (\nabla_1, \nabla_2, \dots, \nabla_N)$ is the $3N$ -dimensional gradient operator, $\mathbf{v}_D(\mathbf{R})$ is the $3N$ -dimensional *drift velocity* defined by

$$\mathbf{v}_D(\mathbf{R}) = \nabla \ln |\Psi_T(\mathbf{R})| = \Psi_T(\mathbf{R})^{-1} \nabla \Psi_T(\mathbf{R}), \quad (3.45)$$

and

$$E_L(\mathbf{R}) = \Psi_T(\mathbf{R})^{-1} \hat{H} \Psi_T(\mathbf{R}) \quad (3.46)$$

is the *local energy* as used in VMC simulations.

The corresponding integral equation is modified accordingly,

$$f(\mathbf{R}, t + \tau) = \int \tilde{G}(\mathbf{R} \leftarrow \mathbf{R}', \tau) f(\mathbf{R}', t) d\mathbf{R}', \quad (3.47)$$

where the modified Green function $\tilde{G}(\mathbf{R} \leftarrow \mathbf{R}', \tau)$ is by definition equal to $\Psi_T(\mathbf{R})G(\mathbf{R} \leftarrow \mathbf{R}', \tau)\Psi_T(\mathbf{R}')^{-1}$. The short-time approximation to $\tilde{G}(\mathbf{R} \leftarrow \mathbf{R}', \tau)$ may be derived using techniques analogous to those employed in the derivation of Eq. (3.34). The result is:

$$\tilde{G}(\mathbf{R} \leftarrow \mathbf{R}', \tau) \approx G_d(\mathbf{R} \leftarrow \mathbf{R}', \tau)G_b(\mathbf{R} \leftarrow \mathbf{R}', \tau), \quad (3.48)$$

where

$$G_d(\mathbf{R} \leftarrow \mathbf{R}', \tau) = (2\pi\tau)^{-3N/2} \exp \left[-\frac{[\mathbf{R} - \mathbf{R}' - \tau\mathbf{v}_D(\mathbf{R}')]^2}{2\tau} \right] \quad (3.49)$$

and

$$G_b(\mathbf{R} \leftarrow \mathbf{R}', \tau) = \exp [-\tau(E_L(\mathbf{R}) + E_L(\mathbf{R}') - 2E_T)/2]. \quad (3.50)$$

Importance sampling has several consequences. First, the density of walkers is enhanced in the regions where Ψ_T is large and vice versa. This is because the drift velocity $\mathbf{v}_D(\mathbf{R})$ carries the walkers along in the direction of increasing $|\Psi_T|$. Second, the exponent in the reweighting term [see Eq. (3.35)] now contains the local energy

$E_L = \Psi_T^{-1} \hat{H} \Psi_T$ instead of the potential energy. This is crucial because for a good trial function the local energy is close to the ground-state energy eigenvalue and roughly constant, so the population fluctuations are much diminished. The overall improvement in efficiency due to importance sampling can be several orders of magnitude. Without this improvement, DMC simulations involving hundreds or, in recent attempts, thousands of electrons would not be possible.

The importance-sampling transformation is also extremely helpful in satisfying the fixed-node constraint. Whenever a walker approaches the nodal surface of Ψ_T , the drift velocity grows and carries it away. In fact, the drift near the nodal surface is proportional to $1/x$, where x is the distance in the direction normal to that surface. Furthermore, unless the trial function happens to be an exact eigenfunction, the local energy also diverges like $1/x$. It is possible to show that the divergence of the drift reduces the probability of crossing the node to zero, and hence, since we also neglect the effects of the infinite local energy on the nodal surface itself, that the importance-sampling transformation automatically enforces the fixed-node constraint.

In real simulations, however, the Green function of Eq. (3.48) is used. This is not exact and so walkers on rare occasions attempt to cross the nodal surface. The problem can be understood by realizing that the drift term in Eq. (3.49) is proportional to τ , while the standard deviation of the Gaussian, which arises from the diffusion of the walker, is proportional to $\sqrt{\tau}$. In order to ensure that the short-time approximation is accurate most simulations use values of τ much smaller than 1 a.u., and so the diffusion can occasionally overcome the drift and the walker can cross the node. This node crossing reflects a failure of the approximate Green function to describe the region close to the node, and of course becomes less likely as the time step τ tends to zero. In order to fulfill the fixed-node constraint one can either eliminate any walker that crosses a node (in other words, apply absorbing boundary conditions as discussed in Sec. III.D.2), or else reject the move and keep the walker at its original position. The choice of procedure has no effect in the limit of small enough time step (when the probability of crossing the nodal surface tends to zero), but it turns out that the rejection algorithm gives smaller time step errors.

The Green function of Eq. (3.48) is usually a reasonable approximation in regions where the trial wave function is smooth and non-zero. Close to the fixed nodal surface or to a singularity in the potential, however, this is not the case, and quantities such as the drift velocity and local energy may even diverge. The approximate Green function of Eq. (3.48) is then poor and the resulting bias can be significant (Umrigar, Nightingale, and Runge, 1993).

The simplest remedy is to take smaller time steps, al-

though this makes the calculation rather inefficient. A better remedy is to improve the Green function. A clever idea due to Ceperley, Kalos, and Lebowitz (1981) and Reynolds *et al.* (1982) is to incorporate a rejection step into the propagation governed by $G_d(\mathbf{R} \leftarrow \mathbf{R}', \tau)$. The rejection step is designed to impose a detailed balance condition, which is motivated by two arguments. First, the *exact* importance-sampled Green function fulfills the detailed balance condition

$$\tilde{G}(\mathbf{R} \leftarrow \mathbf{R}', \tau) \Psi_T(\mathbf{R}')^2 = \tilde{G}(\mathbf{R}' \leftarrow \mathbf{R}, \tau) \Psi_T(\mathbf{R})^2, \quad (3.51)$$

as can be verified from the definition $\tilde{G}(\mathbf{R} \leftarrow \mathbf{R}', \tau) = \Psi_T(\mathbf{R}) G(\mathbf{R} \leftarrow \mathbf{R}', \tau) \Psi_T(\mathbf{R}')^{-1}$ and the observation that the exact $G(\mathbf{R} \leftarrow \mathbf{R}', \tau)$ is symmetric on interchange of \mathbf{R} and \mathbf{R}' . Second, if the trial function is equal to the exact ground state, an attempt to impose this detailed balance condition by accepting trial moves from \mathbf{R}' to \mathbf{R} with probability

$$\begin{aligned} p_{\text{accept}}(\mathbf{R} \leftarrow \mathbf{R}') &= \min \left[1, \frac{G_d(\mathbf{R}' \leftarrow \mathbf{R}, \tau) G_b(\mathbf{R}' \leftarrow \mathbf{R}, \tau) \Psi_T(\mathbf{R})^2}{G_d(\mathbf{R} \leftarrow \mathbf{R}', \tau) G_b(\mathbf{R} \leftarrow \mathbf{R}', \tau) \Psi_T(\mathbf{R}')^2} \right] \\ &= \min \left[1, \frac{G_d(\mathbf{R}' \leftarrow \mathbf{R}, \tau) \Psi_T(\mathbf{R})^2}{G_d(\mathbf{R} \leftarrow \mathbf{R}', \tau) \Psi_T(\mathbf{R}')^2} \right] \end{aligned} \quad (3.52)$$

guarantees the correct sampling *regardless* of the size of the time step, since the DMC algorithm reduces to the VMC algorithm with trial moves sampled from G_d . It has been repeatedly demonstrated that the inclusion of the acceptance/rejection step from Eq. (3.52) significantly decreases the time-step bias resulting from the approximate Green function. Obviously, the acceptance ratio depends on the size of the time step, and tends to unity as the time step tends to zero and the approximate Green function more nearly satisfies the detailed balance conditions of Eq. (3.51). If the time step is large, however, and too many moves are rejected, the introduction of the rejection step will bias the sampling of the unknown distribution $f(\mathbf{R}, t)$. The rule of thumb is to choose the time step such that the acceptance ratio is $\geq 99\%$. The better the approximate Green function, the larger the time step that can be used without a detectable bias (Umrigar, Nightingale, and Runge, 1993).

Several other improvements to the Green function are also normally employed. The use of an *effective* time step in G_b as suggested by Reynolds *et al.* (1982) and by Umrigar, Nightingale, and Runge (1993) further reduces the bias caused by preventing node crossing. As mentioned above, the drift velocity diverges on the nodal surface and so a walker close to a node can make an excessively large move in the configuration space. One can eliminate this problem, and thus improve the approximate Green function, by imposing a cut-off on the magnitude of the drift velocity. Similar cut-offs may also be imposed on the values of the local energy used to compute G_b . Several

different cut-off schemes have been proposed (see, e.g., Garmer and Anderson, 1988; DePasquale, Rothstein, and Vrbik, 1988) but the smooth forms suggested by Umrigar, Nightingale, and Runge (1993) have been found to work well.

The result of the stochastic process described above is a set of walker positions representing the distribution $f(\mathbf{R}, t) = \Phi(\mathbf{R}, t) \Psi_T(\mathbf{R})$. Given these positions, the expectation value of the energy can be evaluated in two ways. One possibility is to measure the energy through the offset E_T that keeps the population constant on average. This is the so-called generational or population-growth estimator. In practice another possibility, known as the mixed estimator, is normally used. This is given by

$$\begin{aligned} E_D &= \lim_{\tau \rightarrow \infty} \frac{\langle e^{-\tau \hat{H}/2} \Psi_T | \hat{H} | e^{-\tau \hat{H}/2} \Psi_T \rangle}{\langle e^{-\tau \hat{H}/2} \Psi_T | e^{-\tau \hat{H}/2} \Psi_T \rangle} \\ &= \lim_{\tau \rightarrow \infty} \frac{\langle e^{-\tau \hat{H}} \Psi_T | \hat{H} | \Psi_T \rangle}{\langle e^{-\tau \hat{H}} \Psi_T | \Psi_T \rangle} = \frac{\langle \Psi_0 | \hat{H} | \Psi_T \rangle}{\langle \Psi_0 | \Psi_T \rangle} \\ &= \lim_{\tau \rightarrow \infty} \frac{\int f(\mathbf{R}, \tau) E_L(\mathbf{R}) d\mathbf{R}}{\int f(\mathbf{R}, \tau) d\mathbf{R}} \\ &\approx \frac{1}{M} \sum_m E_L(\mathbf{R}_m), \end{aligned} \quad (3.53)$$

where $\{\mathbf{R}_m\}$ is the set of M samples of $f(\mathbf{R}, \infty)$ resulting from the DMC run. The accuracies and efficiencies of these two energy estimators were compared by Umrigar, Nightingale, and Runge (1993).

Expectations of quantities that do not commute with the Hamiltonian can be calculated using a combination of the mixed and variational estimators

$$\langle \Phi | \hat{S} | \Phi \rangle \approx 2 \langle \Phi | \hat{S} | \Psi_T \rangle - \langle \Psi_T | \hat{S} | \Psi_T \rangle + \mathcal{O}[(\Phi - \Psi_T)^2], \quad (3.54)$$

where \hat{S} is the operator corresponding to the physical quantity of interest. Such combinations of VMC and DMC estimators are often called extrapolated estimators. For non-negative quantities (e.g., the density) one can also use

$$\langle \Phi | \hat{S} | \Phi \rangle \approx \langle \Phi | \hat{S} | \Psi_T \rangle^2 \langle \Psi_T | \hat{S} | \Psi_T \rangle^{-1} + \mathcal{O}[(\Phi - \Psi_T)^2]. \quad (3.55)$$

The accuracy of an extrapolated estimator depends on the trial function as well as the fixed-node DMC wave function. This is a significant drawback, but alternative approaches such as forward walking, which is clearly explained by Hammond, Reynolds, and Lester (1994), can be used instead. Another possibility may be to use the reptation QMC method of Baroni and Moroni (1999), although this has not yet been tested in large systems. A systematic discussion of the calculation of matrix elements of local, semi-local, and non-local operators that

either do or do not commute with the evolution operator is given by Nightingale (1999).

The most basic version of the DMC algorithm with importance sampling (Reynolds *et al.*, 1992; Hammond, Lester, and Reynolds, 1994) can be briefly outlined as follows:

1. Generate a set of walkers drawn from some initial distribution. (In most cases the initial walkers are sampled from $|\Psi_T|^2$ using variational Monte Carlo.) Calculate the local energies of the walkers.
2. Evaluate the drift velocity \mathbf{v}_D of each walker.
3. Propagate each walker for a time step τ , moving it from its old position \mathbf{R}' to the new position

$$\mathbf{R} = \mathbf{R}' + \chi + \tau \mathbf{v}_D(\mathbf{R}'), \quad (3.56)$$

where χ is a $3N$ -dimensional vector of normally distributed numbers with variance τ and zero mean.

4. Check whether the walker crossed the nodal surface (by checking the sign of the trial wave function) and move it back to the original position if this is the case.
5. Accept the step with the probability given by Eq. (3.52).
6. For each walker, calculate the number of copies that will continue in the evolution using

$$M_{\text{new}} = \text{INT} \left\{ \eta + \exp[-\tau(E_L(\mathbf{R}) + E_L(\mathbf{R}') - 2E_T)/2] \right\}, \quad (3.57)$$

where η is a random number drawn from a uniform distribution on the interval $[0, 1]$.

7. Accumulate the quantities of interest. For example, the energy may be accumulated by averaging the values of E_L over the set of walkers.
8. After an initial equilibration stage, steps 2–7 are repeated until the error bars for averages of interest are sufficiently small. The value of E_T is occasionally adjusted to keep the average walker population roughly constant. A simple formula for adjusting E_T is

$$E_T \leftarrow E_T - C_E \ln(M_{\text{act}}/M_{\text{ave}}), \quad (3.58)$$

where C_E is a positive constant that controls how quickly the actual number of walkers, M_{act} , approaches the desired number M_{ave} . C_E should not be too large because this can bias the expectation values. However, it cannot be too small because the

number of walkers then fluctuates too far from the desired value. Usually, it is reasonable to choose a C_E that rebalances the number of walkers in 10 to 50 time steps. More sophisticated techniques for dealing with E_T are discussed by Umrigar, Runge, and Nightingale (1993).

IV. TRIAL WAVE FUNCTIONS

A. Introduction

The quality of the trial wave function controls the statistical efficiency and limits the final accuracy of any VMC or DMC simulation. The repeated evaluation of the trial wave function (and its gradient and Laplacian) is also the most demanding part of the computation. We therefore seek trial wave functions that are both accurate and easy to evaluate.

In quantum chemical methods it is common to express many-body wave functions as linear combinations of determinants. However, such expansions converge very slowly because of the difficulty in describing the cusps that occur whenever two electrons come into contact. QMC simulations of solids require a much more compact representation and normally use trial wave functions of the Slater-Jastrow type (Jastrow, 1955), consisting of a single Slater determinant multiplied by a totally symmetric non-negative Jastrow correlation factor that includes the cusps. The orbitals in the Slater determinant are usually obtained from accurate LDA or HF calculations, while the Jastrow factor is chosen to have some specific functional form and optimized as explained in Sec. VII. Linear combinations of Slater-Jastrow functions are sometimes required, but the single-determinant form is satisfactory for many purposes. When used in VMC simulations of tetrahedral semiconductors (Fahy, Wang, and Louie, 1988, 1990a; Li, Ceperley, and Martin, 1991; Rajagopal, Needs, Kenny *et al.*, 1994; Rajagopal *et al.*, 1995; Eckstein, Schattke *et al.*, 1996; Malatesta, Fahy, and Bachelet, 1997), single-determinant Slater-Jastrow trial functions generally reproduce the experimental cohesive energies to within about 0.1 eV per atom.

B. Slater-Jastrow wave functions

A single-determinant Slater-Jastrow wave function can be written as

$$\Psi(\mathbf{X}) = e^{J(\mathbf{X})} D(\mathbf{X}), \quad (4.1)$$

where $D(\mathbf{X})$ is a Slater determinant as shown in Eq. (2.3), $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$, and $\mathbf{x}_i = \{\mathbf{r}_i, \sigma_i\}$ denotes the space and spin coordinates of electron i . Most simulations retain only one- and two-body terms in the Jastrow factor,

$$J(\mathbf{X}) = \sum_{i=1}^N \chi(\mathbf{x}_i) - \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ (j \neq i)}}^N u(\mathbf{x}_i, \mathbf{x}_j). \quad (4.2)$$

The u terms describe the electron-electron correlations, while the χ function depends on the positions of the nuclei and describes the electron-nuclear correlation.

In QMC calculations one normally removes the spin variables from the Slater-Jastrow wave function and replaces the single Slater determinant of Eq. (4.1) by a product of determinants of up-spin and down-spin orbitals,

$$\Psi(\mathbf{R}) = e^{J(\mathbf{R})} D^\uparrow(\mathbf{r}_1, \dots, \mathbf{r}_{N_\uparrow}) D^\downarrow(\mathbf{r}_{N_\uparrow+1}, \dots, \mathbf{r}_N), \quad (4.3)$$

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 on exchange of electrons with opposite spins and so differs from the Slater-Jastrow wave function of Eq. (4.1), but in Sec. IV.E we show that it gives the same expectation value for any spin-independent operator. The use of spin-independent wave functions such as $\Psi(\mathbf{R})$ is computationally advantageous because a large determinant is replaced by two smaller ones and no sums over spin variables are required; it also facilitates the imposition of the cusp conditions discussed in Sec. IV.F. By convention, we shall always choose $\mathbf{r}_1, \dots, \mathbf{r}_{N_\uparrow}$ to be the coordinates of the spin-up electrons and $\mathbf{r}_{N_\uparrow+1}, \dots, \mathbf{r}_N$ to be the coordinates of the spin-down electrons.

The physics underlying Slater-Jastrow wave functions is quite straightforward. Figure 6 shows the spin-parallel and spin-antiparallel pair-correlation functions of a three-dimensional uniform electron gas calculated within the HF approximation. The antisymmetry of the wave function creates an exchange hole that keeps parallel-spin electrons apart, but there are no correlations between the positions of electrons with antiparallel spins. There is a significant probability of finding two antiparallel-spin electrons very close to each other and so the electron-electron Coulomb repulsion energy is high. The purpose of the two-body u term in the Jastrow factor is to reduce the magnitude of the many-electron wave function whenever two electrons approach one another; this reduces the probability of finding two electrons very close together and decreases the electron-electron interaction energy.

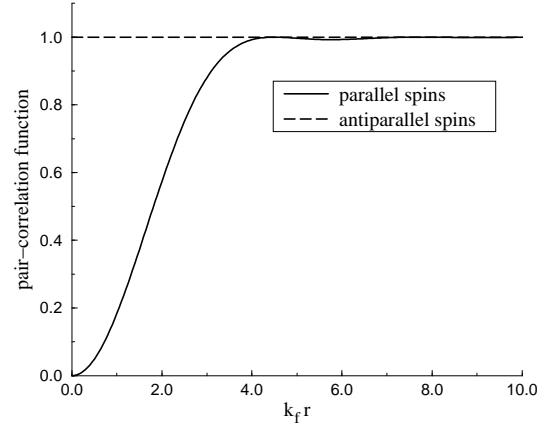


FIG. 6. Spin-parallel and spin-antiparallel pair-correlation functions calculated within the HF approximation for a uniform three-dimensional electron gas. The HF pair-correlation functions depend only on the product of the Fermi wavevector k_f and the inter-particle separation r .

In uniform systems this is all that is required. In non-uniform systems, however, the introduction of the repulsive u term changes the electron density significantly, pushing electrons away from regions of high charge density and into low density regions. This is an unavoidable consequence of introducing a repulsive two-body term, even though the intention was to change the pair-correlation function, not the density. In fact, the charge densities of Slater determinants of HF or LDA orbitals are usually quite accurate, and the change caused by the u term is unwelcome. The one-body χ term shown in Eq. (4.2) alters the charge density without greatly disturbing the pair-correlation function and hence allows this problem to be overcome.

The effect of the χ term is illustrated in Figs. 7 and 8, which show the spherically averaged charge densities of a full core carbon atom. Fig. 7 shows that the LDA charge density (solid line) is significantly altered by the introduction of a two-body u function (dotted line). The u function lowers the energy only slightly because the change in the charge density is unfavorable. Figure 8 shows that the charge density (dotted line) returns almost to the LDA form (solid line again) if a one-body χ function is added to the wave function and optimized variationally. It is the combination of optimized χ and u terms which lowers the calculated atomic energy greatly. Similar behavior is observed in simulations of solids, although with some quantitative differences when the atoms are represented using pseudopotentials. The introduction of the u term alone typically lowers the energy per atom by an eV but suppresses the charge density in the bonding regions and enhances it in the interstitial regions. The introduction of the χ term restores the charge density to

something close to the LDA form and lowers the energy by an additional fraction of an eV.

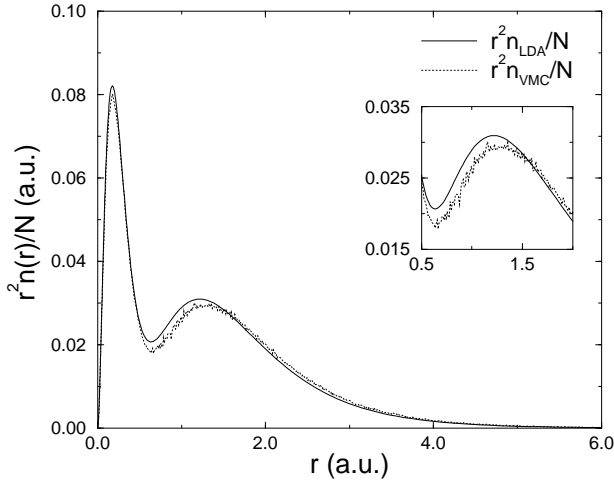


FIG. 7. Charge densities of a full core carbon atom from an LDA calculation (solid line) and a Slater-Jastrow wave function with a two-body correlation term but no one-body correlation term (dotted line). We thank Mark Stedman for providing this figure.

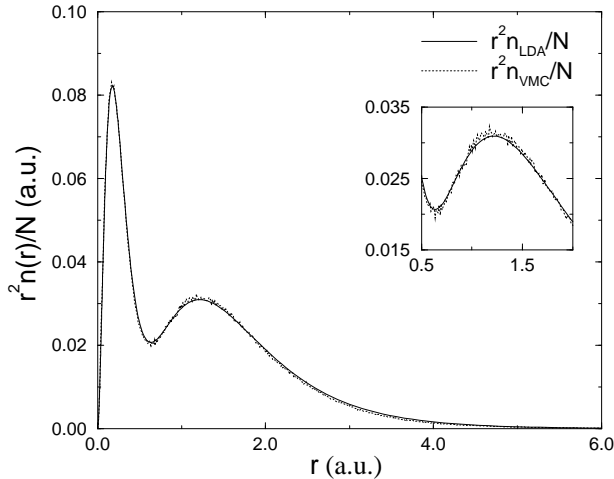


FIG. 8. Charge densities of a full core carbon atom from an LDA calculation (solid line) and a Slater-Jastrow wave function with both one- and two-body correlation terms (dotted line). We thank Mark Stedman for providing this figure.

The importance of the one-body χ term was recognized long ago in the context of Fermi-hypernetted-chain calculations (Feenberg, 1969) and in the correlation factor introduced by Boys and Handy (1969). The introduction of χ functions into QMC simulations of solids was due to Fahy *et al.* (1988, 1990a). Although the e^χ factors can be absorbed into the single-particle orbitals of the Slater determinant, they are so closely linked to the u function that it is normally convenient to keep them in the Jastrow factor. The relationship between the u and

χ functions is discussed in more detail in Sec. IV.D.

The great success of single determinant Slater-Jastrow wave functions is founded on the success of the HF approximation. For example, in the carbon atom HF theory retrieves 99.6% of the total energy. The other 0.4% (4.3 eV) is the correlation energy, which is important for an accurate description of chemical bonding but is small compared with the total energy. QMC calculations for solids almost always use pseudopotentials in which the large energy contribution from the core electrons is removed. In this case the correlation energy can amount to several percent of the total energy, but the HF approximation still accounts for the vast majority of the energy. For example, in the carbon pseudo-atom HF theory retrieves about 98.2% of the total energy. The error of 1.8% (2.7 eV) is a small fraction of the total, but is large on the scale of bonding energies. Even a relatively simple Jastrow factor can lower the total energy of a carbon pseudo-atom by about 2.4 eV, thereby retrieving 89% of the correlation energy and reducing the error in the total energy to 0.2%. The Jastrow factor therefore provides a small but very important correction to the wave function.

C. The Slater determinant

The orbitals for the Slater determinant are most commonly obtained from HF or LDA calculations. This procedure may be motivated by noting that a determinant of HF orbitals gives the lowest energy of all single-determinant wave functions (although it is far from obvious that HF or LDA orbitals remain a good choice in the presence of a Jastrow factor). Here we consider the few attempts that have been made to devise something better.

Direct numerical optimization of the single-particle orbitals in the Slater determinant has a long history in quantum chemistry, but has not often been attempted using QMC methods. To date, most of the work on stochastic orbital optimization has been for fairly small systems (see, e.g., Umrigar, Nightingale, and Runge, 1993), although Eckstein and co-workers successfully optimized the parameters of the very restricted Slater orbital basis sets used in their studies of solid Li (Eckstein and Schattke, 1995) and GaAs (Eckstein, Schattke et al., 1996). More recently, Fahy (1999) and Filippi and Fahy (2000) have developed a new method that makes it possible to optimize the more accurate single particle states used in most current QMC simulations of solids. Although LDA or HF orbitals are usually adequate for real solids, direct numerical optimization will probably become widespread within a few years.

A different tack was taken by Grossman and Mitas (1995), who used a determinant of the *natural orbitals* which diagonalize the one-electron density matrix. The

motivation is that the convergence of configuration interaction expansions is improved by using natural orbitals instead of HF orbitals. Grossman and Mitas (1995) obtained the natural orbitals for Si₂, Si₃, and Si₄ molecules via multiconfiguration HF calculations. They found that the use of natural orbitals lowered the QMC energy significantly, but a similar calculation for bulk silicon (Kent *et al.*, 1998) found little improvement. Further work revealed that for wave functions with little mixing of excited configurations the decrease in energy was rather small. However, for more complicated wave functions the improvements were significant and natural orbitals proved necessary to obtain systematically high accuracy in calculations of molecular reactions (Grossman and Mitas, 1997).

Another approach is based on the idea of backflow correlations, derived from a current conservation argument by Feynman and Cohen (1956) to provide a picture of the excitations in liquid ⁴He. Work on liquid ³He (Schmidt *et al.*, 1981; Panoff and Carlson, 1989), various atoms (Schmidt and Moskowitz, 1992), and the uniform electron gas (Kwon, Ceperley, and Martin, 1993, 1994, 1998) shows that backflow correlations are also helpful in fermionic systems. The essential ingredient of the backflow trial function is the replacement of the electron coordinates \mathbf{r}_i appearing in the Slater determinants of Eq. (4.3) by the corresponding *quasiparticle coordinates*

$$\bar{\mathbf{r}}_i = \mathbf{r}_i + \sum_{\substack{j=1 \\ (j \neq i)}}^N \eta(r_{ij})(\mathbf{r}_i - \mathbf{r}_j), \quad (4.4)$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. In uniform systems, this amounts to replacing the single particle orbitals $\exp(i\mathbf{k}_\ell \cdot \mathbf{r}_i)$ by $\exp(i\mathbf{k}_\ell \cdot \bar{\mathbf{r}}_i)$, which ensures that the correlation between electrons i and j depends on the momentum \mathbf{k}_ℓ of the state occupied by electron i . The optimal function $\eta(r_{ij})$ is usually determined variationally. Kwon, Ceperley, and Martin (1998) found that the introduction of backflow significantly lowered the VMC and DMC energies of the three-dimensional uniform electron gas at high densities (the VMC energy decreased by 0.07 eV per electron when $r_s = 1$ and by 0.02 eV per electron when $r_s = 5$). The same authors (Kwon, Ceperley, and Martin, 1993, 1998) also investigated the importance of three-electron correlation terms in the Jastrow factor, obtaining a small improvement in the energy at low densities. The use of backflow wave functions for inhomogeneous systems is much less explored, but since backflow increases the computational cost of QMC calculations significantly its use will probably be confined to cases where extremely high accuracy is sought.

In some cases it is necessary to use multi-determinant wave functions to preserve important symmetries of the true wave functions. In other cases a single determinant may give the correct symmetry but a significantly more

accurate wave function can be obtained by using a linear combination of a few determinants. For example, in the Be atom the $2s$ and $2p$ orbitals are nearly degenerate and the many-electron wave function contains important contributions from the electronic configurations $1s^2 2s^2$ and $1s^2 2p^2$ (Harrison and Handy, 1985). The accurate variational trial function for Be proposed by Umrigar, Nightingale, and Runge (1993) is a linear combination of the corresponding Slater-Jastrow wave functions. Flad, Caffarel and Savin (1997) have carried out a thorough investigation of the advantages and disadvantages of using multi-determinant wave functions for various other atoms. In solids, however, the number of important configurations may grow very rapidly with system size and the use of multi-reference wave functions can be very demanding.

D. The Jastrow factor

What is an appropriate Jastrow factor for a uniform electron gas? The spatial homogeneity of the system ensures that no χ terms are needed and that u depends only on $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. The cusp conditions (Kato, 1957; Pack and Brown, 1966; and Sec. IV.F) determine the behaviour of the u function in the limit $r_{ij} \rightarrow 0$, while arguments based on the random phase approximation (RPA) of Bohm and Pines (1953) imply that it has a $1/r_{ij}$ tail when r_{ij} is large.

The following simple choice has the correct small and large r_{ij} limits and was used (along with a better form based more closely on the RPA) in some of the earliest work on the homogeneous electron gas (Ceperley, 1978; Ceperley and Alder, 1980):

$$u_{\sigma_i, \sigma_j}(\mathbf{r}_i, \mathbf{r}_j) = \frac{A_{\sigma_i, \sigma_j}}{r_{ij}} \left(1 - e^{-r_{ij}/F_{\sigma_i, \sigma_j}} \right). \quad (4.5)$$

The correlations between pairs of electrons depend on whether they have parallel or antiparallel spins, so the constants A_{σ_i, σ_j} and F_{σ_i, σ_j} are spin dependent. Assuming that the solid is not spin polarized, there are four parameters: $A_{\uparrow\uparrow} = A_{\downarrow\downarrow}$, $A_{\uparrow\downarrow} = A_{\downarrow\uparrow}$, $F_{\uparrow\uparrow} = F_{\downarrow\downarrow}$, and $F_{\uparrow\downarrow} = F_{\downarrow\uparrow}$. Two of the four may be eliminated by imposing the cusp conditions,

$$\left. \frac{du_{\sigma_i, \sigma_j}(r_{ij})}{dr_{ij}} \right|_{r_{ij}=0} = \begin{cases} -1/4 & \sigma_i = \sigma_j, \\ -1/2 & \sigma_i \neq \sigma_j. \end{cases} \quad (4.6)$$

Sec. IV.F shows that as long as the cusp conditions are obeyed the divergence in the Coulomb interaction energy occurring whenever two electrons coincide is exactly cancelled by a divergence in the kinetic energy arising from the cusp in the wave function. For the Jastrow factor of Eq. (4.5) the cusp conditions translate into

$$\frac{A_{\uparrow\uparrow}}{2F_{\uparrow\uparrow}^2} = \frac{1}{4} \quad \text{and} \quad \frac{A_{\uparrow\downarrow}}{2F_{\uparrow\downarrow}^2} = \frac{1}{2}. \quad (4.7)$$

The remaining two parameters, which we take to be $A_{\uparrow\uparrow}$ and $A_{\uparrow\downarrow}$, can be chosen to reproduce the long range behaviour implied by the RPA,

$$u_{\sigma_i, \sigma_j}(r_{ij}) \approx \frac{1}{\omega_p r_{ij}} \quad \text{as } r_{ij} \rightarrow \infty, \quad (4.8)$$

where $\omega_p = \sqrt{4\pi n}$ is the plasma frequency and n is the electron number density. This implies that

$$A_{\uparrow\uparrow} = A_{\uparrow\downarrow} = \frac{1}{\omega_p} = \frac{1}{\sqrt{4\pi n}}. \quad (4.9)$$

An alternative is to optimize the values of $A_{\uparrow\uparrow}$ and $A_{\uparrow\downarrow}$ variationally, but this does not give much improvement in practice. Most solid state QMC calculations these days use Eq. (4.5) or something similar as a starting point and add variational degrees of freedom to improve the accuracy. The optimal values of the variational parameters are determined using the techniques discussed in Sec. VII.

The cusp conditions imply that the Jastrow factor is spin dependent and this produces a small amount of *spin contamination*: a linear combination of Slater-Jastrow trial functions may not be an eigenfunction of \hat{S}^2 even though the corresponding sum of Slater determinants is. It is difficult to clean up this contamination without spoiling the accuracy of the trial function in other ways, but its effects are so small that they can almost always be ignored (Huang, Filippi, and Umrigar, 1998).

In QMC calculations we model extended systems using small simulation cells subject to periodic boundary conditions. This periodicity must be reflected in $u_{\sigma', \sigma}(\mathbf{r}', \mathbf{r})$, which should remain unchanged if either \mathbf{r}' or \mathbf{r} is translated by a simulation cell lattice vector. One way to enforce the correct translational symmetry on the u function of Eq. (4.5) is to sum it over all the periodic “images” of the electrons in an infinite lattice of identical copies of the simulation cell. The resulting sum does not converge because u has a slowly decaying $1/r_{ij}$ tail, but can be made meaningful by adopting the Ewald definition (Ewald, 1921; see Tosi, 1964, for a clear review).

Ortiz and Ballone (1994) and Williamson *et al.* (1996) have used u functions which tend to zero smoothly as r_{ij} approaches the radius r_{WS} of the largest sphere that fits inside the Wigner-Seitz simulation cell. This is computationally efficient because the need for sums over images is eliminated. However, since u is set to zero for $r_{ij} > r_{\text{WS}}$, it no longer has the $1/r_{ij}$ tail implied by the RPA. In practice, the consequences of truncating u appear to be small for reasonably large simulation cells.

A suitable form for the χ function in a periodic solid is a plane wave expansion,

$$\chi_{\sigma}(\mathbf{r}) = \sum_{\mathbf{G}_p} \chi_{\sigma, \mathbf{G}_p} e^{i\mathbf{G}_p \cdot \mathbf{r}}, \quad (4.10)$$

where \mathbf{G}_p is a reciprocal vector of the primitive crystal lattice (*not* the lattice of simulation cells, each of which may contain several primitive unit cells). This form has been used by Fahy and coworkers in studies of *sp* bonded materials (Fahy, Wang, and Louie, 1988, 1990a). In Fahy’s original work (Fahy, Wang, and Louie, 1988, 1990a) the χ function was designed to force the electron density towards the LDA result, which was assumed to be accurate. This *ad hoc* procedure lowers the variational energy by several eV per atom, but has been superseded by direct optimization of the $\chi_{\sigma, \mathbf{G}_p}$ coefficients (Williamson *et al.*, 1996; Malatesta, Fahy, and Bachelet, 1997). The number of independent variational parameters may be reduced by noting that \mathbf{G}_p vectors of the same length are often related by symmetry. When full use is made of such symmetry arguments, only six independent parameters are required to produce an accurate χ function for crystalline Ge (Williamson *et al.*, 1996).

In calculations for finite systems or for solids with many inequivalent atoms the plane-wave representation of χ is inefficient. To overcome this problem one can represent χ by a sum of local atom-centered functions,

$$\chi_{\sigma}(\mathbf{r}) = \sum_{\alpha} g_{\sigma, \alpha}(|\mathbf{r} - \mathbf{d}_{\alpha}|). \quad (4.11)$$

This form proved useful in the study of the formation energies of silicon self-interstitial defects discussed in Sec. V.I. Since the atoms around the defect are not all equivalent, it was found beneficial to allow the χ functions on inequivalent atoms to differ.

An important advance in understanding Jastrow factors was made by Malatesta, Fahy, and Bachelet (1997; see also Fahy, 1999). The discussion of Sec. IV.B suggests that the one- and two-body terms in the Jastrow factor are closely related, and we might expect that a linear relationship exists between the optimal forms of u and χ . Assuming that the u function depends only on the relative positions of the two electrons, $u_{\sigma_i, \sigma_j}(\mathbf{r}_i, \mathbf{r}_j) = u_{\sigma_i, \sigma_j}(\mathbf{r}_i - \mathbf{r}_j)$, the relationship derived by Malatesta, Fahy, and Bachelet (1997) takes the form,

$$\chi_{\sigma, \mathbf{G}_p} = \sum_{\sigma'} \langle \hat{n}_{\sigma', \mathbf{G}_p}^* \rangle u_{\sigma', \sigma, \mathbf{G}_p}, \quad (4.12)$$

where $\chi_{\sigma, \mathbf{G}_p}$ is a Fourier component of $\chi_{\sigma}(\mathbf{r})$ as defined in Eq. (4.10), $u_{\sigma', \sigma, \mathbf{G}_p}$ is the corresponding Fourier component of $u_{\sigma', \sigma}(\mathbf{r})$, and the operators

$$\hat{n}_{\uparrow, \mathbf{G}_p} = \sum_{i=1}^{N_{\uparrow}} e^{-i\mathbf{G}_p \cdot \mathbf{r}_i} \quad \text{and} \quad \hat{n}_{\downarrow, \mathbf{G}_p} = \sum_{i=N_{\uparrow}+1}^{N_{\uparrow}+N_{\downarrow}} e^{-i\mathbf{G}_p \cdot \mathbf{r}_i} \quad (4.13)$$

measure the Fourier components of the up- and down-spin electron number densities [multiplied by the simulation cell volume if the Fourier components are defined in analogy with Eq. (4.10)]. Note that unlike $\chi_{\sigma}(\mathbf{r})$, which

has the full periodicity of the crystal unit cell, the u function only has the periodicity of the simulation cell. The Fourier series for $u_{\sigma',\sigma}(\mathbf{r})$ therefore contains Fourier components for every vector \mathbf{G}_s in the simulation cell reciprocal lattice. The operator $\hat{n}_{\sigma,\mathbf{G}_s}$ is also defined for all simulation cell reciprocal lattice vectors \mathbf{G}_s , but its expectation value $\langle \hat{n}_{\sigma,\mathbf{G}_s} \rangle$ vanishes unless $\mathbf{G}_s \in \{\mathbf{G}_p\}$. The set $\{\mathbf{G}_p\}$ is a subset of $\{\mathbf{G}_s\}$.

Malatesta, Fahy, and Bachelet (1997) used Eq. (4.12) with the LDA charge density and the u function of Eq. (4.5) to construct a χ function for a VMC calculation of cubic boron nitride. The result was almost, but not quite, as good as a χ function obtained (at much greater cost) using the variance minimization method explained in Sec. VII. There is therefore little doubt that Eq. (4.12) provides a valuable insight into the physics of Jastrow functions as well as a practical way of obtaining useful approximations to the χ function.

The reasoning behind Eq. (4.12) was based on the Bohm-Pines (1953) RPA theory of the uniform electron gas, which leads to a spin-independent Jastrow factor. If, for the sake of convenience, we choose to add constant diagonal ($i = j$) terms to the Jastrow factor, we can express it in terms of Fourier components as follows,

$$-\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N u(\mathbf{r}_i - \mathbf{r}_j) = -\frac{1}{2} \sum_{\mathbf{G}_s} \hat{n}_{\mathbf{G}_s}^* u_{\mathbf{G}_s} \hat{n}_{\mathbf{G}_s}. \quad (4.14)$$

The insight of Malatesta, Fahy, and Bachelet (1997) was to see that in extending the RPA to inhomogeneous systems, the Fourier components of the electron number densities should be replaced by the deviations from their mean values, $\Delta \hat{n}_{\mathbf{G}_s} = \hat{n}_{\mathbf{G}_s} - \langle \hat{n}_{\mathbf{G}_s} \rangle$. The Jastrow factor then becomes (Malatesta, Fahy, and Bachelet, 1997)

$$-\frac{1}{2} \sum_{\mathbf{G}_s} \Delta \hat{n}_{\mathbf{G}_s}^* u_{\mathbf{G}_s} \Delta \hat{n}_{\mathbf{G}_s} = -\frac{1}{2} \sum_{\mathbf{G}_s} \hat{n}_{\mathbf{G}_s}^* u_{\mathbf{G}_s} \hat{n}_{\mathbf{G}_s} + \sum_{\mathbf{G}_p} \langle \hat{n}_{\mathbf{G}_p}^* \rangle u_{\mathbf{G}_p} \hat{n}_{\mathbf{G}_p} + \text{constant}, \quad (4.15)$$

where we have assumed that $u(\mathbf{r})$ is real and used our earlier observation that $\langle \hat{n}_{\mathbf{G}_s} \rangle$ is zero unless $\mathbf{G}_s \in \{\mathbf{G}_p\}$. The right hand side contains a two-body term of the form of Eq. (4.14) and a one-body term with Fourier components,

$$\chi_{\mathbf{G}_p} = \langle \hat{n}_{\mathbf{G}_p}^* \rangle u_{\mathbf{G}_p}. \quad (4.16)$$

Equation (4.12) is the natural generalization of Eq. (4.16) to the case when the Jastrow factor is spin dependent.

In real materials, $u_{\sigma_i,\sigma_j}(\mathbf{r}_i, \mathbf{r}_j)$ no longer depends only on r_{ij} but on \mathbf{r}_i and \mathbf{r}_j separately. In atomic and molecular calculations it has become common to take account of this inhomogeneity by including three-body electron-electron-nucleus correlation terms. The Jastrow factor is then of the form

$$J(\mathbf{R}) = \sum_{\alpha} \sum_{i,j} u_{\sigma_i,\sigma_j}(r_{i\alpha}, r_{j\alpha}, r_{ij}), \quad (4.17)$$

where $r_{i\alpha}$ is the separation of the i 'th electron from the α 'th nucleus. Schmidt and Moskowitz (1992) have used the correlation factor of Boys and Handy (1969), while Umrigar and coworkers (Umrigar, Wilson and Wilkins, 1988; Filippi and Umrigar, 1996; Huang, Filippi, and Umrigar, 1998) have used various Padé and polynomial forms. According to Huang, Umrigar, and Nightingale (1997), the inclusion of electron-electron-nucleus correlation terms lowers the VMC energy of the Be atom by 0.16 eV. The same authors also considered four-body electron-electron-electron-nucleus terms, which improved the VMC energy but not the DMC energy, even though the determinantal part of the trial function was reoptimized along with the Jastrow factor. Schmidt and Moskowitz (1992) have shown that the electron-electron-nucleus terms can describe some of the effects of backflow corrections. The inhomogeneity of the Jastrow factor may be less important in pseudopotential calculations, but tests for 3d transition metal atoms (Mitas, 1993) indicate that it can still be significant.

E. Spin

At first sight it may appear that the spin dependence of trial functions such as $\Psi(\mathbf{X})$ from Eq. (4.1) complicates QMC algorithms considerably. This is not the case, as we now explain. Suppose we wish to use $\Psi(\mathbf{X})$ to calculate the expectation value of a spin-independent operator $\hat{O}(\mathbf{R})$, where $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$:

$$\langle \hat{O} \rangle = \frac{\sum_{\sigma} \int \Psi^*(\mathbf{X}) \hat{O}(\mathbf{R}) \Psi(\mathbf{X}) d\mathbf{R}}{\sum_{\sigma} \int \Psi^*(\mathbf{X}) \Psi(\mathbf{X}) d\mathbf{R}}. \quad (4.18)$$

For each spin configuration $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_N)$, we can replace the totally antisymmetric wave function $\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ by a version with permuted arguments $\Psi(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_N})$, choosing the permutation such that the first N_{\uparrow} arguments are spin up and the remaining $N_{\downarrow} = N - N_{\uparrow}$ are spin down:

$$\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N) \rightarrow \Psi(\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_N}) = \Psi(\{\mathbf{r}_{i_1}, \uparrow\}, \dots, \{\mathbf{r}_{i_{N_{\uparrow}}}, \uparrow\}, \{\mathbf{r}_{i_{N_{\uparrow}+1}}, \downarrow\}, \dots, \{\mathbf{r}_{i_N}, \downarrow\}). \quad (4.19)$$

Since \mathbf{R} is a dummy variable of integration, we can now relabel $(\mathbf{r}_{i_1}, \mathbf{r}_{i_2}, \dots, \mathbf{r}_{i_N})$ as $(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$,

$$\Psi(\{\mathbf{r}_{i_1}, \uparrow\}, \dots, \{\mathbf{r}_{i_{N_{\uparrow}}}, \uparrow\}, \{\mathbf{r}_{i_{N_{\uparrow}+1}}, \downarrow\}, \dots, \{\mathbf{r}_{i_N}, \downarrow\}) \rightarrow \Psi(\{\mathbf{r}_1, \uparrow\}, \dots, \{\mathbf{r}_{N_{\uparrow}}, \uparrow\}, \{\mathbf{r}_{N_{\uparrow}+1}, \downarrow\}, \dots, \{\mathbf{r}_N, \downarrow\}). \quad (4.20)$$

The operator \hat{O} is not affected by the relabeling because it is symmetric with respect to exchange of electrons and so each spin configuration gives an identical contribution

to the expectation value. We can therefore remove the sums over spin configurations to get

$$\langle \hat{O} \rangle = \frac{\int \Psi^*(\mathbf{R}) \hat{O}(\mathbf{R}) \Psi(\mathbf{R}) d\mathbf{R}}{\int \Psi^*(\mathbf{R}) \Psi(\mathbf{R}) d\mathbf{R}}, \quad (4.21)$$

where the spatial wave function $\Psi(\mathbf{R})$ is defined by

$$\begin{aligned} \Psi(\mathbf{R}) &= \Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \\ &\Psi(\{\mathbf{r}_1, \uparrow\}, \dots, \{\mathbf{r}_{N_\uparrow}, \uparrow\}, \{\mathbf{r}_{N_\uparrow+1}, \downarrow\}, \dots, \{\mathbf{r}_N, \downarrow\}). \end{aligned} \quad (4.22)$$

The new wave function $\Psi(\mathbf{R})$ gives the same expectation values as $\Psi(\mathbf{X})$ but is much more convenient to use in QMC calculations.

This analysis allows us to replace the spin-dependent trial state $\Psi(\mathbf{X})$ by a spin-independent state $\Psi(\mathbf{R})$. The new state is antisymmetric with respect to exchange of the spatial coordinates of pairs of spin-up electrons or pairs of spin-down electrons, but has no specific symmetry on exchange of the spatial coordinates of electrons with different spins. We are therefore treating spin-up electrons as distinguishable from spin-down electrons. We can reconstruct $\Psi(\mathbf{X})$ by multiplying $\Psi(\mathbf{R})$ by the spin function

$$\delta_{\sigma_1, \uparrow} \dots \delta_{\sigma_{N_\uparrow}, \uparrow} \delta_{\sigma_{N_\uparrow+1}, \downarrow} \dots \delta_{\sigma_N, \downarrow} \quad (4.23)$$

and antisymmetrizing.

F. The cusp conditions

We have made repeated reference to the cusp conditions (Kato, 1957; Pack and Brown, 1966), which we now discuss in more detail. Consider the Schrödinger equation for the ground state of the hydrogen atom,

$$\hat{H}\psi_{1s} = -\frac{1}{2}\nabla^2\psi_{1s} - \frac{1}{r}\psi_{1s} = E_{1s}\psi_{1s}. \quad (4.24)$$

The potential energy diverges as $r \rightarrow 0$ but $\hat{H}\psi_{1s}$ remains finite because there is a cancelling divergence in the kinetic energy. This is reflected in the shape of the $1s$ eigenfunction, which has a sharp cusp at the origin.

The potential energy diverges whenever an electron approaches a nucleus (as long as that nucleus is not represented by a pseudopotential) or another electron. Many-electron wave functions therefore contain similar cusps. Any necessary electron-nuclear cusps are built into the single-particle orbitals obtained by solving the mean-field HF or LDA equations, but these orbitals do not contain the electron-electron cusps. The purpose of applying the cusp conditions (Kato, 1957; Pack and Brown 1966) stated in Eq. (4.6) is to build these missing cusps into the Jastrow factor. Experience shows that enforcing the cusp conditions reduces both the average energy and its variance.

The many-electron ground state $\Psi_0(\mathbf{R})$ satisfies the Schrödinger equation

$$\hat{H}\Psi_0(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = E_0\Psi_0(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N), \quad (4.25)$$

where

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_{\mathbf{r}_i}^2 + V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N). \quad (4.26)$$

If we pick a pair of electrons i and j and introduce the difference and center of mass variables $\mathbf{r} = \mathbf{r}_i - \mathbf{r}_j$ and $\mathbf{r}_{\text{cm}} = (\mathbf{r}_i + \mathbf{r}_j)/2$, \hat{H} may be rewritten as

$$\hat{H} = -\nabla_{\mathbf{r}}^2 - \frac{1}{4}\nabla_{\mathbf{r}_{\text{cm}}}^2 - \frac{1}{2} \sum_{\substack{k=1 \\ (k \neq i, j)}}^N \nabla_{\mathbf{r}_k}^2 + V(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N). \quad (4.27)$$

Since Ψ_0 is an exact eigenfunction the corresponding exact local energy $E_{L0} = \Psi_0^{-1}\hat{H}\Psi_0$ is everywhere equal to the ground state energy E_0 and does not diverge as $r \rightarrow 0$.

The Slater-Jastrow trial function Ψ is not an exact eigenfunction and so the local energy $E_L = \Psi^{-1}\hat{H}\Psi$ is not constant and may diverge as $r \rightarrow 0$. To study this possibility we write Ψ in the form

$$\Psi = e^{-u(r)} f(\mathbf{r}), \quad (4.28)$$

where $f(\mathbf{r})$ includes the Slater determinants and everything from the Jastrow factor except the term involving $u_{\sigma_i, \sigma_j}(r)$. Although $f(\mathbf{r})$ depends on all the variables needed to describe the system, only the dependence on \mathbf{r} has been shown explicitly; the spin labels of $u_{\sigma_i, \sigma_j}(r)$ have also been suppressed. For the sake of simplicity, we shall assume that u is a function of r only and does not depend on \mathbf{r}_i and \mathbf{r}_j separately, although the cusp conditions apply more generally.

Consider the behavior of the local energy

$$\frac{1}{\Psi} \hat{H} \Psi = \frac{1}{e^{-u(r)} f(\mathbf{r})} \hat{H} e^{-u(r)} f(\mathbf{r}) \quad (4.29)$$

as r tends to zero while \mathbf{r}_{cm} and all the other electron positions are held fixed. For almost every possible choice of the fixed coordinates, the function $f(\mathbf{r})$ and all its derivatives are finite as $r \rightarrow 0$, so any divergent terms in the kinetic energy must arise from the action of the $\nabla_{\mathbf{r}}^2$ operator on $e^{-u(r)}$. The local energy therefore remains finite as long as

$$\begin{aligned} &\frac{1}{e^{-u} f} \left(-\nabla^2 + \frac{1}{r} \right) (e^{-u} f) \\ &= u'' + \frac{2u'}{r} - (u')^2 + 2u' \hat{\mathbf{r}} \cdot \frac{\nabla f}{f} - \frac{\nabla^2 f}{f} + \frac{1}{r} \end{aligned} \quad (4.30)$$

remains finite, where $\hat{\mathbf{r}}$ is a unit vector in the \mathbf{r} direction and the primes denote differentiation with respect to r .

If electrons i and j have opposite spins, the value of $f(\mathbf{r} = \mathbf{0})$ is in general non-zero. We insist that u' and u'' tend to finite values as $r \rightarrow 0$, and so only the second and final terms in Eq. (4.30) diverge. These two divergences cancel if we impose the opposite-spin cusp condition

$$\left. \frac{du}{dr} \right|_{r=0} = -\frac{1}{2}. \quad (4.31)$$

If electrons i and j have parallel spins, the Pauli principle ensures that $f(\mathbf{r})$ is an odd function of \mathbf{r} . It therefore has a Taylor expansion of the form,

$$f(\mathbf{r}) = \mathbf{a} \cdot \mathbf{r} + \mathcal{O}(r^3), \quad (4.32)$$

where $\mathbf{a} = \nabla f(\mathbf{r})|_{\mathbf{r}=\mathbf{0}}$. It follows from the form of this series (which is assumed to have a finite radius of convergence) that the Laplacian of f tends to zero as $r \rightarrow 0$, but the second, fourth, and last terms in Eq. (4.30) all diverge. The sum of the divergent contributions is

$$\frac{2u'}{r} + \frac{2u'}{\mathbf{a} \cdot \mathbf{r}} \hat{\mathbf{r}} \cdot \mathbf{a} + \frac{1}{r} = \frac{4u' + 1}{r}, \quad (4.33)$$

and so the parallel-spin cusp condition is

$$\left. \frac{du}{dr} \right|_{r=0} = -\frac{1}{4}. \quad (4.34)$$

V. SELECTED APPLICATIONS OF QUANTUM MONTE CARLO TO GROUND STATES

A. Cohesive energies of solids

The cohesive energy of a solid is the difference between the energy of a dispersed gas of the constituent atoms or molecules and the energy of the solid at zero temperature. Cohesive energies can be obtained by measuring the latent heat of sublimation and extrapolating to zero temperature. Calculating a cohesive energy is a severe test of QMC techniques because it is the energy difference between two very different systems. To obtain an accurate cohesive energy within VMC one must use trial wave functions of closely matching accuracy for the atom and solid. DMC calculations are less sensitive to the trial wave function but the quality of the fixed nodal surfaces must still be comparable.

In the pioneering work of Fahy, Wang, and Louie (1988, 1990a), VMC methods with non-local pseudopotentials were used to calculate cohesive energies of diamond-structure C and Si. The calculations were performed with simulation cells containing 16 and 54 atoms and used trial wave functions of the Slater-Jastrow type. The VMC results were surprisingly accurate, giving cohesive

energies within about 0.2 eV of the experimental values. This success was followed by a VMC and DMC study of diamond-structure Si by Li, Ceperley, and Martin (1991). This calculation represented another significant step forward because it was the first application of the DMC method to a heavy-atom solid. The DMC cohesive energy of Si obtained by Li, Ceperley, and Martin (1991) differs from experiment by only about 0.1 eV, although a correction of about 0.2 eV was applied to compensate for an error introduced by using a pseudo-Hamiltonian of the type discussed in Sec. VIII.F to eliminate the core electrons. Another important advance was the development of the locality approximation for evaluating the energy of a non-local pseudopotential within DMC calculations (Mitas, Shirley, and Ceperley 1991). This scheme has significantly increased the scope of DMC calculations and has made it possible to perform accurate DMC calculations of the cohesive energies of heavy-atom solids.

QMC calculations of the cohesive energies of a number of solids have been performed, and in Table I we have compiled the available results on the tetrahedrally bonded semiconductors, C, Si, Ge, and cubic BN. The calculated results include corrections for the finite sizes of the simulation cells and for the zero-point motion of the nuclei. We refer the reader to the original references for details of these corrections. The finite-size corrections are particularly important and were very carefully studied by Kent, Hood, *et al.* (1999). DMC results for cells containing up to 250 atoms were used to obtain a cohesive energy for Si of 4.63(2) eV per atom, which is very close to the experimental value of 4.62(8) eV per atom. This indicates that the DMC method can give cohesive energies to better than 0.1 eV per atom in *sp*-bonded systems. The data in Table I shows the well-known trend that the cohesive energies calculated within the local spin density approximation (LSDA) are too large in *sp*-bonded systems. The largest contribution to the error arises from the atom; for example, in silicon the LSDA energy is about 1.1 eV per atom too high in the pseudo-atom and about 0.4 eV too high in the pseudo-solid, resulting in an overestimation of the cohesive energy by about 0.7 eV.

Very few groups have yet applied QMC methods to more strongly correlated real solids, but examples include the study of NiO by Tanaka (1993) and a study of the Fe atom by Mitas (1994). The results so far suggest that accuracies better than 1 eV per atom can be obtained. This is still a far cry from the accuracy required to study high temperature superconductivity, for example, which would need an improvement of at least two orders of magnitude. The major difficulty is to construct trial wavefunctions that capture the essence of the physics of such complicated systems.

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.

| Method | Si | Ge | C | BN |
|--------|----------------------|----------------------|----------------------|-----------------------|
| LSDA | 5.28 ^a | 4.59 ^a | 8.61 ^a | 15.07 ^e |
| VMC | 4.38(4) ^c | 3.80(2) ^b | 7.27(7) ^d | 12.85(9) ^e |
| | 4.82(7) ^d | | 7.36(1) ^f | |
| | 4.48(1) ^g | | | |
| DMC | 4.51(3) ^c | 3.85(2) ^b | 7.46(1) ^f | |
| | 4.63(2) ^g | | | |
| Exp. | 4.62(8) ^a | 3.85 ^a | 7.37 ^a | 12.9 ^h |

^aFarid and Needs (1992), and references therein.

^bRajagopal *et al.* (1995).

^cLi, Ceperley, and Martin (1991).

^dFahy, Wang, and Louie (1990a). 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.

^eMalatesta, Fahy, and Bachelet (1997).

^fKent *et al.* (2000).

^gLeung *et al.* (1999).

^hEstimated by Knittle *et al.* (1989) from experimental results on hexagonal BN.

B. Phases of the electron gas

The homogeneous electron gas is the simplest realistic model of interacting electrons and studies of its properties are still yielding new insights into electronic many-body phenomena. Consider a uniform electron gas with number density $n = (4\pi r_s^3/3)^{-1}$. In terms of the scaled variables $\mathbf{r}'_i = \mathbf{r}_i/r_s$ the Hamiltonian is:

$$\hat{H} = -\frac{1}{2r_s^2} \sum_{i=1}^N \nabla_{\mathbf{r}'_i}^2 + \frac{1}{2r_s} \sum_{i=1}^N \sum_{\substack{j=1 \\ (j \neq i)}}^N \frac{1}{|\mathbf{r}'_i - \mathbf{r}'_j|}. \quad (5.1)$$

(In real calculations it is necessary to add a uniform positive background and use a periodic Coulomb interaction as discussed in Sec. IX.C.) Note that the kinetic and potential energy operators in Eq. (5.1) scale differently as functions of r_s . This means that in the very high density (small r_s) limit the interactions become negligible and the wave function may be approximated as a HF determinant, while at very low density (large r_s) the Coulomb interactions dominate and the kinetic energy can be ignored. In the low density limit the electrons behave like classical charges at zero temperature and should freeze into a so-called Wigner crystal (Wigner, 1934). The electron gas must therefore undergo a first-order phase transition as the density is lowered.

At intermediate densities another possibility is that the electron gas may become partially or wholly ferromagnetic. This was first predicted within the HF approximation by Bloch (1929), who considered the behavior of the HF energy as a function of the spin polarization $\zeta = (N_\uparrow - N_\downarrow)/N$. He found that if the density was high the minimum energy was at $\zeta = 0$ and the gas was paramagnetic, but that if r_s was greater than approximately 5.45 the minimum jumped to $\zeta = 1$ and the gas became completely polarized. (For comparison, Cs has an r_s of 5.62.) The HF approximation is not accurate for such large values of r_s and the suggestion that the electron gas polarizes at low density has remained controversial.

The first attempt to study the phases of the electron gas using DMC was made by Ceperley and Alder (1980). Such calculations require fine comparisons of the energies of phases with very different trial wave functions, symmetries, and finite-size errors. The energy differences of interest are so small that it is dangerous to make precise quantitative statements, but Ceperley and Alder did indeed observe the expected paramagnetic \rightarrow ferromagnetic \rightarrow Wigner crystal sequence as the density was lowered. They found that the $\zeta = 1$ ferromagnetic gas was stable for r_s values from 75 to 100, at which point the bcc Wigner crystal became more favorable. The possibility of partial polarization was considered by Alder, Ceperley, and Pollock (1982), who found that the 50% polarized gas was more stable than the 100% polarized gas for all densities above that at which Wigner crystallization was observed. In the region in which the $\zeta = 1$ ferromagnetic gas was more stable than both the unpolarized gas and the Wigner crystal, the 50% polarized gas was therefore even more stable.

Very recently, Ortiz, Harris, and Ballone (1999) recalculated the energies of various phases for a range of low to intermediate electron densities, using much larger systems and a more careful data analysis than in earlier work. Figure 9 shows the energy differences between the paramagnetic and ferromagnetic gas (filled circles), and between the ferromagnetic bcc crystal and the ferromagnetic gas (filled squares) as functions of r_s . Further calculations showed that in fact the spin polarization grows continuously from zero at $r_s = 20 \pm 5$ to 100% at $r_s = 40 \pm 5$, and that the Wigner crystal is stable for $r_s > 65 \pm 10$. Ortiz, Harris, and Ballone (1999) also made the controversial suggestion that the phase with partial spin polarization may explain the recent experimental results on $\text{Ca}_{1-x}\text{La}_x\text{B}_6$, which exhibits very weak ferromagnetism in a narrow range of concentrations, but with a surprisingly high Curie temperature of ≈ 600 K (Young *et al.*, 1999; Ceperley 1999).

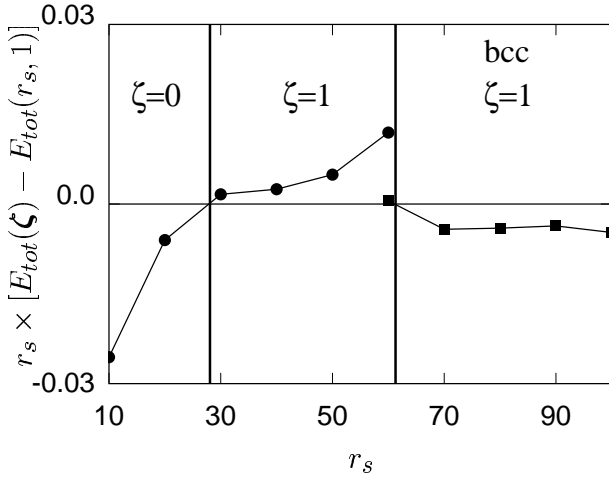


FIG. 9. Total energy difference (eV) times r_s of: (a) the paramagnetic and the ferromagnetic gas (filled circles); and (b) the ferromagnetic bcc crystal and the ferromagnetic gas (filled squares). The statistical error bar is comparable to the size of the symbols. The ferromagnetic gas is stable when both symbols are above zero. (Ortiz, Harris, and Ballone, 1999.)

The results of Ortiz, Harris and Ballone (1999) are significantly different from the results of earlier authors, presumably because the earlier work suffered from larger finite-size errors, although there were also other differences in the technical details of the calculations. Evidently, there is a need for even more accurate studies of the residual finite-size and fixed-node errors. In addition, the relative accuracies of the trial functions used for different phases remain to be thoroughly investigated.

C. Static response of the electron gas

Many experimental techniques, including most types of elastic and inelastic photon, neutron, and electron scattering, measure how solids respond to small perturbations. The results of such experiments are dynamical response functions, the most familiar of which is the dynamical susceptibility $\chi(\mathbf{r}, \mathbf{r}', t - t')$. Although time-dependent response functions such as $\chi(\mathbf{r}, \mathbf{r}', t - t')$ are difficult to obtain using QMC techniques (see Sec. VI.C), both the linear and non-linear responses to *static* perturbations are straightforward to calculate. If, for example, one applies a time-independent external potential $\phi_{\text{ext}}(\mathbf{r}) = 2\phi_{\mathbf{q}} \cos(\mathbf{q} \cdot \mathbf{r})$ to a uniform electron gas with electronic charge density ρ_0 , the charge density response at wave vector \mathbf{q} takes the form $2\rho_{\mathbf{q}} \cos(\mathbf{q} \cdot \mathbf{r})$, where $\rho_{\mathbf{q}}$ may be expanded in powers of $\phi_{\mathbf{q}}$ as follows:

$$\rho_{\mathbf{q}} = \chi(q)\phi_{\mathbf{q}} + \frac{1}{2}\chi^{(3)}(\mathbf{q}, \mathbf{q}, -\mathbf{q})\phi_{\mathbf{q}}^3 + \dots \quad (5.2)$$

The prefactors χ and $\chi^{(3)}$ (not to be confused with the one-body χ term in the Jastrow factor) are known as

the linear and cubic response functions, respectively, and may be obtained by applying potentials of several different strengths and fitting the calculated values of $\rho_{\mathbf{q}}$ to a polynomial in $\phi_{\mathbf{q}}$. Alternatively, one can calculate the change in total energy per electron, which may be expressed in terms of χ and $\chi^{(3)}$ as follows (Senatore, Moroni, and Ceperley, 1999),

$$\frac{\delta E}{N} = \frac{\chi(q)}{\rho_0}\phi_{\mathbf{q}}^2 + \frac{\chi^{(3)}(\mathbf{q}, \mathbf{q}, -\mathbf{q})}{4\rho_0}\phi_{\mathbf{q}}^4 + \dots, \quad (5.3)$$

and fit this to a polynomial in $\phi_{\mathbf{q}}$. Because the DMC method can calculate energies without the need for extrapolated estimation (see Sec. III.D.3), the latter method turns out to be the most accurate.

Calculations of this type have been carried out for both the two-dimensional (Moroni, Ceperley, and Senatore, 1992) and three-dimensional electron gases (Bowen, Sugiyama, and Alder, 1994; Moroni, Ceperley, and Senatore, 1995), but not yet for a real solid, where the lack of translational symmetry complicates matters considerably. Although the static susceptibility is less interesting than the full dynamical susceptibility, it has still been the focus of an enormous amount of theoretical work using many different techniques (see, e.g., Singwi and Tosi, 1981). It was therefore surprising that the QMC results exposed significant flaws in the best known approximations. For further details see the recent review by Senatore, Moroni, and Ceperley (1999).

D. The relativistic electron gas

Calculating quantum relativistic effects for interacting systems is a formidable task, but one simple approach is to perturb about the non-relativistic limit. The relativistic correction to the interaction energy can be obtained to order $1/c^2$ by evaluating the Breit interaction within first order perturbation theory, see Bethe and Salpeter (1957). Kenny *et al.* (1996) used this approach to calculate the relativistic correction to the exchange-correlation energy of the homogeneous electron gas in the density range $r_s = 0.1$ –10. The expectation values of the Breit interaction and the mass-velocity terms were calculated with non-relativistic wave functions using VMC and DMC methods, and extrapolated estimation (see Sec. III.D.3) was used to obtain more accurate expectation values.

The results enabled Kenny *et al.* (1996) to parameterize the relativistic corrections to the LDA exchange and correlation functional. Comparisons of relativistic LDA and accurate relativistic calculations for atoms (Kenny, Rajagopal and Needs, 1995) revealed that the Darwin part of the Breit interaction was well described by a local density approximation, but the term arising from the retardation correction to the Coulomb interaction was very poorly described. The results provided both new insights

and useful expressions for evaluating the relativistic corrections which are important for heavier elements and dense plasmas.

E. Exchange and correlation energies

Section II.D sketched the main features of density-functional theory (DFT), which is by far the most popular method for calculating the electronic properties of solids. The explosion of interest in DFT during the eighties was built on the success of the local density approximation (LDA), which in turn was based on the accurate parameterizations of $\epsilon_{xc}^{\text{hom}}(n)$ published by Vosko, Wilk, and Nusair (1980) and Perdew and Zunger (1981). Both these parameterizations were constructed using values of the XC energy of the uniform electron gas obtained from the pioneering DMC simulations of Ceperley and Alder (1980). It is fair to say that these early DMC results made the success of DFT possible, and that without them DFT may never have grown into the leviathan we know today.

The main drawback of approximate XC energy functionals such as the LDA and the generalized gradient approximation (GGA) (Langreth and Mehl, 1983; Becke, 1988; Perdew *et al.*, 1992) is that it is very difficult to judge their accuracy or to improve them when necessary. Most attempts to devise better approximations have been based on the coupling-constant integration formula (see, e.g., Parr and Yang, 1989):

$$E_{xc}[n] = \frac{1}{2} \int \int \frac{n(\mathbf{r}) \bar{n}_{xc}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}', \quad (5.4)$$

where

$$\bar{n}_{xc}(\mathbf{r}, \mathbf{r}') = \int_0^1 n_{xc}^\lambda(\mathbf{r}, \mathbf{r}') d\lambda, \quad (5.5)$$

and $n_{xc}^\lambda(\mathbf{r}, \mathbf{r}')$ is the XC hole of a fictitious system in which the strength of the electron-electron interaction has been reduced by a factor λ while the external potential has been adjusted to keep the electron density fixed at $n(\mathbf{r})$. The XC hole describes the small “exclusion zone” that forms around an electron at \mathbf{r} due to the effects of the Pauli principle and (for $\lambda \neq 0$) the Coulomb interaction. A sum rule guarantees that the integral of $n_{xc}^\lambda(\mathbf{r}, \mathbf{r}')$ over \mathbf{r}' is always equal to -1 , independent of the position \mathbf{r} or the value of λ .

Equation (5.4) suggests a natural (although by no means unique) definition of the XC energy density:

$$e_{xc}(\mathbf{r}) = \frac{1}{2} \int \frac{n(\mathbf{r}) \bar{n}_{xc}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'. \quad (5.6)$$

One possible way to obtain this quantity is to use VMC to calculate $n_{xc}^\lambda(\mathbf{r}, \mathbf{r}')$ for several different values of λ and

to evaluate the integrals in Eqs. (5.5) and (5.6) numerically. This has recently been done for Si (Hood *et al.*, 1997, 1998) and for inhomogeneous electron gases subject to cosine-wave external potentials (Nekovee *et al.*, 1999). Filippi and co-workers have calculated exchange-correlation energy densities for a model simple harmonic system (Filippi, Umrigar, and Taut, 1994) and for various atoms (Filippi, Gonze, and Umrigar, 1996) using a different definition of e_{xc} that is not based on the coupling-constant integration formula. Because they were studying small systems, these authors were also able to calculate the exact exchange-correlation potential. The comparisons of the QMC results with the various approximate XC energy functionals are instructive and may aid the development of better approximations in the future.

Figure 10 (Hood *et al.*, 1998) shows how the exchange and correlation components of the accurate XC energy density calculated using VMC differ from those assumed in the LDA. The plot shows a cross section in the (110) plane of crystalline silicon. It can be seen that the LDA exchange energy density is not negative enough on average, while the LDA correlation energy density is too negative. When integrated over the unit cell the two errors largely cancel, and thus the total XC energy is very accurate; this explains why LDA calculations of the properties of silicon work so well. Nekovee *et al.* (1999) observed a similar cancellation of errors in cosine-wave jellium, so this behavior seems to be fairly common.

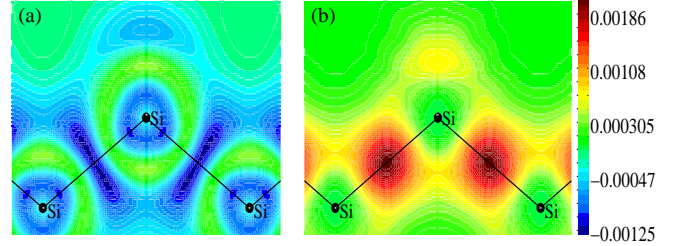


FIG. 10. Contour plots in the (110) plane of Si in the diamond structure: (a) $e_x^{VMC}(\mathbf{r}) - e_x^{LDA}(\mathbf{r})$; (b) $e_c^{VMC}(\mathbf{r}) - e_c^{LDA}(\mathbf{r})$. The chains of atoms and bonds are represented schematically. The contours are in atomic units. (Hood *et al.*, 1998.)

Figure 11 (Nekovee, Foulkes, and Needs, 2000) shows $e_{xc}^{VMC} - e_{xc}^{LDA}$ and $e_{xc}^{VMC} - e_{xc}^{ADA}$ for three different strongly inhomogeneous cosine-wave jellium systems. The ADA (average density approximation) is a simple non-local XC energy functional devised by Gunnarsson, Jonson, and Lundqvist (1979). All three systems have the same average electron density as a uniform electron gas with $r_s = 2$. The Fermi wave vector, Fermi wavelength, and Fermi energy of this uniform system are denoted by k_F^0 , λ_F^0 , and E_F^0 respectively. The applied cosine-wave potentials are of the form $V_q \cos(\mathbf{q} \cdot \mathbf{r})$ with $V_q = 2.08 E_F^0$, but the wave vector q is different for each system.

The most striking feature of Fig. 11 is the strong similarity between the LDA errors and the Laplacian of the electron density. This suggests that gradient-corrected XC energy functionals should include a Laplacian dependence as well as the more familiar gradient terms. Other authors, including Engel and Vosko (1993), Umrigar and Gonze (1994), and Becke (1998) have considered this possibility before, but not for similar reasons. The definition of the XC energy density given in Eq. (5.6) is not the one on which the GGA is based (Burke, Cruz, and Lam, 1998), and so it is not certain that the GGA can be improved by the addition of Laplacian terms, but the possibility is worth investigating. The non-local ADA functional is clearly more accurate than the LDA at most points \mathbf{r} . However, the cancellation of errors that explains the success of the LDA in Si does not occur for the ADA, and the ADA total energy of Si is less accurate than the LDA total energy (Hood *et al.*, 1997).

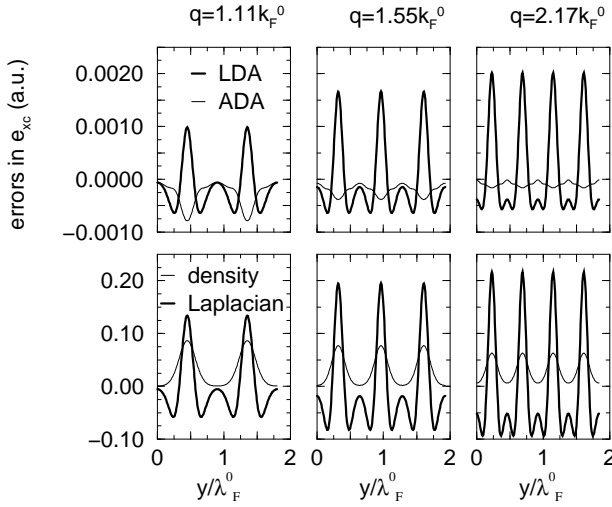


FIG. 11. The upper graphs show $e_{xc}^{VMC} - e_{xc}^{LDA}$ and $e_{xc}^{VMC} - e_{xc}^{ADA}$ for three different strongly inhomogeneous cosine-wave jellium systems. The lower graphs show the corresponding electron densities $n(\mathbf{r})$ and (the negatives of the) Laplacians $\nabla^2 n(\mathbf{r})$. All three systems have the same average electron density as a uniform electron gas with $r_s = 2$. The Fermi wave vector, Fermi wavelength, and Fermi energy of this uniform system are denoted by k_F^0 , λ_F^0 , and E_F^0 respectively. The applied cosine-wave potentials are of the form $V_q \cos(\mathbf{q} \cdot \mathbf{r})$ with $V_q = 2.08 E_F^0$. The systems have different wave vectors q as shown. (Nekovee, Foulkes, and Needs, 2000.)

F. Compton scattering in Si and Li

QMC methods have recently been used to study inelastic x-ray scattering at large momentum transfers (Compton scattering) by Kralik, Delaney, and Louie (1998) and Filippi and Ceperley (1999). Compton scattering is a powerful technique for probing electron correlation effects in solids and the Fermi surfaces of metals. Within

the impulse approximation (Eisenberger and Platzman, 1970) the recoiling electron is treated as free, and the scattering cross section is proportional to the Compton profile. The Compton profile at momentum p for direction $\hat{\mathbf{e}}$ is given by

$$J(p) = \int d\mathbf{k} n(\mathbf{k}) \delta(\mathbf{k} \cdot \hat{\mathbf{e}} - p), \quad (5.7)$$

where the momentum distribution $n(\mathbf{k})$ can be expressed in terms of the wave function Ψ :

$$n(\mathbf{k}) = \frac{N}{V} \int d\mathbf{r}_1, \dots, d\mathbf{r}_N \int d\mathbf{r} e^{i\mathbf{k} \cdot \mathbf{r}} \times \Psi^*(\mathbf{r}_1, \dots, \mathbf{r}_N) \Psi(\mathbf{r}_1 + \mathbf{r}, \dots, \mathbf{r}_N), \quad (5.8)$$

and there are N electrons in volume V . Kralik, Delaney, and Louie (1998) performed VMC calculations of the momentum density of Si, while Filippi and Ceperley (1999) performed similar calculations for Li using both the VMC and DMC methods.

The primary interest is in the valence electrons, although all the electrons contribute to the scattering. Pseudopotentials were used in the QMC calculations so that only the valence electrons were included, and corrections were added to account for the core-valence orthogonality. Experimental valence Compton profiles are deduced by subtracting a contribution calculated for the core electrons. Fig. 12 shows experimental valence Compton profiles for Li (Sakurai *et al.*, 1995) together with the valence Compton profiles calculated by Filippi and Ceperley (1999) using the LDA and VMC methods. The overall agreement between the shapes of the calculated and experimental valence Compton profiles is good, but both the LDA and VMC profiles are too large at low momenta and too small above the Fermi momentum ($p_F = 0.5905 \text{ a.u.}$). Further calculations using the more accurate DMC method confirmed these results.

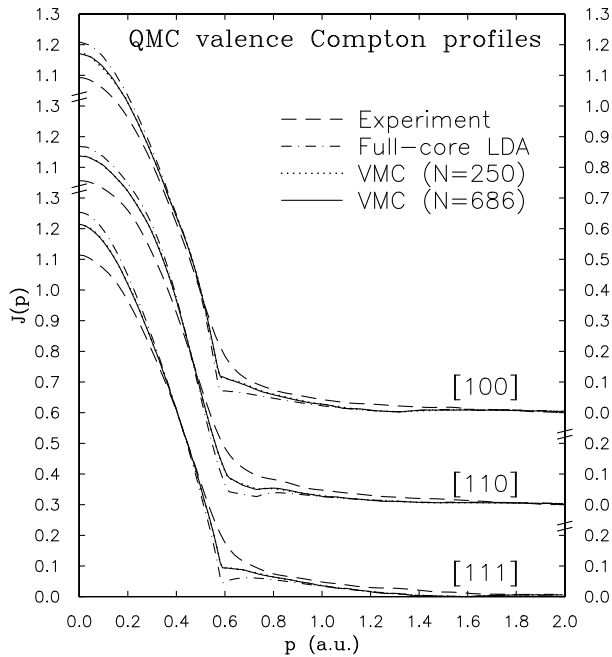


FIG. 12. Valence Compton profiles for Li in the [100], [110], and [111] directions. VMC results for the cells with 250 and 686 atoms are compared with the LDA valence profiles from an all-electron LDA calculation and with experiment. From Filippi and Ceperley (1999).

Independent-particle calculations of Compton profiles may be corrected for electron correlation effects using the theory of Lam and Platzman (1974). This correction is in principle exact within the framework of density functional theory, although it is approximate within the LDA. Both Kralik, Delaney, and Louie (1998) and Filippi and Ceperley (1999) found that the Lam-Platzman correction gave a good account of the small differences between the QMC and LDA valence Compton profiles. This indicates that the approximate treatment of the valence electron correlations is not the primary reason for the discrepancies between the calculated and measured Compton profiles of Si and Li.

Before the advent of the QMC calculations it was widely believed that the approximate treatment of electron correlation was responsible for the discrepancies between LDA and measured Compton profiles. These QMC studies will force a more serious consideration of other possibilities such as thermal effects and the possible inadequacy of the impulse approximation for calculating the contribution from the core electrons.

G. Solid Hydrogen

Hydrogen, the most abundant of all elements, has been the focus of a great deal of experimental and theoretical research. One interesting area of application is to the interiors of giant planets. For example, Jupiter is

about 90% hydrogen at high pressures and temperatures, most of which exists in a fluid metallic state. At low temperatures and pressures hydrogen solidifies into the only known example of a molecular quantum crystal, in which the molecules rotate freely due to quantum rather than thermal effects. With sufficient applied pressure the molecules lock into preferred orientations and at very high pressures hydrogen is expected to form a metallic solid, which could exhibit interesting properties such as high temperature superconductivity. Metallization, defined as the existence of a finite DC electrical conductivity at low temperatures, has not, however, been observed in terrestrial static pressure experiments on hydrogen. Understanding the behaviour of hydrogen at high pressures is a challenge to experimentalists and theorists alike. High pressures are simple to achieve in calculations, but quantum effects due to the small mass of the proton become increasingly important. In addition the energy differences between candidate structural phases are very small and predictions from mean-field calculations may not be trustworthy.

QMC techniques are ideal for studying the hydrogen problem because they allow an accurate treatment of the quantum motion of both the electrons and protons. Among the very first applications of QMC methods to real solids was the DMC work on solid hydrogen of Ceperley and Alder (1981, 1987), while further DMC studies have been performed by Natoli, Martin, and Ceperley (1993, 1995). These studies encompassed molecular and monatomic phases in which the protons are localized around lattice sites, so that exchange effects between protons are negligible and the wave function can be taken to be symmetric in the proton coordinates. The trial wave functions used are of the form

$$\Psi = \Psi_e \exp \left[- \sum_{\alpha \neq \beta} u_{pp}(\mathbf{d}_\alpha - \mathbf{d}_\beta) - \sum_{\alpha} c_\alpha |\mathbf{d}_\alpha - \bar{\mathbf{d}}_\alpha|^2 \right], \quad (5.9)$$

where \mathbf{d}_α and $\bar{\mathbf{d}}_\alpha$ denote, respectively, the proton positions and the lattice sites. The factor Ψ_e is the electronic part of the wave function for fixed proton positions, and is taken to be a standard Slater-Jastrow form, while the symmetric proton part consists of a proton-proton correlation term, u_{pp} , and a product of Gaussian functions which localize the protons around the lattice sites. Ceperley and Alder (1987) explored both molecular and atomic phases, predicting a transition from the freely rotating molecular quantum crystal to a rotationally ordered molecular phase at about 100 GPa, which was subsequently identified in experiments (Lorenzana, Silvera, and Goettel, 1990), and a further transition to atomic phases at about 300 GPa. The DMC calculations of Natoli, Martin, and Ceperley (1993) showed that the zero-point motion of the protons is a very important

factor in determining the relative stabilities of the high pressure monatomic phases of hydrogen. The electronic energy favours low coordination numbers at these densities, but this is counteracted by the effect of the zero point energy which favours high coordination and symmetry. At still higher pressures the Madelung energy of the proton lattice should result in a transition to close packed structures. Natoli, Martin, and Ceperley (1993) found the diamond structure to be the most stable of the monatomic phases up to pressures of about 400 GPa. They concluded that the usual estimations of zero-point energies from mean-field theories assuming the harmonic approximation were inaccurate, in some cases by a factor of two. Natoli, Martin, and Ceperley (1995) found that insulating molecular phases with canted orientations of molecules on an hcp lattice were favoured at pressures above about 123 GPa. Combining this data with their earlier DMC results on monatomic phases of hydrogen, they predicted that the transition to a monatomic diamond-structure phase would occur at around 300 GPa (see Fig. 13). Such a transition pressure would be consistent with values estimated by extrapolating the measured frequency of the molecular stretching mode to zero.

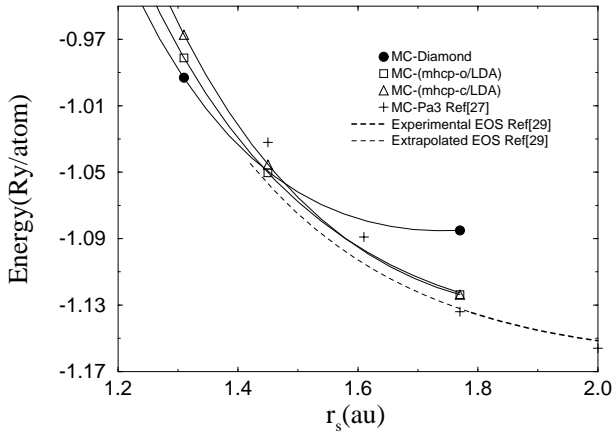


FIG. 13. The equation of state of hydrogen determined by DMC calculations and compared with extrapolations from experiment due to Hemley *et al.* (1990). DMC data are shown for the low-pressure *Pa3* molecular structure (Ceperley and Alder, 1987), the diamond structure (Natoli, Martin, and Ceperley, 1995), and two molecular phases consisting of differing orientations of molecules on an hcp lattice (mhcp-*o* and mhcp-*c*) (Natoli, Martin, and Ceperley, 1993). From Natoli, Martin, and Ceperley (1995).

Both experimental and theoretical results suggest that the metallization of hydrogen occurs at a pressure not too far above the largest static pressure currently attainable of about 300 GPa. The unique capabilities of the DMC method have already made substantial contributions to the theoretical understanding of solid hydrogen at zero temperature. The path integral Monte Carlo method has been applied to hydrogen at high temperatures (Pierleoni

et al., 1994; Magro *et al.*, 1996; Militzer, Magro, and Ceperley, 1999), and it is hoped that further DMC and path integral Monte Carlo studies will help in obtaining a more complete picture of the phase diagram of hydrogen.

H. Clusters

Clusters of atoms are intermediate between small molecules and solids and exhibit a rich variety of physical and chemical properties. They also form an important testing ground for QMC methods because of the absence of the finite-size errors which plague periodic boundary conditions calculations. This facilitates comparisons between QMC and other computational methods and with experimental data.

Encouraging progress in applying QMC methods was demonstrated in studies of silicon and carbon clusters of intermediate sizes up to 20 atoms (Grossman and Mitas, 1995; Grossman, Mitas, and Raghavachari, 1995; Mitas and Grossman, 1997). These studies focused on the binding (atomization) energies and the energetic ordering of competing isomers. Perhaps the most striking result was obtained by Grossman and coworkers in evaluating the energetic ordering of three isomers (ring, graphitic bowl, dodecahedron cage, see Fig. 14) of C_{20} in an attempt to search for the smallest carbon fullerene (Grossman, Mitas, and Raghavachari, 1995). Raghavachari and coworkers (1993) had earlier discovered large discrepancies in the energy differences between the three isomers as calculated using the LDA and GGA methods. For example, the LDA predicts the cage structure to be about 4 eV lower in energy than the ring while the BLYP-GGA functional predicts essentially the opposite (see Fig. 15). Such differences are unusually large for moderate sized systems. DMC calculations with HF geometries showed that the graphitic bowl was the lowest energy isomer. The results in Fig. 15 were met with surprise and even disbelief in the electronic structure community, and so it is reassuring that an independent method confirmed the DMC results. Calculations by Murphy and Friesner (1998), using their newly developed high level perturbation approach based on generalized valence bond wave functions, agreed with the DMC results to within the error bars. The QMC calculations on C_{20} therefore provided new and unexpected results that would have been very difficult to obtain by any other method, and highlighted the limited predictive power of LDA/GGA approaches for systems that were thought to be well understood.

The study of carbon clusters within DMC was further extended to C_{24} , C_{26} , C_{28} , and C_{32} by Kent *et al.* (2000). These authors found that the caged structure of C_{24} is higher in energy than other isomers, but that for C_{26} and C_{28} caged structures are slightly favored over other

isomers. For C_{32} the caged geometry is very clearly the most stable.

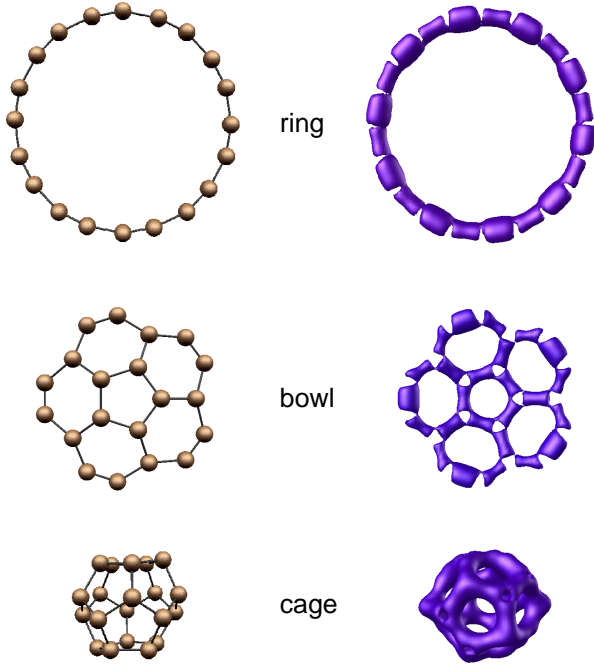


FIG. 14. Geometries and charge densities of C_{20} isomers calculated using the HF method (Grossman, Mitas, and Raghavachari, 1995).

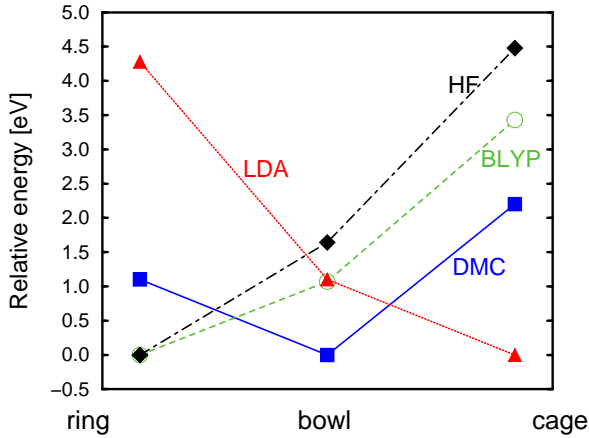


FIG. 15. Relative energies of C_{20} isomers from the HF, LDA, BLYP, and DMC methods. The energies are given relative to the lowest energy isomer within the given theory (Grossman, Mitas, and Raghavachari, 1995).

I. Formation energies of silicon self-interstitials

Silicon is the most important material in the micro-electronics industry, and the diffusion of dopant impu-

rity atoms during thermal processing is one of the factors that limit how small semiconductor devices can be made. The diffusion of impurity atoms in silicon is critically influenced by intrinsic defects such as self-interstitials and vacancies, and it is therefore of great importance to improve our understanding of these defects. Unfortunately it has not been possible to detect self-interstitials directly, although their presence has been inferred using various techniques (Fahey, Griffin, and Plummer, 1989). Measurements of the self-diffusivity, D_{SD} , of silicon at high temperatures, using radioactive isotopes of silicon as tracers, have established an Arrhenius behaviour with an activation energy in the range 4.1–5.1 eV (Frank *et al.*, 1985). D_{SD} is usually written as the sum of contributions from independent diffusive mechanisms, and these contributions can each be written as the product of the diffusivity, D_i , and the concentration, C_i , of the relevant defect, i.e.,

$$D_{SD} = \sum_i D_i C_i. \quad (5.10)$$

It is widely believed that self-interstitial diffusion is important at higher temperatures, while vacancy diffusion is important at lower temperatures (see the review by Gösele, Plössl, and Tan, 1996). The experimental situation regarding self-diffusion in silicon is, however, still highly controversial, especially when it comes to the individual values of D_i and C_i . Indeed, experimental data have been used to support values of the diffusivity of the silicon self-interstitial that differ by ten orders of magnitude at the temperatures of around 800°C at which silicon is processed (Eaglesham, 1995).

Many theoretical studies of self-interstitials in silicon have been carried out. The most advanced of these have used LDA-DFT methods to calculate the defect formation energies and energy barriers to diffusion (Bar-Yam and Joannopoulos, 1984; Blöchl *et al.*, 1993). The consensus view emerging from these calculations is that the split- $\langle 110 \rangle$, hexagonal, and tetrahedral self-interstitial defects are the lowest in energy. Another interesting suggestion is that self-diffusion could occur without point defects via exchange of neighbouring atoms in the perfect lattice, and Pandey (1986) has proposed such a concerted exchange mechanism for self-diffusion in silicon. The structures of the split- $\langle 110 \rangle$, hexagonal, and tetrahedral interstitial defects, and of the saddle point of Pandey's concerted-exchange mechanism are illustrated in Fig. 16.

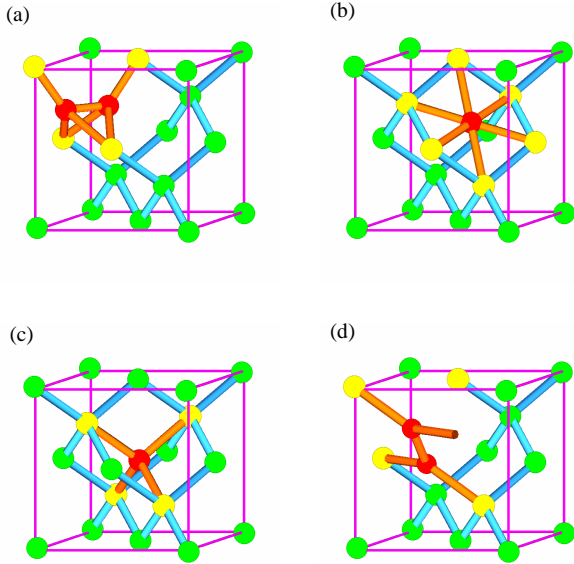


FIG. 16. Self-interstitial defects in silicon: (a) the split- $\langle 110 \rangle$, (b) hexagonal, and (c) tetrahedral interstitial defects, and (d) the saddle point of the concerted-exchange mechanism. The atom(s) forming the defect are shown in red, while the nearest neighbours to the defect atoms are shown in yellow. The bonds between the defect and nearest neighbour atoms are shown in orange. (Leung *et al.*, 1999)

Leung *et al.* (1999) performed fixed-node DMC calculations and DFT calculations using the LDA and the PW91 GGA to determine the formation energies of self-interstitials in silicon. Within each method they found the split- $\langle 110 \rangle$ and hexagonal interstitials to be the most stable, although the formation energies were significantly different in the three methods (see Table II). The DMC formation energies are about 1 eV larger than the PW91-GGA values and 1.5 eV larger than the LDA values. Leung *et al.* (1999) used these DMC data to estimate a value for the activation energy for self-interstitial diffusion of about 5 eV, which is consistent with the value deduced from experiment of 4.84 eV (Gösele, Plössl, and Tan, 1996). The activation energies predicted by the LDA and PW91-GGA density functionals are, however, considerably lower than the experimental value, and do not provide a satisfactory explanation of self-diffusion in silicon. This study has highlighted the importance of a proper treatment of electron correlation when treating such systems.

TABLE II. LDA, PW91-GGA, and DMC formation energies in eV of the self-interstitial defects and the saddle point of the concerted-exchange mechanism. ^a16 atom supercell, ^b54 atom supercell. Note that the DMC results for the 16 and 54 atom simulation cells are consistent, indicating that the residual finite-size effects are small.

| Defect | LDA | GGA | DMC ^a | DMC ^b |
|------------------------------|------|------|------------------|------------------|
| Split- $\langle 110 \rangle$ | 3.31 | 3.84 | 4.96(24) | 4.96(28) |
| Hexagonal | 3.31 | 3.80 | 4.70(24) | 4.82(28) |
| Tetrahedral | 3.43 | 4.07 | 5.50(24) | 5.40(28) |
| Concerted Exchange | 4.45 | 4.80 | 5.85(23) | 5.78(27) |

J. Jellium surfaces

The jellium model with the positive background terminated at a plane provides the simplest model of a metal surface. Many theoretical techniques have been used to investigate jellium surfaces, which have become a testing ground for developing new methods of studying correlation effects in inhomogeneous systems. Li *et al.* (1992) used DMC to study the surface of jellium at a density of $r_s = 2.07$, which is the average valence charge density of aluminium. Acioli and Ceperley (1996) considered five densities in the range $r_s = 1.87 - 3.93$, calculating the surface energies, work functions, charge densities and pair correlation functions. The trial wave functions were of the Slater-Jastrow type with a determinant of LDA orbitals. The final QMC charge densities agreed with the LDA densities to within 2%, supporting the use of LDA orbitals for the trial wave function. The DMC surface energies were higher than the LDA and GGA energies, although quite close to the Fermi-hypernetted-chain (FHNC) results of Krotscheck, Kohn, and Qian (1985) at low densities. The work functions were in reasonable agreement with the FHNC results and were lower than the LDA values. Inside the jellium the pair correlation functions were nearly spherical, but as the electron moved towards the surface they showed significant anisotropy.

The large discrepancy between the DMC and density functional surface energies is surprising and has recently been questioned by Yan *et al.* (2000). They observe that the density functional and DMC energies (Ballone, Umrigar, and Delaly, 1992) of finite jellium spheres are in rather close agreement, and suggest that the discrepancy in the case of the surface may arise from finite-size errors in the DMC simulations.

VI. EXCITED STATES

A. Introduction

Although the VMC and DMC methods were designed to study ground states, they can also provide some information about excited states. The lowest band gap of a solid, which may be measured from a combination of photoemission and inverse photoemission experiments, is the difference between the energies to add and subtract an electron from the N -electron system: $E_g = (E_{N+1} - E_N) - (E_N - E_{N-1})$. This expression involves only ground-state energies and hence is immediately accessible to QMC methods. The first application of DMC to an energy gap in a solid was by Ceperley and Alder (1987), who calculated the minimum energy zone center gap of the $Pa3$ molecular hydrogen crystal as a function of pressure. To evaluate the energies of the $N + 1$ and $N - 1$ electron systems they added uniform neutralizing background charge densities to preserve the charge neutrality of the simulation cell.

Another straightforward way of obtaining excitation energies is to devise a many-electron trial wave function that models an excited state and use it in a VMC calculation. If the chosen trial wave function has a specific symmetry, the variational principle guarantees that the energy obtained is greater than or equal to the eigenvalue of the lowest exact eigenstate of that symmetry.

At first sight it might appear that the DMC method is inapplicable to excited states because the wave function always evolves towards the ground state. In fixed-node DMC, however, the nodal constraint ensures convergence to the lowest energy state compatible with the imposed nodal surface, not to the overall ground state. It is straightforward to show that if the nodal surface of the trial wave function is the same as that of an exact eigenstate then the fixed-node DMC algorithm gives the exact energy of that eigenstate.

Until recently, it was widely believed that a symmetry-constrained variational theorem analogous to the one for VMC also held for fixed-node DMC calculations. It was assumed that the symmetry of the fixed-node ground state produced by the imaginary-time evolution was always the same as that of the trial function used to define the nodes, and hence that the DMC energy was always greater than or equal to the eigenvalue of the lowest exact eigenstate of that symmetry. It has now been shown (Foulkes, Hood, and Needs, 1999) that this is not necessarily the case unless the symmetry of interest corresponds to a one-dimensional irreducible representation of the symmetry group of the Hamiltonian. In other words, the eigenstate should be non-degenerate or have only accidental degeneracies. For multi-dimensional irreducible representations the symmetry of the fixed-node ground state may differ from that of the trial state, and hence

the DMC energy may lie below the eigenvalue of the lowest exact eigenstate with the same symmetry as the trial state. In such cases it is possible to obtain weaker variational principles by choosing trial functions that transform according to one-dimensional irreducible representations of subgroups of the full symmetry group (Foulkes, Hood, and Needs, 1999). In practice, the surprise is that the fixed-node DMC method often works well for excited states regardless of the dimension of the irreducible representation; it even works quite well for excited states that are not the lowest energy states of any particular symmetry.

B. VMC and DMC calculations of excitations in solids

VMC and DMC calculations of excitations in solids are computationally demanding because excitation energies are “ $1/N$ ” effects; that is, the fractional change in the total energy due to the presence of the excitation is inversely proportional to the number of electrons in the simulation cell. Excited-state VMC and DMC simulations therefore require great statistical accuracy and a careful treatment of finite-size effects.

Mitas and Martin (1994) used the DMC method to estimate the band gap of atomic solid nitrogen in the $I2_13$ structure. Simulations were carried out to obtain the total energy of the ground state and of an excited state in which a single electron had been promoted from the highest occupied Γ state to the lowest unoccupied H state. This can be interpreted as a calculation of the energy of an electron-hole pair (Mott-Wannier exciton). The exciton binding energy is usually small (~ 0.01 – 0.1 eV) and so it is easy to obtain the excitation energy once the formation energy of the exciton is known.

Mitas (1996) later used the DMC method to calculate a selection of excitation energies in diamond, and a more complete study has recently been completed by Towler *et al.* (2000), see Table III. The DMC approach gives a good estimation of the $\Gamma_{25'} \rightarrow X_{1c}$ band gap, but somewhat overestimates the valence band width.

TABLE III. Excitation energies (eV) of diamond calculated using the HF, LDA, *GW*, and DMC methods, and compared with experimental data.

| Method | Band gap | Band width |
|-----------|---|---|
| | $\Gamma_{25'} \rightarrow X_{1c}$ | $\Gamma_{1v} \rightarrow \Gamma_{25'}$ |
| HF | 13.2 ^a | 29.4 ^a |
| LDA | 4.6 ^a , 4.63 ^b | 22.1 ^a , 21.35 ^b |
| <i>GW</i> | 6.3 ^c | 22.88 ^c , 23.0 ^d |
| DMC | 6.0(4) ^a , 5.71(20) ^b | 23.9(7) ^a , 24.98(20) ^b |
| Exp. | 6.1 ^f | 23.0(2) ^e |

^aMitas (1996). ^bTowler *et al.* (2000). ^cRohlfing *et al.* (1993). ^dHybertsen and Louie (1986). ^eJiménez *et al.* (1997). ^fEstimated by correcting the measured minimum band gap of 5.48 eV.

Williamson *et al.* (1998) used the DMC method to calculate a large number of excitation energies in Si. By studying many different electron-hole pair excitations, Williamson *et al.* obtained the quasiparticle energies shown in Fig. 17. The solid lines show an empirical pseudopotential bandstructure that may be taken as an accurate representation of the experimental results. The DMC energy of the state at the top of the valence band is set equal to zero by definition, but the rest of the results are meaningful. The low-lying quasiparticle energies are accurate, but the energies of holes lying deeper in the valence bands are significantly overestimated due to deficiencies of the trial wave functions. A new feature of this work was the successful calculation of several different excited states at the same k-point. Moreover, it was found that the DMC method produced equally good results whether or not the electron and hole had the same crystal momentum.

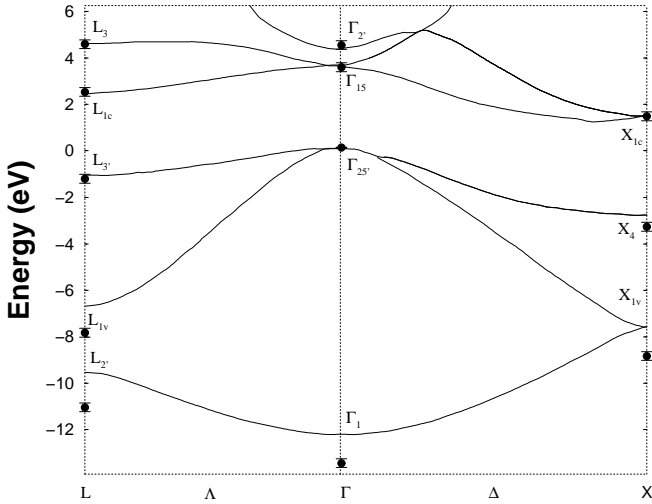


FIG. 17. The DMC bandstructure of Si (Williamson *et al.*, 1998). The solid lines show the empirical pseudopotential bandstructure of Chelikowsky and Cohen (1976).

C. Other QMC methods for excited states

There are a number of alternative methods for calculating excitation energies within QMC. Ceperley and Bernu (1988) combined the idea of the generalized variational principle (i.e., the variational principle for the energies of a set of orthogonal trial functions) with the DMC algorithm to derive a method for calculating the eigenvalues of several different excited states simultaneously. The first application of the Ceperley-Bernu method was to vibrational excited states (Bernu, Ceperley, and Lester, 1990), but it has also been used to investigate electronic excitations of the two-dimensional uniform electron gas (Kwon, Ceperley, and Martin, 1996) and of He atoms in strong magnetic fields (Jones, Ortiz, and Ceperley, 1997). Correlated sampling techniques (Kwon, Ceperley, and Martin, 1996; Jones, Ortiz, and Ceperley, 1997) can be used to reduce the variance, but the Ceperley-Bernu method has stability problems in large systems and has not been applied to a real solid.

Another method that uses the generalized variational principle is based on the extended Koopmans' theorem (EKT) derived independently by Day, Smith, and Garrod (1974) and Morrell, Parr, and Levy (1975). The EKT leads to an approximate expression for the ground- and excited-state energies of the $N + 1$ and $N - 1$ electron systems relative to the ground-state energy of the N electron system. Recently, Kent *et al.* (1998) used this expression in conjunction with the VMC algorithm to calculate the bandstructure of Si and obtained results not much worse than those found using the direct DMC method. The main advantage of the EKT approach is that it allows many quasiparticle energies to be calculated simultaneously. Within the so-called diagonal approximation, which is accurate in Si (Kent *et al.*, 1998), the EKT method reduces to the scheme used previously by Fahy, Wang and Louie (1990b) to calculate hole energies in Si, and by Tanaka (1995) to calculate hole energies in NiO.

VII. WAVE FUNCTION OPTIMIZATION

A. Introduction

The quality of the trial wave function controls the statistical efficiency of the VMC and DMC algorithms and determines the final accuracy obtained. Clearly one would like to use a high quality trial wave function, but there is also an issue of computational efficiency. The most costly part of VMC and DMC calculations is normally the repeated evaluation of the trial wave function (and its gradient and Laplacian). It is therefore important to use a trial wave function that is as accurate as possible yet can be computed rapidly.

By far the most common type of trial wave function used in VMC and DMC calculations for atoms, molecules and solids is the Slater-Jastrow form discussed in Sec. IV. Ideally one would like to perform a simultaneous optimization of the u and χ functions in the Jastrow factor and the orbitals in the Slater determinants. However, almost all work on large systems has so far involved optimizing only the u and χ functions. Typical solid state problems currently involve optimizing of order 10^2 parameters for 10^2 – 10^3 electrons.

B. The cost function

To optimize a wave function containing a set of parameters $\{\alpha\}$ requires some criterion for deciding on the quality of a particular parameter set. To this end we require a *cost function* which is to be minimized with respect to the values of the parameters. The choice of cost function may depend on the application. For instance, if one wants to calculate the best variational bound on the energy in a VMC calculation one should minimize the variational energy,

$$E_V(\alpha) = \frac{\int \Psi_T^2(\alpha) E_L(\alpha) d\mathbf{R}}{\int \Psi_T^2(\alpha) d\mathbf{R}}, \quad (7.1)$$

where $E_L = \Psi_T^{-1}(\alpha) \hat{H} \Psi_T(\alpha)$ is the local energy. It has also been suggested that minimizing the energy will maximize the efficiency of a DMC calculation (Ceperley, 1986). An alternative is to minimize the variance of the energy,

$$\sigma_E^2(\alpha) = \frac{\int \Psi_T^2(\alpha) [E_L(\alpha) - E_V(\alpha)]^2 d\mathbf{R}}{\int \Psi_T^2(\alpha) d\mathbf{R}}, \quad (7.2)$$

which minimizes the statistical error bar on the VMC energy. Another option is to minimize some combination of the variance and energy, such as σ_E^2/E_V^2 (Meierovich, Mushinski and Nightingale, 1996). To maximize the accuracy of an extrapolated estimator (see Subsec. III.D.3) one wants the trial and DMC wave functions to be as close as possible, which can be achieved by maximizing their overlap,

$$\frac{\int \Psi_T(\alpha) \Phi(t \rightarrow \infty) d\mathbf{R}}{\int \Psi_T^2(\alpha) d\mathbf{R}}. \quad (7.3)$$

Maximization of the overlap (Ceperley and Alder, 1987) has rarely been used because it requires a computationally intensive DMC calculation at each step of the optimization.

In practice almost all wave function optimizations have been performed by minimizing the variance of the energy. A number of reasons have been advanced for preferring variance minimization to energy minimization, but the

most important is that it is more stable in large systems as explained below. Optimizing wave functions by minimizing the variance of the energy is actually a very old idea, having been used in the 1930's. The first application using Monte Carlo techniques to evaluate the integrals appears to have been by Conroy (1964), but the present popularity of the method derives from the work of Umrigar and collaborators (Umrigar, Wilson, and Wilkins, 1988).

Let us look at the variance of the energy and the minimization procedure in more detail. The minimum possible value of σ_E^2 is zero, which is obtained if and only if $\Psi_T(\alpha)$ is an exact eigenstate of \hat{H} . The variance therefore has a minimum at every eigenstate of \hat{H} , which makes it a suitable cost function for optimizing both ground and excited states, although it has not been applied widely to excited states. Minimization of σ_E^2 is normally carried out via a correlated-sampling approach in which a set of configurations distributed according to $\Psi_T^2(\alpha_0)$ is generated, where α_0 is an initial set of parameter values. The variance $\sigma_E^2(\alpha)$ is then evaluated as

$$\sigma_E^2(\alpha) = \frac{\int \Psi_T^2(\alpha_0) w(\alpha) [E_L(\alpha) - E_V(\alpha)]^2 d\mathbf{R}}{\int \Psi_T^2(\alpha_0) w(\alpha) d\mathbf{R}}, \quad (7.4)$$

where

$$E_V(\alpha) = \frac{\int \Psi_T^2(\alpha_0) w(\alpha) E_L(\alpha) d\mathbf{R}}{\int \Psi_T^2(\alpha_0) w(\alpha) d\mathbf{R}}, \quad (7.5)$$

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

$$w(\alpha) = \frac{\Psi_T^2(\alpha)}{\Psi_T^2(\alpha_0)}. \quad (7.6)$$

The parameters $\{\alpha\}$ are adjusted until $\sigma_E^2(\alpha)$ is minimized.

The advantage of the correlated-sampling approach is that one does not have to generate a new set of configurations every time the parameter values are changed. In practice the set of configurations is normally regenerated a few times (typically 3 or 4) during the optimization procedure because the estimate of the variance is poor if $|\{\alpha_0\} - \{\alpha\}|$ is too large. A variant of this scheme is to replace the energy $E_V(\alpha)$ by a fixed value, \bar{E} , that is a little below the ground-state energy. Minimization of this modified cost function is equivalent to minimizing a linear combination of E_V and σ_E^2 .

C. Numerical stability of variance minimization

The cost function is evaluated as an average over a set of M configurations, \mathbf{R}_m , drawn from the distribution $\Psi_T^2(\alpha_0)$,

$$\sigma_E^2(\alpha) \simeq \frac{\sum_m^M w(\mathbf{R}_m; \alpha) [E_L(\mathbf{R}_m; \alpha) - E_V(\{\mathbf{R}_m\}; \alpha)]^2}{\sum_m^M w(\mathbf{R}_m; \alpha)}. \quad (7.7)$$

The eigenstates of \hat{H} give the minimum value, $\sigma_E^2 = 0$, for *any* set of configurations because $E_L(\mathbf{R})$ is independent of \mathbf{R} for an eigenstate (Nightingale and Umrigar, 1997; Kent, Needs, and Rajagopal, 1999). This highly desirable feature is not shared by the energy E_V . The fact that the positions of the global minima of the variance are robust to finite sampling is an important advantage of variance minimization over energy minimization.

Direct minimization of the variance of Eq. (7.7) has often been successful. However, in large systems the procedure often exhibits a numerical instability: as the minimization proceeds, a few configurations (often only one) acquire a very large weight. The estimate of the variance is then reduced almost to zero by a set of parameters that usually give extremely poor results in subsequent QMC calculations. This instability has been noticed by many researchers and also occurs in energy minimization. In principle one could overcome it by using more configurations, but the number required is normally impossibly large. Fortunately, in variance minimization (though not in energy minimization) there is a simple remedy, which relies on the observation that the positions of the minima of the variance are not affected by the values of the weights as long as they are positive (Kent, Needs, and Rajagopal, 1999). This is because the exact variance of Eq. (7.2) reaches its minimum value of zero if and only if the trial function is an exact eigenfunction, in which case the local energy is independent of position. The numerical instability may therefore be eliminated by restricting the values of the weights. In some calculations the upper value of the weight is limited (Filippi and Umrigar, 1996), while in others the weights are set to unity (Schmidt and Moskowitz, 1990; Williamson *et al.*, 1996). This freedom is not available in energy minimization because altering the weights alters the positions of the minima.

D. Minimization procedures

The minimization itself may be carried out using standard optimization techniques such as the Levenberg-Marquardt method (Press *et al.*, 1992), which finds the unconstrained minimum of a sum of squares and requires only the function values. If one is optimizing only the Jastrow factor then the storage and CPU requirements may be greatly reduced by using u and χ functions which are linear in the parameters $\{\alpha\}$ (Williamson *et al.*, 1996). This allows the sums over the electron coordinates to be calculated once for each set of configurations, instead of every time the parameter values are changed. An alternative method for performing the optimization

is to use the stochastic gradient approximation (SGA), which was introduced into QMC calculations by Harju *et al.* (1997). The SGA has the appealing feature that it was designed for optimizations in the presence of noise. However, our experience is that deterministic methods of minimization over a fixed set of configurations perform much better than the SGA.

The single-particle orbitals in the Slater determinants are normally obtained from mean-field calculations and will not be optimal in the presence of the Jastrow factor. Direct optimization of the single-particle orbitals using variance minimization is possible in small systems (Umrigar, Nightingale, and Runge, 1993), but would be very expensive in large systems because the number of parameters increases rapidly with system size. A more promising technique (Fahy, 1999; Filippi and Fahy, 2000) is to optimize the potential that generates the orbitals rather than the orbitals themselves. The equation determining the orbitals is obtained by minimizing the fluctuations in the local energy. This method corresponds to energy minimization and is limited in the sense that the orbitals are constrained to be the lowest energy eigenstates of a one-electron Hamiltonian. However, it has the significant advantage that the number of parameters in the potential does not increase as rapidly with system size as the number of parameters in the single-particle orbitals.

VIII. PSEUDOPOTENTIALS

A. The need for pseudopotentials

QMC methods have been applied very successfully to first-row atoms, but the computational effort increases rapidly with the atomic number Z . Various estimates of the scaling of the computational cost with Z have been made, with results ranging from $Z^{5.5}$ (Ceperley, 1986) to $Z^{6.5}$ (Hammond, Reynolds, and Lester, 1987). This scaling effectively rules out applications to heavy atoms. The presence of core electrons causes two related problems. Because of the shorter length scales associated with variations in the wave function near a nucleus of large Z the time step should be decreased. This problem can be significantly reduced by the use of acceleration schemes, such as those devised by Umrigar (1993) and Stedman, Foulkes, and Nekovee (1998). The other 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. Although these fluctuations can be reduced by a judicious choice of trial wave function, in practice they are large for heavier atoms.

However, many properties of interest, including the interatomic bonding and the low energy excitations, are determined by the behavior of the valence electrons, and just as in other electronic structure techniques one can use pseudopotentials to remove the core electrons from

the problem. This approach serves to reduce the effective value of Z . Errors are inevitably introduced, but the gain in computational efficiency is very large and is sufficient to make applications to heavy atoms feasible.

B. Non-local pseudopotentials

The ideas behind pseudopotentials are perhaps best explained in the context of independent-electron schemes such as the Hartree-Fock (HF) method and Hohenberg-Kohn-Sham density-functional theory (DFT). In both these theories the wave function consists of a determinant of orbitals, which can be partitioned into core and valence orbitals on chemical grounds. For example, the ground state configuration of the silicon atom is $[1s^2 2s^2 2p^6 3s^2 3p^2]$. The $3s$ and $3p$ orbitals take part in the chemical bonding and are designated as valence orbitals, while the $1s$, $2s$, and $2p$ orbitals largely retain their atomic identities and are designated as core orbitals. The idea is to create an effective potential (the pseudopotential) 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. For silicon the lowest energy states of the pseudopotential with angular momenta $l = 0, 1, 2$ are the pseudo- $3s$, $3p$, and $3d$ orbitals, respectively.

It is conventional to divide the pseudopotential $V_l^{\text{ps}}(r)$ for any given 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 angular momentum l . The final result is an effective interaction potential between a valence electron at \mathbf{r} and the (pseudo) ion at the origin. The electron-ion potential energy term in the full many-electron Hamiltonian of the atom then takes the form

$$V_{\text{loc}}(\mathbf{R}) + \hat{V}_{\text{nl}} = \sum_i V_{\text{loc}}^{\text{ps}}(r_i) + \sum_i \hat{V}_{\text{nl},i}^{\text{ps}}, \quad (8.1)$$

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

$$\hat{V}_{\text{nl},i}^{\text{ps}} f(\mathbf{r}_i) = \sum_{l,m} V_{\text{nl},l}^{\text{ps}}(r_i) Y_{lm}(\Omega_i) \int_{4\pi} Y_{lm}^*(\Omega'_i) f(\mathbf{r}'_i) d\Omega'_i. \quad (8.2)$$

The angular integrals pick out the different angular momentum components (s, p, d, \dots) of the function $f(\mathbf{r}_i)$, and so guarantee that each symmetry channel (s, p, d, \dots) “feels” its own potential $V_{\text{nl},l}^{\text{ps}}(r)$.

The first norm-conserving LDA pseudopotentials were generated by Hamann, Schlüter, and Chiang (1979). In a similar development, Christiansen, Lee, and Pitzer (1979) generated HF pseudopotentials, which are called

effective core potentials. Both of these sets of pseudopotentials have been used very successfully in numerous applications, including QMC simulations.

Within independent-electron theories such as DFT and HF the distinction between core and valence electrons (the “core-valence partition”) is exact and is achieved by simply designating some orbitals as core orbitals and others as valence orbitals. In a many-electron framework, however, the electrons are identical particles and the core-valence partition cannot be exact. The use of pseudopotentials in many-electron calculations therefore involves a further approximation. The accurate results obtained in pseudopotential QMC simulations show that this extra approximation is usually of minor importance.

C. Core polarization potentials

Correlation between electrons within a core has an indirect effect upon the valence electrons, which is usually unimportant. The neglect of core-valence correlation can, however, be significant whenever the number of valence electrons is small and the core is large; the core then exhibits core polarization and relaxation effects as a response to changes in the valence environment. Examples of systems with large core polarization effects include alkali metals with one or two valence electrons.

One can include some of the effects of core-valence correlation within a pseudopotential framework by introducing a Core Polarization Potential (CPP). This can be obtained within a core-valence partition scheme using the theory of Callaway (1957), although in practice empirical forms have been employed. Callaway (1957) showed that, after a series of approximations, the CPP could be approximated by a dipole term, which represents the polarization of the core due to the electric field of the valence electrons (and other ions, if present). It is consistent to use a CPP in conjunction with a HF pseudopotential.

Rather than calculate the CPP from first principles, it has normally been described by a simple analytic form. CPPs have been generated for a wide range of elements using several different methods (Müller and Meyer, 1984; Müller, Flesch, and Meyer, 1984; Shirley and Martin, 1993), and it has been shown (Shirley and Martin, 1993) that the combination of HF pseudopotentials and CPPs works very well for the first two rows of the periodic table. The use of CPP pseudopotentials within QMC methods was tested in the case of the sodium dimer by Shirley, Mitas, and Martin (1991), and the results were significantly better than those obtained using pseudopotentials that neglected core polarization effects.

D. Pseudopotentials in variational Monte Carlo

In this subsection we describe the technical aspects of using non-local pseudopotentials in VMC. The first pseudopotential QMC calculations made the approximation that the non-local angular momentum projection operators acted only on the determinantal part of the wave function (Hurley and Christiansen, 1987; Hammond, Reynolds, and Lester, 1987; Christiansen, 1988). Fahy, Wang, and Louie (1988, 1990a) formulated and applied the VMC method with non-local pseudopotentials acting on the full correlated wave function.

The action of the non-local pseudopotential on the wave function can be written as a sum of contributions from each electron and each angular momentum channel. The contribution to the local energy $E_L = \Psi_T^{-1} \hat{H} \Psi_T$ made by the non-local pseudopotential terms is

$$\begin{aligned} V_{\text{nl}} &= \Psi_T^{-1} \hat{V}_{\text{nl}} \Psi_T \\ &= \sum_i \Psi_T^{-1} \hat{V}_{\text{nl},i}^{\text{ps}} \Psi_T = \sum_i V_{\text{nl},i}, \end{aligned} \quad (8.3)$$

where for simplicity we consider the case of a single atom placed at the origin. Using Eq. (8.2) we can write the non-local contribution of electron i to the local energy as

$$\begin{aligned} V_{\text{nl},i} &= \sum_l V_{\text{nl},l}^{\text{ps}}(r_i) \sum_{m=-l}^l Y_{lm}(\Omega_{\mathbf{r}_i}) \int Y_{lm}^*(\Omega_{\mathbf{r}'_i}) \\ &\times \frac{\Psi_T(\mathbf{r}_1, \dots, \mathbf{r}_{i-1}, \mathbf{r}'_i, \mathbf{r}_{i+1}, \dots, \mathbf{r}_N)}{\Psi_T(\mathbf{r}_1, \dots, \mathbf{r}_{i-1}, \mathbf{r}_i, \mathbf{r}_{i+1}, \dots, \mathbf{r}_N)} d\Omega_{\mathbf{r}'_i}, \end{aligned} \quad (8.4)$$

where the angular integration is over the sphere passing through the i 'th electron and centered on the origin. Equation (8.4) 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

$$\begin{aligned} V_{\text{nl},i} &= \sum_l V_{\text{nl},l}^{\text{ps}}(r_i) \frac{2l+1}{4\pi} \int P_l[\cos(\theta'_i)] \\ &\times \frac{\Psi_T(\mathbf{r}_1, \dots, \mathbf{r}_{i-1}, \mathbf{r}'_i, \mathbf{r}_{i+1}, \dots, \mathbf{r}_N)}{\Psi_T(\mathbf{r}_1, \dots, \mathbf{r}_{i-1}, \mathbf{r}_i, \mathbf{r}_{i+1}, \dots, \mathbf{r}_N)} d\Omega_{\mathbf{r}'_i}, \end{aligned} \quad (8.5)$$

where P_l denotes a Legendre polynomial.

The integral over the surface of the sphere in Eq. (8.5) must be evaluated numerically. The \mathbf{r}' dependence of the many-body wave function is expected to have predominantly the angular momentum character of the orbitals in the determinantal part of the wave function. A suitable integration scheme is therefore to use a quadrature rule that integrates products of spherical harmonics exactly up to some maximum value l_{max} . Appropriate values of l_{max} may be deduced on chemical grounds, but one must remember that even in an atom the Jastrow factor introduces higher angular momentum components than occur in the determinantal part of the wave function. To

avoid bias the orientation of the axes is chosen randomly each time such an integral is evaluated. Mitas, Shirley, and Ceperley (1991) have tested a number of quadrature grids for the Si atom and the Cu^+ ion. Taking Si as an example, a grid containing 6 points with $l_{\text{max}} = 3$ is sufficient for reasonable accuracy, but considerably higher accuracy can be obtained using a grid containing 12 points with $l_{\text{max}} = 5$. In earlier work (Fahy, Wang, and Louie, 1988, 1990a) it was suggested that each non-local integral should be evaluated by sampling the quadrature grid several times using randomly oriented axes, but it has since been established that it is more efficient to sample a higher order rule once (Mitas, Shirley, and Ceperley, 1991). Within a VMC calculation it is often possible to use a low-order quadrature rule because the error cancels over the run, but higher accuracy is required for wave function optimization and DMC calculations, which are biased by errors in the non-local integration.

In principle the non-local energy should be summed over all the ionic cores and all electrons in the system. However, since the non-local potential of each ion is short ranged, one need only sum over the few atoms nearest to each electron. The additional angular integrations are computationally inexpensive within VMC because the local energy is not required during the one-electron moves and need not be evaluated frequently. Indeed, if the local energy is evaluated too frequently the values obtained will be correlated and little or no benefit will accrue.

E. Pseudopotentials in diffusion Monte Carlo

The use of non-local pseudopotentials in DMC is more problematic. If the Hamiltonian contains the non-local operator $\hat{V}_{\text{nl}} = \sum_i \hat{V}_{\text{nl},i}^{\text{ps}}$ the approximate propagator includes matrix elements of the form $\langle \mathbf{R} | \exp(-\tau \hat{V}_{\text{nl}}) | \mathbf{R}' \rangle$, which are *not* guaranteed to be non-negative for arbitrary \mathbf{R}' , \mathbf{R} , and τ . Consequently, as the population of walkers evolves according to the imaginary-time Schrödinger equation,

$$\begin{aligned} \partial_t f &= \frac{1}{2} \nabla^2 f - \nabla \cdot (\mathbf{v}_D f) - \frac{(\hat{H} - E_T) \Psi_T}{\Psi_T} f \\ &+ \left\{ \frac{\hat{V}_{\text{nl}} \Psi_T}{\Psi_T} - \frac{\hat{V}_{\text{nl}} \Phi}{\Phi} \right\} f, \end{aligned} \quad (8.6)$$

the sign of a walker can change as the time evolves. After a few time steps the sign becomes random and we encounter a sign problem analogous to the fermion sign problem discussed in Sec. III.D.2 and just as serious.

To circumvent this difficulty the so-called pseudopotential *localization approximation* has been introduced: the term in curly brackets is neglected, thus making Eq. (8.6) formally equivalent to an imaginary-time Schrödinger equation with local potentials. If $\Psi_T \approx \Psi_0$

the error introduced by this approximation is small and is proportional to $(\Psi_T - \Psi_0)^2$ (Mitas, Shirley, and Ceperley, 1991). Because the localization approximation depends on the accuracy of Ψ_T , it is necessary to use accurate trial functions to ensure that the calculation is in the low variance/quadratic convergence regime. Fortunately, as was explained in Sec. IV, accurate trial functions are available for simple solids. It is not easy to establish the exact size of the error introduced by the localization approximation since no better calculational technique is available for solids, but comparisons with experiments and other calculations for small systems show that in most cases it is smaller than the fixed-node error.

The crucial speed-up resulting from the use of pseudopotentials is clearly demonstrated by the example of the iron atom (Mitas, 1994). Table IV give values of the total energy; the variance of the local energy, $\sigma_E^2 = \langle \Psi_T | [E_L(\mathbf{R}) - E_V]^2 | \Psi_T \rangle / \langle \Psi_T | \Psi_T \rangle$; the energy autocorrelation time, κ , divided by the all-electron value, κ_{all} (values of the local energy calculated at closely separated times are statistically correlated; κ is a measure of how long the interval between two energy measurements must be to ensure that these correlations are negligible); and, finally, the efficiency, which is proportional to $1/(\kappa\sigma_E^2)$. The results were obtained using VMC but a comparison of DMC results would be qualitatively similar. It is evident that the efficiency improves dramatically as the size of the core is increased, but that the systematic errors introduced by the pseudopotentials also increase. In the case of the Fe atom, if we seek an accuracy of ~ 0.1 eV, the best compromise is the Ne core. This comparison provides a quantitative example but should not be taken as definitive: to some extent one can change the quantities shown in Table IV through improvement of the trial function, more efficient sampling, and more efficient coding.

TABLE IV. VMC of the Fe atom with decreasing number of valence electrons.

| | all-electron | Ne core | Ar core |
|------------------------------------|--------------|---------------|----------------|
| $E_{\text{HF}}(a.u.)$ | -1262.444 | -123.114 | -21.387 |
| $E_{\text{VMC}}(a.u.)$ | -1263.20(2) | -123.708(2) | -21.660(1) |
| σ_E^2 | ≈ 50 | 1.54 | 0.16 |
| $\kappa/\kappa_{\text{all}}$ | 1 | ≈ 0.3 | ≈ 0.05 |
| efficiency= $1/(\kappa\sigma_E^2)$ | 0.02 | 2.1 | 125. |
| valence errors (eV) | 0. | ≈ 0.1 | ≈ 0.5 |

Very recently, a new algorithm was suggested that produces an upper bound on the energy even when the Hamiltonian contains non-local operators. This new approach combines the localization approximation with non-local operator sampling (ten Haaf, van Bemmelen, van Leeuwen, van Saarloos, and Ceperley, 1994), and has proved very useful in lattice models. Unfortunately there is a price to be paid: the algorithm has no zero variance

property (Ceperley and Mitas, 1996). This might be an issue for calculations of realistic systems using very accurate wave functions, when one is usually working in the low variance regime.

F. Alternatives to non-local pseudopotentials

The use of non-local pseudopotentials in DMC is both problematic and computationally demanding. This has motivated the development of alternative approaches, the most important of which are the pseudo-Hamiltonian method of Bachelet, Ceperley, and Chiochetti (1989) and the damped-core method of Hammond, Reynolds, and Lester (1988).

In the pseudo-Hamiltonian approach the core is removed and the local potential modified in the normal way. However, instead of introducing non-local potentials, the kinetic energy operator (Laplacian) is replaced by a general second-order differential operator with a position-dependent electron effective mass tensor. The fundamental advantage of this construction is that the Hamiltonian remains *local*. The pseudo-Hamiltonian approach is easily incorporated in both VMC and DMC, but unfortunately it has inherent limitations (Foulkes and Schlüter, 1990) connected with the required positive definiteness of the effective mass tensor. The consequence is that pseudo-Hamiltonians can be constructed only for a restricted set of elements in which the non-local effects are not very strong. Nevertheless, successful simulations employing pseudo-Hamiltonians have been carried out (Li, Ceperley, and Martin, 1991).

In the damped-core technique (Hammond, Reynolds, and Lester, 1988) the space around an atom is divided into core (inner) and valence (outer) regions. All the core and valence electrons are included in the simulation, but electrons in the core regions are moved according to the VMC algorithm while electrons in the valence regions are moved using DMC. When an electron moves from one region to the other it experiences a smooth transition between these two regimes. Although very good results were obtained for atoms (Hammond, Reynolds, and Lester, 1988), this technique is difficult to apply to large systems. The reason is that the key problem, namely the impact of the large energy fluctuations from the core states, is diminished but not eliminated.

There is clearly a need for further development of many-body pseudopotentials for use in QMC calculations. A step in this direction was taken by Acioli and Ceperley (1994), who showed that the quality of a many-body pseudopotential depends on an accurate description of the one-electron, two-electron, etc., density matrices outside the core region. The most important quantity is the one-electron density matrix, the eigenfunctions of which are the natural orbitals. One therefore has to

ensure that the natural orbitals of the pseudo-atom are correct outside the core region. This work represents an advance in understanding, but has not so far been developed into a practical scheme for generating accurate pseudopotentials for heavy atoms.

IX. PERIODIC BOUNDARY CONDITIONS AND FINITE-SIZE ERRORS

A. Introduction

Current algorithms and computational resources permit QMC calculations on systems with up to about 1000 electrons. Although finite clusters of this size are very unlike bulk material, periodic systems containing a few hundred electrons are large enough to model many phenomena in condensed matter with good accuracy. The limited system size inevitably introduces finite-size errors, which decrease as the number of electrons in the periodically repeated region increases, so QMC simulation cells for perfect crystals usually consist of several primitive unit cells. One can also model non-periodic systems within this framework by using the supercell approach, which has proved very successful within independent-particle methods.

B. Bloch's theorem for many-body systems

Bloch's theorem is one of the cornerstones of the independent-particle electronic structure theory of solids (see, e.g., Ashcroft and Mermin, 1976). It defines the crystal momentum used to label the eigenstates in a periodic solid, and allows the independent-particle problem to be solved by considering only a single primitive unit cell. The reduction of the problem to one within the primitive cell is only possible in independent-particle theories; in many-electron QMC simulations Bloch's theorem does not obviate the need to solve the Schrödinger equation over the entire simulation cell. However, Bloch wave functions have the property that the local energy $\Psi_T^{-1} \hat{H} \Psi_T$ has the full translational symmetry of the Hamiltonian. Bloch functions are normally complex and so cannot be used in fixed-node DMC calculations. If the Hamiltonian has time reversal symmetry, however, the real part of a complex Bloch eigenfunction is also an eigenfunction (although not in general a Bloch function) and so this is not a serious problem. The Bloch functions themselves can always be chosen real if the Bloch wave vector is equal to half a reciprocal lattice vector, and the use of such trial functions is advantageous because the corresponding fixed-node DMC energies are guaranteed to be greater than or equal to the energy of the lowest exact eigenstate with that wave vector (Foulkes, Hood,

and Needs, 1999). A real trial function also implies a real local energy at every point in configuration space.

Bloch's theorem arises from the translational symmetry of the many-body Hamiltonian, which may be written schematically as

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 + \sum_{i=1}^N V_{\text{ion}}(\mathbf{r}_i) + \frac{1}{2} \sum_{\mathbf{R}_s} \sum_{i=1}^N \sum_{j=1}^N{}' \frac{1}{|\mathbf{r}_i - \mathbf{r}_j - \mathbf{R}_s|}, \quad (9.1)$$

where $\{\mathbf{R}_s\}$ is the set of translation vectors of the simulation-cell lattice, the ionic potential $V_{\text{ion}}(\mathbf{r})$ has the periodicity of $\{\mathbf{R}_s\}$, N is the number of electrons in the simulation cell, and the prime on the j summation indicates that the $j = i$ term is omitted when $\mathbf{R}_s = \mathbf{0}$. The electron-electron term includes interactions with the electron “images”, which are copies of the electrons displaced by simulation-cell lattice vectors \mathbf{R}_s ; these images model the potential due to the electrons “outside” the simulation cell. The ionic potential includes contributions from all the ions within the simulation cell and their images. Strictly, the potentials in Eq. (9.1) are divergent and additional conditions have to be imposed to make them well defined as discussed in Sec. IX.C. At the moment, however, we are only interested in the translational symmetry, for which Eq. (9.1) will suffice.

The invariance of \hat{H} under the translation of any electron by a vector in $\{\mathbf{R}_s\}$ leads to the many-body Bloch condition (Rajagopal, Needs, Kenny *et al.*, 1994; Rajagopal *et al.*, 1995):

$$\Psi_{\mathbf{k}_s}(\{\mathbf{r}_i\}) = U_{\mathbf{k}_s}(\{\mathbf{r}_i\}) \exp\left(i\mathbf{k}_s \cdot \sum_{i=1}^N \mathbf{r}_i\right), \quad (9.2)$$

where the simulation-cell periodic part, $U_{\mathbf{k}_s}$, is invariant under the translation of any electron position by a vector in $\{\mathbf{R}_s\}$ and is antisymmetric under particle exchange. (We have omitted the band index in Eq. (9.2) and elsewhere in this section.) The “simulation-cell wave vector” \mathbf{k}_s may always be reduced into the first Brillouin Zone (BZ) of the simulation-cell reciprocal lattice. If the simulation cell has translation vectors $N_1\mathbf{a}_1$, $N_2\mathbf{a}_2$, and $N_3\mathbf{a}_3$, where the N_i are integers and the \mathbf{a}_i are the primitive translation vectors of the underlying crystal lattice, the first BZ of the primitive lattice contains a grid of $N_1N_2N_3$ points with the same (reduced) value of \mathbf{k}_s . This grid becomes finer as the simulation cell is made larger, and is analogous to the k-point sampling grids used in independent-electron calculations. Note that the symmetries leading to the existence of the good quantum number \mathbf{k}_s arise from the artificial periodicity of the simulation-cell Hamiltonian; Eq. (9.2) does not hold in a real solid.

As long as N_1 , N_2 , and N_3 are not all equal to 1, the simulation-cell Hamiltonian has more translational symmetry than assumed so far, since it is also unchanged when all the electrons are *simultaneously* translated by any primitive lattice vector \mathbf{R}_p . This additional symmetry is quite different from the translations of a single electron considered in the discussion of \mathbf{k}_s , and leads to a second Bloch condition (Rajagopal, Needs, Kenny *et al.*, 1994; Rajagopal *et al.*, 1995):

$$\Psi_{\mathbf{k}_p}(\{\mathbf{r}_i\}) = W_{\mathbf{k}_p}(\{\mathbf{r}_i\}) \exp\left(i\mathbf{k}_p \cdot \frac{1}{N} \sum_{i=1}^N \mathbf{r}_i\right), \quad (9.3)$$

where $W_{\mathbf{k}_p}$ is invariant under the simultaneous translation of all electrons by any primitive lattice vector \mathbf{R}_p and is antisymmetric under particle exchange. The “primitive wave vector” \mathbf{k}_p , which can be reduced into the first BZ of the primitive reciprocal lattice, gives the physical crystal momentum measured in experiments. The wave function can be chosen to satisfy the simulation-cell and primitive-cell Bloch conditions simultaneously, with \mathbf{k}_s and \mathbf{k}_p linked by the condition that the wave functions of Eqs. (9.2) and (9.3) must be identical if we translate all electrons by a simulation-cell reciprocal lattice vector (Rajagopal, Needs, Kenny *et al.*, 1994; Rajagopal *et al.*, 1995). However, as long as the simulation cell contains more than one primitive unit cell, the two wave vectors are distinct labels of the many-body Bloch wave function and both are required to specify the full translational symmetry.

C. Periodic boundary conditions and Coulomb interactions

The potential energy of a macroscopic but finite system containing charges (electrons and nuclei) q_i at positions ℓ_i may be written as

$$V = \frac{1}{2} \sum_i \sum_{j(\neq i)} \frac{q_i q_j}{|\ell_i - \ell_j|} = \frac{1}{2} \sum_i q_i V_i, \quad (9.4)$$

where the sums extend over all charges in the system and

$$V_i = \sum_{j(\neq i)} \frac{q_j}{|\ell_i - \ell_j|} \quad (9.5)$$

is the potential at ℓ_i due to every charge except q_i . The potential energy of the QMC simulation cell is defined by analogy with Eq. (9.4):

$$V(\ell_1, \ell_2, \dots, \ell_N) = \frac{1}{2} \sum_{i=1}^N q_i V_i, \quad (9.6)$$

where N is the number of charges in the simulation cell and V_i is the potential felt by charge q_i at position ℓ_i

in an infinite periodic lattice of identical copies of the simulation cell. Thus

$$V_i = \sum_{\mathbf{R}_s} \sum_{j=1}^{N'} \frac{q_j}{|\ell_i - (\ell_j + \mathbf{R}_s)|}, \quad (9.7)$$

where the prime on the j summation indicates that the $j = i$ term is omitted when $\mathbf{R}_s = \mathbf{0}$. The potential V_i only depends on the positions of the few hundred charges in the simulation cell and is at best a poor imitation of the potential of a real solid, which depends on the positions of approximately 10^{23} electrons and ions.

Note that Eq. (9.7) includes the interactions of q_i with its own images; if it did not, the copies of the simulation cell in the infinite periodic lattice would have a net charge and the summation would be divergent. Unfortunately, even when the self-interactions are included, the sum in Eq. (9.7) is not absolutely convergent. Consider, for example, the sum of contributions from simulation cells lying entirely within a large but finite sphere. For almost every configuration sampled in a QMC simulation, the simulation cell will have a non-zero dipole moment, and hence, from the viewpoint of macroscopic electrostatics, the surfaces of the sphere will be covered with polarization charges. The resulting depolarization fields are well known to affect the total energy per unit cell (see, e.g., Kittel, 1966, Chapter 12), even in the limit as the sphere size goes to infinity. If, instead of spheres, we had considered larger and larger clusters of a different shape (perhaps cubes or ellipsoids), the depolarization fields would have been different and we would have obtained a different potential energy.

Since the depolarization potential contribution to V_i is not periodic, it is convenient to set the depolarization field to zero. This can be thought of as surrounding the sphere by a conducting medium and is sometimes known as applying tin foil boundary conditions. The solution of Poisson’s equation within the sphere with zero depolarization field is unique to within a constant, which does not affect the Coulomb energy of Eq. (9.6) because the simulation cell has no net charge. This recipe therefore produces an unambiguous result, called the Ewald energy. For a full mathematical discussion of the difficulties associated with defining Coulomb potentials in periodic systems see De Leeuw, Perram, and Smith (1980), while a simpler treatment is given by Fraser *et al.* (1996).

The Ewald energy may be calculated in a number of different ways (see Sec. X.D), but QMC simulations of solids generally use the original Ewald method (Ewald, 1921; Tosi, 1964; De Leeuw, Perram, and Smith, 1980) or one of its variants (Rajagopal and Needs, 1994; Natoli and Ceperley, 1995). The Ewald energy is written in the form,

$$V(\ell_1, \dots, \ell_N) = \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ (j \neq i)}}^N q_i q_j V^E(\ell_i - \ell_j) + \frac{1}{2} \sum_{i=1}^N q_i^2 \xi, \quad (9.8)$$

where $V^E(\mathbf{r} - \ell_j)$ is the potential at \mathbf{r} due to a periodic lattice of unit charges at positions $\ell_j + \mathbf{R}_s$ plus a cancelling uniform background, and

$$\xi = \lim_{\mathbf{r} \rightarrow \ell_i} \left(V^E(\mathbf{r} - \ell_i) - \frac{1}{|\mathbf{r} - \ell_i|} \right) \quad (9.9)$$

is the potential at ℓ_i due to its own images and background. The Ewald interaction V^E is calculated using the expression:

$$V^E(\mathbf{r} - \ell_i) = \sum_{\mathbf{R}_s} \frac{\text{erfc}(\kappa|\mathbf{r} - (\ell_i + \mathbf{R}_s)|)}{|\mathbf{r} - (\ell_i + \mathbf{R}_s)|} - \frac{\pi}{\kappa^2 \Omega_s} + \frac{4\pi}{\Omega_s} \sum_{\mathbf{G}_s \neq 0} \frac{\exp(-G_s^2/4\kappa^2)}{G_s^2} e^{i\mathbf{G}_s \cdot (\mathbf{r} - \ell_i)}, \quad (9.10)$$

where Ω_s is the volume of the simulation cell, $\{\mathbf{G}_s\}$ is the set of vectors reciprocal to $\{\mathbf{R}_s\}$, and κ is an adjustable parameter that does not affect the value of V^E and may be chosen to optimize the convergence rates of the real and reciprocal space sums.

Depolarization fields in real solids are almost always negligibly small, either because the solid is unpolarized or because the polarization charges are screened by charged dust particles adhering to the surfaces. The Ewald method therefore gives the correct Hartree and electron energies in almost all cases. (These energies are the same as are obtained using the k-space approach familiar from band structure calculations.) The electrons in real solids are in different places in every unit cell, however, and so the correlations between electrons are not the same as in a periodic array of charges. The spurious correlations built into the Ewald interaction produce large finite-size errors in the exchange-correlation energy, the treatment of which is discussed in the next section.

D. Coping with finite-size errors

1. Introduction

Using a finite simulation cell to model an extended system introduces finite-size errors. These are one of the major problems encountered in the application of accurate many-body techniques to extended systems, and it is vital that they are treated properly.

Finite-size errors also occur in independent-particle calculations, but their nature is subtly different from the many-body case. Within the supercell approach, a point

defect in an otherwise perfect crystal is modelled by taking a simulation cell containing a single point defect and repeating the cell periodically throughout space. If the simulation cell is too small the interactions between the defects in neighboring simulation cells produce a significant finite-size error. This type of finite-size error is common to both independent-particle and many-body calculations. A second source of error within independent-particle calculations arises from inaccuracies in the integration over the Brillouin Zone (BZ). This error can also be thought of as type of finite-size effect. Most many-body simulations for periodic systems use only one k-point in the BZ of the simulation cell (i.e., one value of \mathbf{k}_s), and therefore they suffer from significant finite-size effects of this type.

There is, however, an additional important contribution to the finite-size errors in many-body calculations of solids. The periodic Ewald interaction used to model the electron-electron interactions depends on the size and shape of the simulation cell, and this produces errors that are not encountered in density-functional calculations. Hartree-Fock (HF) theory also requires an explicit electron-electron interaction, and using the Ewald form produces an analogous finite-size error in the exchange term (but not the Hartree term). In DFT, however, the exchange-correlation energy is obtained from a formula derived from DMC data that already contains an extrapolation to infinite simulation-cell size (Ceperley and Alder, 1980). DFT calculations for periodic systems therefore do not contain a finite-size error arising from the Ewald interaction.

2. Finite-size correction and extrapolation formulae

What practical schemes are available for coping with finite-size errors in QMC or other many-body calculations? One idea is to use a finite-size correction formula in which the energy for the infinite system is written as

$$E_\infty = E_N + (E_\infty - E_N), \quad (9.11)$$

where the subscript denotes the system size. A QMC calculation for the N -particle system is then performed to obtain an accurate value for E_N , and the correction term in brackets is approximated using a much less expensive scheme that can be applied to very large systems. An alternative is to carry out QMC calculations for a range of different system sizes, fit the results to some chosen function of N , and attempt to extrapolate to infinite system size. Correction and extrapolation procedures can be combined to give

$$E_\infty \simeq E_N + (E'_\infty - E'_N) + F(N), \quad (9.12)$$

where the prime indicates that a less expensive scheme is used and $F(N)$ is an extrapolation function. The optimal

form of $F(N)$ depends on the method used to calculate the correction term. Extrapolation is costly because it involves more calculations, and is prone to inaccuracy because one has to perform a fit with only a few noisy data points. In designing a correction/extrapolation procedure one therefore tries to make the extrapolation term as small as possible.

Ceperley and coworkers (Ceperley, 1978; Ceperley and Alder, 1987; Tanatar and Ceperley, 1989; Kwon, Ceperley, and Martin, 1998) have successfully used an extrapolation technique, fitting to the formula

$$E_{\text{DMC},\infty} \simeq E_{\text{DMC},N} + a(T_{\infty} - T_N) + \frac{b}{N}, \quad (9.13)$$

where a and b are parameters and T is the kinetic energy of the non-interacting electron gas. The b/N term accounts for the finite-size effects arising from the interaction energy, and the difference of the parameter a from unity accounts for the difference between the kinetic energies of the interacting and non-interacting systems. In general, the non-interacting gas correction can be replaced by an independent-particle LDA or HF correction.

Fahy, Wang, and Louie (1990a) have used a finite-size correction procedure in which the correction term is evaluated within the LDA using an accurate BZ integration; however, this method does not include a correction for the interaction energy. A related idea is to average QMC results over different values of the simulation-cell wave vector \mathbf{k}_s , which amounts to averaging over the boundary conditions on the simulation cell. This method has been used for lattice Hamiltonians (see, for example, Valentí *et al.*, 1991, and Gros, 1992), and was used in the VMC calculations of Kralik, Delaney, and Louie (1998) to evaluate the Compton profile of Si (see Sec. V.F). Averaging over \mathbf{k}_s has the advantage that results from more approximate methods, such as the LDA, are not required, but it is expensive and does not include a correction for the interaction energy.

In practice, the BZ integration and interaction errors appear to be almost independent and can be corrected for separately (Kent, Hood *et al.*, 1999). The BZ integration errors are usually well accounted for by corrections derived from LDA results, but these corrections do not remove the finite-size errors in the interaction energy. Recently, Fraser *et al.* (1996), Williamson *et al.* (1997), and Kent, Hood *et al.* (1999) have shown how the finite-size errors in the interaction energy may be reduced by replacing the Ewald interaction by a “model periodic Coulomb” interaction. This technique is very effective, reducing or even eliminating the need for extrapolation. The model periodic Coulomb interaction is described in Sec. IX.D.4.

3. Choosing the simulation-cell wave vector

The computational cost of an LDA calculation is proportional to the number of k-points sampled within the primitive BZ. In a QMC calculation, however, the volume of the simulation cell is proportional to the number of k-points sampled within the primitive BZ, and the computational cost increases approximately as the cube of the volume of the simulation cell. This makes it much more difficult to reduce the BZ integration errors by improving the k-point sampling. The problem may be ameliorated by applying corrections calculated using a simpler method such as the LDA, but one nevertheless expects to obtain more accurate answers if the required corrections are as small as possible. Since the BZ integration errors for a given simulation cell depend quite sensitively on the choice of the simulation-cell wave vector \mathbf{k}_s , it is important to choose \mathbf{k}_s carefully.

The problem of choosing \mathbf{k}_s is entirely analogous to that of choosing “special k-points” (Baldereschi, 1973; Monkhorst and Pack, 1976) in band structure calculations, and ideas from that field can be taken over directly into many-body calculations (Rajagopal, Needs, Kenny *et al.*, 1994; Rajagopal *et al.*, 1995; Kent, Hood *et al.*, 1999). The theory of special k-points was first developed by Baldereschi (1973), who defined the “mean-value point”, which is a k-point at which smooth periodic functions of the wave vector accurately approximate their averages over the BZ. However, since the Baldereschi mean-value point is not equal to half a reciprocal lattice vector, one cannot construct real Bloch wave functions at that point. A more convenient and very successful approach is to choose \mathbf{k}_s from among those points that allow real Bloch wave functions according to the symmetrized plane-wave test of BZ integration quality introduced by Baldereschi (1973). The relationship of this approach to the widely used Monkhorst-Pack scheme of BZ integration has been discussed by Kent, Hood *et al.* (1999).

4. Interaction energy finite-size effects

Where does the large finite-size effect in the Ewald interaction energy come from? Since the LDA corrections to the QMC energies are effective in removing the finite-size errors in the non-interacting kinetic energy, the interaction of the electrons with the external potential, and the Hartree energy, the remaining finite-size errors must arise from the exchange-correlation (XC) energy. This can be written as the energy of interaction of the electrons with their own XC holes, which are normally short ranged and quite insensitive to the system size. The large finite-size effect must therefore arise from the dependence of the interaction itself on the system size.

This conclusion can be rationalized by the following analysis. Expanding the Ewald interaction around zero

separation gives

$$V^E(\mathbf{r}) = \frac{1}{r} + c + \frac{2\pi}{3\Omega_s} \mathbf{r}^T \cdot \mathbf{D} \cdot \mathbf{r} + \mathcal{O}\left(\frac{r^4}{\Omega_s^{5/3}}\right), \quad (9.14)$$

where c is a constant, Ω_s is the volume of the simulation cell, and the tensor \mathbf{D} depends on the shape of the simulation cell. (For a cell with cubic symmetry $\mathbf{D}=\mathbf{I}$.) The value of the constant c does not affect the total energy because the simulation cell is neutral. The other deviations from $1/r$ make the Ewald interaction periodic, but are also responsible for the spurious contribution to the XC energy (Fraser *et al.*, 1996; Williamson *et al.*, 1997). For cubic cells these terms are positive for small r and so the XC energy is more negative than it should be. Since the leading correction is proportional to the inverse of the simulation-cell volume, the error per electron is inversely proportional to the number of electrons in the cell. Detailed calculations have confirmed this picture (Fraser *et al.*, 1996; Williamson *et al.*, 1997; Kent, Hood *et al.*, 1999).

Clearly it would be desirable to remove this spurious contribution to the XC energy, but we must remember that the Hartree energy is given correctly by the Ewald interaction. The key requirements for a model Coulomb interaction giving small interaction-energy finite-size effects in simulations using periodic boundary conditions are therefore: (i) it should give the Ewald interaction for the Hartree terms, and (ii) it should be exactly $1/r$ for the interaction with the XC hole. Unfortunately, the only periodic solution of Poisson's equation for a periodic array of charges is the Ewald interaction, which obeys criterion (ii) only in the limit of an infinitely large simulation cell.

This problem was solved by Williamson *et al.* (1997) by abandoning the Ewald interaction and instead using a model periodic Coulomb (MPC) interaction that satisfies both criteria,

$$\hat{H}_{e-e} = \sum_{i>j} f(\mathbf{r}_i - \mathbf{r}_j) + \sum_i \int_{\text{WS}} [V^E(\mathbf{r}_i - \mathbf{r}) - f(\mathbf{r}_i - \mathbf{r})] \rho(\mathbf{r}) d\mathbf{r}, \quad (9.15)$$

where ρ is the electronic charge density and

$$f(\mathbf{r}) = \frac{1}{r_m}. \quad (9.16)$$

The definition of the cutoff Coulomb function f involves a minimum image convention whereby the inter-electron vector \mathbf{r} is reduced into the Wigner-Seitz (WS) cell of the simulation-cell lattice by removal of simulation-cell lattice vectors, leaving a vector \mathbf{r}_m . This ensures that \hat{H}_{e-e} has the correct translational and rotational symmetries. The second term of Eq. (9.15) is a mean-field-like one-electron potential that accounts for the difference between the Hartree energies calculated using the Ewald

and MPC interactions. As in the case of the Hartree potential in HF and DFT calculations, a double-counting correction has to be added to the expression for the total electron-electron interaction energy to prevent this term being counted twice. The electron-electron energy as calculated with the MPC interaction then consists of the sum of the Hartree energy calculated with the Ewald interaction and the XC energy calculated with the cut-off interaction f (Williamson *et al.*, 1997). The mean-field-like potential depends on the electron density, which could in principle be calculated self-consistently; in practice, however, the LDA (or HF) density is normally accurate enough to serve as a good approximation.

Tests using the MPC interaction have shown that it dramatically reduces the finite-size effects in the interaction energy (Fraser *et al.*, 1996; Williamson *et al.*, 1997; Kent, Hood *et al.*, 1999). Wave functions generated by variance minimization using the Ewald and MPC interactions are almost identical, and so properties other than the energy, such as pair-correlation functions, are hardly affected by the choice of interaction. As the MPC interaction gives the correct interaction between the electrons at short distances it may give a better account of, for example, the short distance behavior of the pair-correlation function, but this point has not yet been tested by calculations.

X. COMPUTATIONAL ISSUES

A. Representation of the single-particle orbitals

It is very important to use accurate single-particle orbitals to form the Slater determinants in the trial wave function. To this end we use sophisticated codes to perform the required DFT and HF calculations such as the molecular Gaussian orbital packages GAUSSIAN (Frisch *et al.*, 1995) and GAMESS (Schmidt *et al.*, 1993), the molecular and solid CRYSTAL package (Dovesi *et al.*, 1996), and various plane-wave pseudopotential packages. Within a QMC calculation itself the evaluation of the trial wave function and its derivatives can take up to half the computing time. It is therefore necessary to develop efficient methods to evaluate single-particle orbitals, Slater determinants, and Jastrow factors. A central issue is how to represent the single-particle orbitals. This question is logically separate from the basis used in the HF and DFT calculations, although re-expansion in a different basis set is undesirable because of the loss of accuracy. Because we require the single-particle orbitals at points in real space, the most natural procedure is to use a basis which is localized in real space, so that the cost of evaluating the single-particle orbitals will not increase rapidly with system size. Tabulating the orbitals on a grid is feasible for small systems, but for bigger systems the storage requirements are too large. Plane-wave

expansions are inefficient in most cases, but if there is a short repeat length in the problem they can still be the best solution. The best all round compromise appears to be to use an expansion in Gaussian orbitals, which are reasonably well localized in space and may be obtained from standard electronic structure packages.

B. Evaluation of the trial wave function

One of the main tasks in a QMC program is to evaluate the ratio $\Psi(\mathbf{R}^{\text{new}})/\Psi(\mathbf{R}^{\text{old}})$ needed to calculate the acceptance probability of a trial move from \mathbf{R}^{old} to \mathbf{R}^{new} . In our QMC calculations electrons are moved one at a time and therefore \mathbf{R}^{old} and \mathbf{R}^{new} differ in the position of one electron only.

Consider a Slater-Jastrow wave function of the form

$$\Psi(\mathbf{R}) = e^{J(\mathbf{R})} D^\uparrow(\mathbf{r}_1, \dots, \mathbf{r}_{N_\uparrow}) D^\downarrow(\mathbf{r}_{N_\uparrow+1}, \dots, \mathbf{r}_N), \quad (10.1)$$

where D^\uparrow and D^\downarrow are Slater determinants for the spin-up and spin-down electrons respectively, and

$$J(\mathbf{R}) = \sum_{i=1}^N \chi_{\sigma_i}(\mathbf{r}_i) - \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ (j \neq i)}}^N u_{\sigma_i, \sigma_j}(\mathbf{r}_i, \mathbf{r}_j). \quad (10.2)$$

How does this wave function change when the i 'th spin-up electron is moved from $\mathbf{r}_i^{\text{old}}$ to $\mathbf{r}_i^{\text{new}}$? Since the spin-down Slater determinant does not alter, the wave function ratio may be written as

$$\frac{\Psi(\mathbf{R}^{\text{new}})}{\Psi(\mathbf{R}^{\text{old}})} = q^\uparrow \frac{\exp \left[\chi_\uparrow(\mathbf{r}_i^{\text{new}}) - \sum_{j(\neq i)} u_{\uparrow, \sigma_j}(\mathbf{r}_i^{\text{new}}, \mathbf{r}_j) \right]}{\exp \left[\chi_\uparrow(\mathbf{r}_i^{\text{old}}) - \sum_{j(\neq i)} u_{\uparrow, \sigma_j}(\mathbf{r}_i^{\text{old}}, \mathbf{r}_j) \right]}, \quad (10.3)$$

where

$$q^\uparrow = \frac{D^\uparrow(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i^{\text{new}}, \dots, \mathbf{r}_N)}{D^\uparrow(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i^{\text{old}}, \dots, \mathbf{r}_N)}. \quad (10.4)$$

Assuming that the matrix of cofactors corresponding to the determinant D^\uparrow is available, there are efficient algorithms for evaluating q^\uparrow in a time proportional to N . If the trial move is accepted, the matrix of cofactors may then be updated in a time proportional to N^2 . These algorithms, which are well explained by Fahy, Wang, and Louie (1990a), make use of the Sherman-Morrison formula (see, e.g., Press *et al.*, 1992), and were first used in QMC simulations by Ceperley, Chester, and Kalos (1977). The contributions from the Jastrow factor may also be evaluated in a time proportional to N .

C. Evaluation of the local energy

Another important part of any Monte Carlo program is the evaluation of the local energy

$$E_L(\mathbf{R}) = \frac{\hat{H}\Psi(\mathbf{R})}{\Psi(\mathbf{R})} = \frac{\hat{T}\Psi(\mathbf{R})}{\Psi(\mathbf{R})} + \frac{\hat{V}_{\text{nl}}\Psi(\mathbf{R})}{\Psi(\mathbf{R})} + V_{\text{loc}}(\mathbf{R}), \quad (10.5)$$

where \hat{T} is the kinetic energy operator, $V_{\text{loc}}(\mathbf{R})$ is the local part of the potential energy, and \hat{V}_{nl} is the non-local part arising from the pseudopotential. In DMC simulations, in particular, the local energy determines the branching ratio and has to be evaluated after every one-electron trial move. This is a very time-consuming process and so efficient algorithms are needed.

The evaluation of $V_{\text{loc}}(\mathbf{R})$ is in principle straightforward (although some subtleties connected with the use of periodic boundary conditions were discussed in Sec. IX.C). Efficient algorithms for evaluating Coulomb potentials in periodic systems are discussed in Sec. X.D. The treatment of the non-local pseudopotential term was discussed in Secs. VIII.D and VIII.E.

The kinetic contribution to the local energy may be decomposed into contributions from each electron,

$$K = \sum_{i=1}^N K_i = \sum_{i=1}^N \Psi^{-1} \left(-\frac{1}{2} \nabla_i^2 \right) \Psi. \quad (10.6)$$

Because of the exponential form of the Jastrow factor, it is convenient to re-express K_i in terms of the logarithm of Ψ :

$$K_i = 2T_i - |\mathbf{F}_i|^2, \quad (10.7)$$

where

$$\mathbf{F}_i = \frac{1}{\sqrt{2}} \nabla_i (\ln |\Psi|) = \frac{1}{\sqrt{2}} \frac{\nabla_i \Psi}{\Psi}, \quad (10.8)$$

$$T_i = -\frac{1}{4} \nabla_i^2 (\ln |\Psi|) = -\frac{1}{4} \frac{\nabla_i^2 \Psi}{\Psi} + \frac{1}{4} \left(\frac{\nabla_i \Psi}{\Psi} \right)^2. \quad (10.9)$$

The three-dimensional vector \mathbf{F}_i is proportional to that part of the $3N$ -dimensional drift velocity vector \mathbf{v}_D associated with electron i . The drift velocity plays an important role in the DMC algorithm as explained in Sec. III.D. Integration by parts shows that

$$\langle K_i \rangle = \langle |\mathbf{F}_i|^2 \rangle = \langle T_i \rangle, \quad (10.10)$$

where the angular brackets denote averages over the probability density $|\Psi(\mathbf{R})|^2$. The kinetic energy may be evaluated using any of these three estimators, but $\langle K \rangle$ is the most useful because it has the lowest variance. Moreover, when $\langle K \rangle$ is added to the average of the potential energy, the variance of the sum tends to zero as the wave

function approaches an eigenstate. Eq. (10.10) holds for any trial function $\Psi(\mathbf{R})$ and provides a very useful consistency check of QMC programs. T_i and \mathbf{F}_i (and hence K_i and \mathbf{v}_D) may be evaluated efficiently using the cofactor methods mentioned in Sec. X.B.

D. Efficient treatment of the Coulomb interaction

As explained in Sec. IX.C, the Coulomb energy of the simulation cell is defined in terms of the Ewald potential $V^E(\mathbf{r})$, which is obtained by solving Poisson's equation subject to periodic boundary conditions. The Ewald potential may be calculated using the original Ewald method (Ewald, 1921; Tosi, 1964; De Leeuw, Perram, and Smith, 1980), which was invented long before the age of computers but is still in widespread use today. The idea is to write $V^E(\mathbf{r})$ as the sum of two separate series, one in real space and one in k-space, both of which can be made to converge fairly rapidly. In QMC simulations the electrons are moved one at a time and so it is necessary to calculate $V^E(\mathbf{r})$ at a point in space, not just the total potential energy of the cell. This makes other well known methods such as the particle-particle and particle-mesh schemes (Hockney and Eastwood, 1981) and Greengard's multipole algorithm (Greengard and Rokhlin, 1987) less effective. In addition, Greengard's method is restricted to Coulomb potentials, and, although $O(N)$, is inefficient unless the simulation cell contains at least several thousand particles. Consequently, the original Ewald method or variants thereof are normally used. We favor the efficient variants developed by Rajagopal and Needs (1994) and by Natoli and Ceperley (1995) based on an optimized division between the real- and k-space summations.

E. Scaling with the number of particles

The scaling laws discussed in Sec. X.B were for a move of a single electron. We now address the scaling of the *total* computational cost with N . Consider, for example, a Monte Carlo calculation of the total energy. The variance σ_{run}^2 of the calculated average energy is equal to σ^2/M , where M is the number of statistically independent samples of E_L and σ^2 is the variance of the sampled values. The total computer time T_{run} required by the simulation is MT_{sample} , where T_{sample} is the time to obtain a single statistically independent sample of E_L . Hence

$$T_{\text{run}} = \frac{\sigma^2 T_{\text{sample}}}{\sigma_{\text{run}}^2}. \quad (10.11)$$

The local energy may always be written as a sum of contributions from each electron, and in large enough systems these contributions are approximately independent. The value of σ^2 is therefore proportional to the number of electrons N . For T_{sample} we can write

$$T_{\text{sample}} \propto N^2 + \varepsilon N^3 \quad (10.12)$$

where the N^2 term comes from (i) evaluating all N single-particle orbitals and their derivatives at the N new electron positions (using localized basis functions such as Gaussians), (ii) evaluating the Jastrow factor, and (iii) evaluating the interaction energy. The N^3 contribution arises from the N updates of the matrix of cofactors, each of which takes a time proportional to N^2 . Because ε is of the order of 10^{-4} the quadratic term dominates in practice. Remembering the factor of N from σ^2 , the total computer time required to reach a given target variance σ_{run}^2 is therefore

$$T_{\text{run}} \propto N^3. \quad (10.13)$$

F. QMC on Parallel Computers

Monte Carlo calculations are inherently parallel in nature. At their most basic they involve the calculation of a large set of independent random numbers and then the averaging of a set of results produced by each of these random numbers. Coupled with the fact that QMC calculations, especially for solids, are relatively expensive to perform on today's workstations, they are ideal for parallel architecture machines, which offer two to three orders of magnitude more computational power. In fact, the major proportion of the work highlighted in Secs. V and VI was carried out on massively parallel computers, which are crucial for studying these computationally demanding problems.

It is important that any practitioner of electronic structure theory should acquire the ability to use parallel computers effectively. Simply speaking, a broad definition of a parallel computer (encompassing machines with hundreds/thousands of CPUs, networked workstations, multiprocessor workstations, etc.) is that it is a machine with a set of processors able to work cooperatively to solve a computational problem. Our QMC codes have been ported to both symmetric multi-processing and massively parallel processing platforms. Interprocessor communications involving sending packets of data between processors are carried out using the Message Passing Interface (MPI) standard (Snir *et al.*, 1998). Key issues for optimal performance include tuning code on a single node as well as optimizing the distribution of work between nodes. Important factors such as the proper arrangement of data to maximize memory access to registers from primary and secondary cache and main memory strongly influence the performance of programs, especially in large systems.

Central to any Monte Carlo computation is the efficient generation of random numbers. Although it is difficult to generate truly random numbers using a computer, many deterministic techniques exist for generating

sequences of so-called *pseudo-random* numbers (Knuth, 1997; Press *et al.*, 1992) that pass most of the obvious statistical tests for randomness. On parallel computers, the process of generating random numbers becomes more complicated because there are many concurrently executing tasks. Recently, Ceperley, Srinivasan, and Mascagni have developed a scalable library for pseudo-random number generation, called **SPRNG** (for Scalable Parallel Random Number Generators), which is freely available on the Web (URL: www.ncsa.uiuc.edu/Apps/SPRNG). The library consists of a collection of generators gathered together into a single easy-to-use package that runs on almost any computing architecture. In general, however, it is not possible to specify a set of tests sufficient to guarantee that the random numbers in a sequence are statistically independent samples of a prescribed probability distribution. As a result, there is always the possibility that results obtained using a Monte Carlo simulation may be artifacts of the random number generator used. In practice, the best way to ensure that this is not the case is to rerun the calculations with different generators.

The central idea in doing QMC on massively parallel processors involves one processor orchestrating the whole simulation, i.e., the “master-slave” paradigm. In VMC, each processor independently runs a simulation and accumulates its own set of observables; the observables from different processors are gathered, averaged, and written out by the master processor at the end of the run. The DMC algorithm described in Sec. III.D is also intrinsically parallel. Since DMC simulations require an order of magnitude more computer time than comparable VMC simulations, the argument for using parallel computers is even more compelling. The ensemble of walkers is distributed across the nodes of the parallel machine and each processor carries out the various stages of the DMC algorithm (the diffusion, drift, and creation/annihilation of walkers) on its own subset. After all the walkers have been advanced for a block of time steps, the accumulated mean walker energies from all the nodes are gathered and used to update the trial energy. It is important to load balance the algorithm by keeping the numbers of walkers on different nodes approximately equal. Like VMC, the parallel DMC algorithm achieves almost linear scaling on machines with a few hundred nodes.

XI. CONCLUSIONS

As the previous sections have demonstrated, QMC methods have been through a remarkably productive period during the last ten or twelve years. A number of theoretical and algorithmic improvements, such as the use of pseudopotentials in both VMC and DMC, the investigation and accurate correction of finite-size effects, and

the adoption of better optimization methods and sampling techniques, have raised the usefulness of QMC to a new level. These advances have made possible groundbreaking applications such as the evaluation of band gaps in insulating solids, studies of exchange-correlation holes, simulations of the jellium surface, calculations of the properties of defects in Si, and investigations of the geometries of Si and C clusters.

We hope that we have been able to show that QMC is becoming a *practical and powerful* alternative to more traditional electronic structure methods. The most important characteristics and advantages of the QMC methodology can be summarized as follows:

- (a) it gives a direct and accurate wave-function based treatment of quantum many-body effects;
- (b) it is a very general approach, applicable to solids and molecules and able to calculate almost any ground state expectation value, including energies and static correlation functions;
- (c) the N^3 scaling of the computational cost is very favorable when compared with other correlated wave function methods;
- (d) it has the significant computational advantages of easily achieved scalability on parallel architectures and low storage requirements;
- (d) the DMC method does not suffer from the basis set errors inherent in other correlated wave function methods;
- (e) it has a track record of unique calculations: the homogeneous electron gas, superfluidity in He, exchange-correlation holes and energies, etc.

To put the DMC method into a broader perspective we provide a comparison of methods in Table V. This table compares the variational and diffusion Monte Carlo (VMC and DMC), Hartree-Fock (HF), local density approximation (LDA), and coupled cluster (CC) approaches. The comparison, which is by no means exhaustive, is based on results gathered from the literature as well as on our own tests, which are mostly for *sp* systems with a limited number of elements such as C, N, H and Si. The timings given are typical, but the relative speeds of the different methods could easily vary by a factor of 3 or more depending on the programs, computers, and implementations used. The following quantities are listed:

- (a) the percentage of the valence correlation energy ($E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}}$) recovered, estimated by comparison with high accuracy calculations for small systems and by combining HF, correlated atomic, and experimental data for larger systems;

- (b) the errors in binding or cohesive energies;
- (c) the scaling of the computational cost to calculate the total energy with the number of electrons;
- (d) the actual computational time for calculations on a C₁₀ cluster: the GAMESS code (Schmidt *et al.*, 1993) was used for the HF calculations and GAUSSIAN (Frisch *et al.*, 1995) for CCSD(T); the target QMC error bar was 0.01 eV/atom and the absolute unit of cpu time was 90 seconds on a Cray C90 processor.

TABLE V. Comparison of methods

| Method | E _{corr} | E _{coh/bind} % errors | Scaling with # of electrons | Total time for C ₁₀ |
|---------|--------------------|-----------------------------------|--------------------------------------|-----------------------------------|
| HF | 0 | ≈50% | N^3 | 14 |
| LDA | N/A | 15-25% | N^3 ^a | 1 |
| VMC | ≈ 85% | 2-10% | $N^3 + \varepsilon N^4$ ^b | 16 |
| DMC | ≈ 95% | 1-4% | $N^3 + \varepsilon N^4$ ^b | 300 |
| CCSD(T) | ≈ 75% ^c | 10-15% ^c | N^7 | 1500 ^c |

^aLDA algorithms with scaling proportional to N have been proposed.

^b $\varepsilon \approx 10^{-4}$.

^cwith a 6-311G* basis set; in the limit of infinite basis set and order of excitations the CC method is formally exact.

The CC results are quoted for 6-311G* basis sets and the SD(T) level of excitations, which was the largest affordable run for C₁₀. The DMC method is considerably more computationally demanding than the HF and LDA methods, but can achieve much higher accuracy. For C₁₀ clusters, DMC is both more accurate and less computationally demanding than CCSD(T). QMC calculations also scale much better with system size than CCSD(T), and therefore we believe that QMC is the best available correlated wave function approach for calculating the energies of solids and reasonably large molecules.

QMC methods also have limitations and problems, the most important of which are:

- (a) it is demanding to calculate first and second derivatives of the total energy with respect to the atomic positions, and hence to estimate interatomic forces and force constants;
- (b) the results are obtained with a statistical error bar that decays only as the inverse square root of the computer time;
- (c) for solids the results also suffer from systematic finite-size errors;
- (d) only limited information about excited states is available;

- (e) in some cases the impact of the fixed-node approximation can bias the results.

All of these problem areas are topics of current research. For example, the calculation of derivatives using QMC techniques is being investigated by several research groups and significant advances can be expected. The statistical nature is of course inherent in QMC methods, and while the development of better trial wave functions and computers will reduce the attainable error bars, QMC methods will remain expensive compared with more approximate methods such as HF and LDA.

What are the main tasks in a DMC calculation? In Table VI we outline the various tasks together with their human and computer time demands. The table shows that the activities most demanding of human time are the generation of the single-particle orbitals and the optimization of the many-body wave function. This highlights the importance of integrating quantum chemistry and condensed matter electronic structure software packages with QMC methods, and of improving optimization methods and developing more efficient functional forms for variational wave functions. The computer time is dominated by the DMC calculations themselves, which, however, require little human intervention.

TABLE VI. Human and computational costs of a typical DMC calculation.

| Human time | Task | Computer time |
|------------|--|---------------|
| 10% | Choice of basis set, pseudopotential, etc. | 0% |
| 40% | Calculation of HF and/or post-HF orbitals | 10% |
| 40% | Many-electron trial wave function optimization | 20% |
| 10% | DMC | 70% |

As Table V illustrates, VMC calculations using Slater-Jastrow wave functions with ~ 30 variational parameters can recover between 75 and 85% of the valence correlation energy, and DMC calculations can recover roughly 95%. The remaining 5% is the fixed-node error. Although no exact solution to the sign problem is in sight, it has been shown that it is possible to decrease fixed-node errors significantly, and there is reason to hope that in the future we may be able to do calculations in which the fixed-node errors, although still present, are negligible (say, less than 1% of the correlation energy). In any case, for systems with more than a few atoms, fixed-node DMC is already much more accurate than any competing method.

It is now becoming apparent that the accuracy of the VMC and DMC methods is *systematic*: the same

accuracy pattern has been demonstrated for atoms, molecules, solids, and surfaces. This observation is perhaps the most important outcome of recent work on QMC; a decade ago there was no general expectation that QMC would be systematically accurate and no clear computational evidence to justify such a conclusion.

If high accuracy is required, QMC is already the method of choice for tackling large quantum many-body problems. In very large systems such as solids, it is also the *only* practical method based on many-body correlated wave functions, the variational principle, and the original many-electron Schrödinger equation. It is obvious, however, that much remains to be done to make QMC as flexible and easy to use as traditional electronic structure methods. This will require a sizeable investment in the further development of methods and algorithms, but we are confident that the results will justify the effort.

The largely unanticipated success of QMC simulations of solids is beginning to change the way the electronic structure community thinks about electronic correlation. With its emphasis on many-electron wave functions and probabilities, QMC has shown that it is possible to study interacting electrons in real solids using very direct computational techniques; there is no need to resort to perturbation theory or mean-field approximations. This realization is slowly overturning decades of accepted wisdom.

ACKNOWLEDGMENTS

We have benefited from the exchange of ideas and results with a large number of students, postdocs, and other collaborators, to all of whom we offer our thanks. The work of WMCF, RJN, and GR is supported by the Engineering and Physical Sciences Research Council of Great Britain. WMCF has also received support from the European Union. GR gratefully acknowledges financial support from Hitachi Ltd. (Japan). LM is supported by the State of Illinois, DOE, and NSF funds. The calculations reported in this review would not have been possible without access to powerful computing resources. We thank the staff of the National Center for Supercomputing Applications (NCSA), the University of Cambridge High Performance Computing Facility, the Edinburgh Parallel Computing Centre, and the UK Research Councils' CSAR Service for their generous help and support.

REFERENCES

- Acioli, P. H., and D. M. Ceperley, 1994, *J. Chem. Phys.* **100**, 8169.
- Alder, B. J., and D. M. Ceperley, 1996, *Phys. Rev. B* **54**, 17199.
- Alder, B. J., D. M. Ceperley, and E. L. Pollock, 1982, *Int. J. Quantum Chem.* **S16**, 49.
- Anderson, J. B., 1975, *J. Chem. Phys.* **63**, 1499.
- Anderson, J. B., 1976, *J. Chem. Phys.* **65**, 4121.
- Anderson, J. B., 1999, *Reviews in Computational Chemistry* **13**, 133.
- Ashcroft, N. W., and N. D. Mermin, 1976, *Solid State Physics* (Holt Saunders, Philadelphia), p. 133.
- Assaraf, R., and M. Caffarel, 1999, *Phys. Rev. Lett.* **83**, 4682.
- Bachelet, G. B., D. M. Ceperley, and M. G. B. Chiochetti, 1989, *Phys. Rev. Lett.* **62**, 2088.
- Baer, R., M. Head-Gordon, and D. Neuhauser, 1998, *J. Chem. Phys.* **109**, 6219.
- Baer, R., and D. Neuhauser, 2000, *J. Chem. Phys.* **112**, 1679.
- Baldereschi, A., 1973, *Phys. Rev. B* **7**, 5212.
- Ballone, P., C. J. Umrigar, and P. Delaly, 1992, *Phys. Rev. B* **45**, 6293.
- Baroni, S., and S. Moroni, 1999, *Phys. Rev. Lett.* **82**, 4745.
- Bar-Yam, Y. and J. D. Joannopoulos, 1984, *Phys. Rev. B* **30**, 1844.
- Becke, A. D., 1988, *Phys. Rev. A* **38**, 3098.
- Becke, A. D., 1998, *J. Chem. Phys.* **109**, 2092.
- Bernu, B., D. M. Ceperley, and W. A. Lester, Jr., 1990, *J. Chem. Phys.* **93**, 552.
- Bethe, H. A., and E. E. Salpeter, 1957, *Quantum Mechanics of one- and two-electron atoms* (Springer-Verlag, Berlin).
- Blankenbecler, R., D. J. Scalapino, and R. L. Sugar, 1981, *Phys. Rev. D* **24**, 2278.
- Bloch, F., 1928, *Z. Phys.* **52**, 555.
- Bloch, F., 1929, *Z. Phys.* **57**, 545.
- Blöchl, E. Smargiassi, R. Car, D. B. Laks, W. Andreoni, and S. T. Pantelides, 1993, *Phys. Rev. Lett.* **70**, 2435.
- Bohm, D., and D. Pines, 1953, *Phys. Rev.* **92**, 609.
- Bowen, C., G. Sugiyama, and B. J. Alder, 1994, *Phys. Rev. B* **50**, 14838.
- Boys, S. F., and N. C. Handy, 1969, *Proc. R. Soc. London Ser. A* **310**, 63.
- Burke, K., F. G. Cruz, and K. C. Lam, 1998, *J. Chem. Phys.* **109**, 8161.
- Caffarel, M., and P. Claverie, 1988, *J. Chem. Phys.* **88**, 1100.
- Callaway, J., 1957, *Phys. Rev.* **106**, 868.
- Ceperley, D. M., 1978, *Phys. Rev. B* **18**, 3126.
- Ceperley, D. M., 1986, *J. Stat. Phys.* **43**, 815.
- Ceperley, D. M., 1991, *J. Stat. Phys.* **63**, 1237.
- Ceperley, D. M., 1995a, *Rev. Mod. Phys.* **67**, 279.
- Ceperley, D. M., 1995b, in *Strongly Interacting Fermions and High T_c Superconductivity*, Proceedings of the Les Houches Summer School, Session LVI, edited by B. Douçot and J. Zinn-Justin (Elsevier, Amsterdam, The Netherlands), p. 427.
- Ceperley, D. M., 1999, *Nature*, **397**, 386.
- Ceperley, D. M., and B. J. Alder, 1980, *Phys. Rev. Lett.* **45**, 566.
- Ceperley, D. M., and B. J. Alder, 1981, *Physica* **108B**, 875.
- Ceperley, D. M., and B. J. Alder, 1987, *Phys. Rev. B* **36**, 2092.
- Ceperley, D. M., and B. Bernu, 1988, *J. Chem. Phys.* **89**, 6316.

- Ceperley, D. M., G. V. Chester, and M. H. Kalos, 1977, Phys. Rev. B **16**, 3081.
- Ceperley, D. M., and M. H. Kalos, 1979, in *Monte Carlo Methods in Statistical Physics*, 2nd ed., edited by K. Binder (Springer, Berlin, Germany), p. 145.
- Ceperley, D., M. H. Kalos, and J. L. Lebowitz, 1981, *Macromolecules* **14**, 1472.
- Ceperley, D. M., and L. Mitas, 1996, in *Advances in Chemical Physics*, Vol. XCIII, edited by I. Prigogine and S. A. Rice (Wiley, New York), p. 1.
- Chelikowsky, J. R., and M. L. Cohen, 1976, Phys. Rev. B **14**, 556.
- Christiansen, P. A., 1988, J. Chem. Phys. **88**, 4867.
- Christiansen, P. A., Y. S. Lee, and K. S. Pitzer, 1979, J. Chem. Phys. **71**, 4445.
- Čížek, J., 1969, Adv. Chem. Phys. **14**, 35.
- Conroy, H., 1964, J. Chem. Phys. **41**, 1331.
- Coulson, C. A., and I. Fisher, 1949, Philos. Mag. **40**, 386.
- Day, O. W., D. W. Smith, and C. Garrod, 1974, Int. J. Quantum Chem. Symp. **8**, 501.
- De Leeuw, S. W., J. W. Perram, and E. R. Smith, 1980, Proc. R. Soc. London **A373**, 27.
- DePasquale, M. F., S. M. Rothstein, and J. Vrbik, 1988, J. Chem. Phys. **89**, 3629.
- Diedrich, D. L., and J. B. Anderson, 1992, Science **258**, 786.
- Dirac, P. A. M., 1929, Proc. R. Soc. Lond. Ser. A **123**, 714.
- Donsker, M. D., and M. Kac, 1950, J. Res. Natl. Bur. Stan. **44**, 551.
- Dovesi, R., *et al.*, 1996, *CRYSTAL95 User's Manual* (University of Torino, Torino, Italy).
- Dreizler, R. M., and E. K. U. Gross, 1990, *Density Functional Theory* (Springer, Berlin).
- Eaglesham, D., 1995, Physics World **8**, No. 11, 41.
- Eckstein, H., and W. Schattke, 1995, Physica A **216**, 151.
- Eckstein, H., W. Schattke, M. Reigrotzki, and R. Redmer, 1996, Phys. Rev. B **54**, 5512.
- Eisenberger, P., and P. M. Platzman, 1970, Phys. Rev. A **2**, 415.
- Engel, E., and S. H. Vosko, 1993, Phys. Rev. B **47**, 13164.
- Ewald, P. P., 1921, Ann. Phys. **64**, 25.
- Fahy, P. M., P. B. Griffin, and J. D. Plummer, 1989, Rev. Mod. Phys. **61**, 289.
- Fahy, S., 1999, in *Quantum Monte Carlo Methods in Physics and Chemistry*, Nato Science Series C: Mathematical and Physical Sciences, Vol. 525, edited by M. P. Nightingale and C. J. Umrigar (Kluwer Academic, Dordrecht, The Netherlands), p. 101.
- Fahy, S., X. W. Wang, and S. G. Louie, 1988, Phys. Rev. Lett. **61**, 1631.
- Fahy, S., X. W. Wang, and S. G. Louie, 1990a, Phys. Rev. B **42**, 3503.
- Fahy, S., X. W. Wang, and S. G. Louie, 1990b, Phys. Rev. Lett. **65**, 1478.
- Farid, B., and R. J. Needs, 1992, Phys. Rev. B **45**, 1067.
- Feenberg, E., 1969, *Theory of Quantum Fluids* (Academic Press, New York).
- Feller, W., 1968, *An Introduction to Probability Theory and its Applications*, Vol. 1, 3rd ed. (Wiley, New York).
- Feynman, R. P., and M. Cohen, 1956, Phys. Rev. **102**, 1189.
- Filippi, C., and D. M. Ceperley, 1999, Phys. Rev. B **59**, 7909.
- Filippi, C., and S. Fahy, 2000, J. Chem. Phys. **112**, 3523.
- Filippi, C., X. Gonze, and C. J. Umrigar, 1996, in *Recent Developments and Applications of Density Functional Theory*, edited by J. M. Seminario (Elsevier, Amsterdam, The Netherlands), p. 295.
- Filippi, C., and C. J. Umrigar, 1996, J. Chem. Phys. **105**, 213.
- Filippi, C., C. J. Umrigar, and M. Taut, 1994, J. Chem. Phys. **100**, 1290.
- Flad, H.-J., M. Caffarel, and A. Savin, 1997, in *Recent Advances in Quantum Monte Carlo Methods*, edited by W. A. Lester, Jr., Recent Advances in Computational Chemistry, Vol. 2, (World Scientific, Singapore), p. 73.
- Fock, V., 1930, Z. Phys. **61**, 126.
- Foulkes, W. M. C., R. Q. Hood, and R. J. Needs, 1999, Phys. Rev. B **60**, 4558.
- Foulkes, W. M. C., and M. Schlüter, 1990, Phys. Rev. B **42**, 11505.
- Frank, W., U. Gösele, H. Mehrer, and A. Seeger, 1985, in *Diffusion in Crystalline Solids*, edited by G. E. Murch and A. S. Nowick (Academic, Orlando, FL), p. 64.
- Fraser, L. M., W. M. C. Foulkes, G. Rajagopal, R. J. Needs, S. D. Kenny, and A. J. Williamson, 1996, Phys. Rev. B **53**, 1814.
- Frisch, M. J., *et al.*, 1995, *Gaussian 94* (Gaussian Inc., Pittsburgh, PA).
- Garmer, D. R., and J. B. Anderson, 1988, J. Chem. Phys. **89**, 3050.
- Glauser, W. A., W. R. Brown, W. A. Lester, Jr., D. Bresanini, B. L. Hammond, and M. L. Koszykowski, 1992, J. Chem. Phys. **97**, 9200.
- Gösele, U., A. Plössl, and T. Y. Tan, 1996, in *Process Physics and Modeling in Semiconductor Technology*, edited by G. R. Srinivasan, C. S. Murthy, and S. T. Dunham (Electrochemical Society, Pennington, NJ), p. 309.
- Greengard, L., and V. Rokhlin, 1987, J. Comput. Phys. **73**, 325.
- Grimm, R. C., and R. G. Storer, 1971, J. Comput. Phys. **7**, 134.
- Gros, C., 1992, Z. Phys. B **86**, 359.
- Grossman, J. C., and L. Mitas, 1995, Phys. Rev. Lett. **74**, 1323.
- Grossman, J. C., and L. Mitas, 1997, Phys. Rev. Lett. **79**, 4353.
- Grossman, J. C., L. Mitas, and K. Raghavachari, 1995, Phys. Rev. Lett. **75**, 3870.
- Gunnarsson, O., M. Jonson, and B. I. Lundqvist, 1979, Phys. Rev. B **20**, 3136.
- Hamann, D. R., M. Schlüter, and C. Chiang, 1979, Phys. Rev. Lett. **43**, 1494.
- Hammond, B. L., W. A. Lester, Jr., and P. J. Reynolds, 1994, *Monte Carlo Methods in Ab Initio Quantum Chemistry* (World Scientific, Singapore).
- Hammond, B. L., P. J. Reynolds, and W. A. Lester, Jr., 1987, J. Chem. Phys. **87**, 1130.
- Hammond, B. L., P. J. Reynolds, and W. A. Lester, Jr., 1988, Phys. Rev. Lett. **61**, 2312.
- Harju, A., B. Barbiellini, S. Siljamäki, R. M. Nieminen, and G. Ortiz, 1997, Phys. Rev. Lett. **79**, 1173.
- Harrison, R. J., and N. C. Handy, 1985, Chem. Phys. Lett.

- 113**, 257.
- Hartree, D. R., 1928, Proc. Cambridge Philos. Soc. **24**, 89.
- Heitler, W., and F. London, 1927, Z. Phys. **35**, 557.
- Hemley, R. J., H. K. Mao, L. W. Finger, A. P. Jephcoat, R. M. Hazen, and C. S. Zha, 1990, Phys. Rev. B **42**, 6458.
- Hockney, R. W., and J. W. Eastwood, 1981, *Computer Simulation Using Particles* (McGraw-Hill, Maidenhead, NJ).
- Hohenberg, P., and W. Kohn, 1964, Phys. Rev. **136**, B864.
- Hood, R. Q., M.-Y. Chou, A. J. Williamson, G. Rajagopal, R. J. Needs, and W. M. C. Foulkes, 1997, Phys. Rev. Lett. **78**, 3350.
- Hood, R. Q., M.-Y. Chou, A. J. Williamson, G. Rajagopal, and R. J. Needs, 1998, Phys. Rev. B **57**, 8972.
- Huang, C.-J., C. Filippi, and C. J. Umrigar, 1998, J. Chem. Phys. **108**, 8838.
- Huang, C.-J., C. J. Umrigar, and M. P. Nightingale, 1997, J. Chem. Phys. **107**, 3007.
- Hückel, E., 1931a, Z. Phys. **70**, 204.
- Hückel, E., 1931b, Z. Phys. **72**, 310.
- Hückel, E., 1932, Z. Phys. **76**, 628.
- Hund, F., 1928, Z. Phys. **51**, 759.
- Hurley, M. M., and P. A. Christiansen, 1987, J. Chem. Phys. **86**, 1069.
- Hybertsen, M. S., and S. G. Louie, 1986, Phys. Rev. B **34**, 2920.
- Jastrow, R. J., 1955, Phys. Rev. **98**, 1479.
- Jiménez, I., L. J. Terminello, D. G. J. Sutherland, J. A. Carlisle, E. L. Shirley, and F. J. Himpsel, 1997, Phys. Rev. B **56**, 7215.
- Jones, M. D., G. Ortiz, and D. M. Ceperley, 1997, Phys. Rev. E **55**, 6202.
- Jones, R. O., and O. Gunnarsson, 1989, Rev. Mod. Phys. **61**, 689.
- Kalos, M. H., 1962, Phys. Rev. **128**, 1791.
- Kalos, M. H., 1967, J. Comput. Phys. **2**, 257.
- Kalos, M. H., D. Levesque and L. Verlet, 1974, Phys. Rev. A **9**, 257.
- Kalos, M., and F. Pederiva, 1999, in *Quantum Monte Carlo Methods in Physics and Chemistry*, Nato Science Series C: Mathematical and Physical Sciences, Vol. 525, edited by M. P. Nightingale and C. J. Umrigar (Kluwer Academic, Dordrecht, The Netherlands), p. 263.
- Kalos, M. H., and P. A. Whitlock, 1986, *Monte Carlo Methods Volume 1: Basics* (Wiley, New York).
- Karlin S., and H. M. Taylor, 1981, *A Second Course in Stochastic Processes* (Academic Press, New York).
- Kato, T., 1957, Communications on Pure and Applied Mathematics **10**, 151.
- Kenny, S. D., G. Rajagopal, and R. J. Needs, 1995, Phys. Rev. A **51**, 1898.
- Kenny, S. D., G. Rajagopal, R. J. Needs, W.-K. Leung, M. J. Godfrey, A. J. Williamson, and W. M. C. Foulkes, 1996, Phys. Rev. Lett. **77**, 1099.
- Kent, P. R. C., R. Q. Hood, M. D. Towler, R. J. Needs, and G. Rajagopal, 1998, Phys. Rev. B **57**, 15293.
- Kent, P. R. C., R. Q. Hood, A. J. Williamson, R. J. Needs, W. M. C. Foulkes, and G. Rajagopal, 1999, Phys. Rev. B **59**, 1917.
- Kent, P. R. C., R. J. Needs, and G. Rajagopal, 1999, Phys. Rev. B **59**, 12344.
- Kent, P. R. C., M. D. Towler, R. J. Needs, and G. Rajagopal, 2000, unpublished.
- Kittel, C., 1966 *Introduction to Solid State Physics*, 3rd ed. (Wiley, New York, 1966).
- Knittle, E., R. Wentzcovitch, R. Jeanloz, and M. L. Cohen, 1989, Nature (London) **337**, 349.
- Knuth, D. E., 1997, *The Art of Computer Programming*, 3rd ed., Vol. 2, (Addison-Wesley, Reading, Massachusetts), p. 61.
- Kohn, W., and L. J. Sham, 1965, Phys. Rev. **140**, A1133.
- Kralik, B., P. Delaney, and S. G. Louie, 1998, Phys. Rev. Lett. **80**, 4253.
- Krotscheck, E., W. Kohn and G.-X. Qian, 1985, Phys. Rev. B **32**, 5693.
- Kwon, Y., D. M. Ceperley, and R. M. Martin, 1993, Phys. Rev. B **48**, 12037.
- Kwon, Y., D. M. Ceperley, and R. M. Martin, 1994, Phys. Rev. B **50**, 1684.
- Kwon, Y., D. M. Ceperley, and R. M. Martin, 1996, Phys. Rev. B **53**, 7376.
- Kwon, Y., D. M. Ceperley, and R. M. Martin, 1998, Phys. Rev. B **58**, 6800.
- Lam, L., and P. M. Platzman, 1974, Phys. Rev. B **9**, 5122.
- Langreth, D. C., and M. J. Mehl, 1983, Phys. Rev. B **28**, 1809.
- Lester, W. A., 1997, Ed. *Recent Advances in Quantum Monte Carlo Methods*, Recent Advances in Computational Chemistry, Vol. 2 (World Scientific, Singapore).
- Leung, W.-K., R. J. Needs, G. Rajagopal, S. Itoh, and S. Ihara, 1999, Phys. Rev. Lett. **83**, 2351.
- Li, X.-P., D. M. Ceperley, and R. M. Martin, 1991, Phys. Rev. B **44**, 10929.
- Li, X.-P., R. J. Needs, R. M. Martin, and D. M. Ceperley, 1992, Phys. Rev. B **45**, 6124.
- Lorenzana, H. E., I. F. Silvera, and K. A. Goettel, 1990, Phys. Rev. Lett. **64**, 1939.
- Magro, W. R., D. M. Ceperley, C. Pierleoni, and B. Bernu, 1996, Phys. Rev. Lett. **76**, 1240.
- Malatesta, A., S. Fahy, and G. B. Bachelet, 1997, Phys. Rev. B **56**, 12201.
- McMillan, W. L., 1965, Phys. Rev. **A138**, 442.
- Meierovich, M., A. Mushinski, and M. P. Nightingale, 1996, J. Chem. Phys. **105**, 6498.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953, J. Chem. Phys. **21**, 1087.
- Militzer, B., W. Magro, and D. M. Ceperley, 1999, Contributions to Plasma Physics **39**, 151.
- Mitas, L., 1993, in *Computer Simulation Studies in Condensed Matter Physics V*, edited by D. P. Landau, K. K. Mon, and H. B. Schuttler (Springer, Berlin).
- Mitas, L., 1994, Phys. Rev. A **49**, 4411.
- Mitas, L., 1996, Comp. Phys. Commun. **96**, 107.
- Mitas, L., and J. C. Grossman, 1997, in *Recent Advances in Quantum Monte Carlo Methods*, Recent Advances in Computational Chemistry, Vol. 2, edited by W. A. Lester, Jr. (World Scientific, Singapore), p. 133.
- Mitas, L., and R. M. Martin, 1994, Phys. Rev. Lett. **72**, 2438.
- Mitas, L., E. L. Shirley, and D. M. Ceperley, 1991, J. Chem. Phys. **95**, 3467.
- Monkhorst, H. J., and J. D. Pack, 1976, Phys. Rev. B **13**,

- 5188.
- Moroni, S., D. M. Ceperley, and G. Senatore, 1992, Phys. Rev. Lett. **69**, 1837.
- Moroni, S., D. M. Ceperley, and G. Senatore, 1995, Phys. Rev. Lett. **75**, 689.
- Morrell, M. M., R. G. Parr, and M. Levy, 1975, J. Chem. Phys. **62**, 549.
- Moskowitz, J. W., K. E. Schmidt, M. A. Lee, and M. H. Kalos, 1982, J. Chem. Phys. **77**, 349.
- Müller, W., J. Flesch, and W. Meyer, 1984, J. Chem. Phys. **80**, 3297.
- Müller, W., and W. Meyer, 1984, J. Chem. Phys. **80**, 3311.
- Mulliken, R. S., 1928, Phys. Rev. **32**, 186.
- Murphy, R. B., and R. A. Friesner, 1998, Chem. Phys. Lett. **288**, 403.
- Natoli, V., and D. M. Ceperley, 1995, J. Comp. Phys. **117**, 171.
- Natoli, V., R. M. Martin, and D. M. Ceperley, 1993, Phys. Rev. Lett. **70**, 1952.
- Natoli, V., R. M. Martin, and D. M. Ceperley, 1995, Phys. Rev. Lett. **74**, 1601.
- Nekovee, M., W. M. C. Foulkes, and R. J. Needs, 2000, unpublished.
- Nekovee, M., W. M. C. Foulkes, A. J. Williamson, G. Rajagopal, and R. J. Needs, 1999, Adv. Quantum Chem. **33**, 189.
- Nightingale, M. P., 1999, in *Quantum Monte Carlo Methods in Physics and Chemistry*, Nato Science Series C: Mathematical and Physical Sciences, Vol. 525, edited by M. P. Nightingale and C. J. Umrigar (Kluwer Academic, Dordrecht, The Netherlands), p. 1.
- Nightingale, M. P., and C. J. Umrigar, 1997, in *Recent Advances in Quantum Monte Carlo Methods*, Recent Advances in Computational Chemistry, Vol. 2, edited by W. A. Lester, Jr. (World Scientific, Singapore), p. 201.
- Nightingale, M. P., and C. J. Umrigar, 1999, Eds., *Quantum Monte Carlo Methods in Physics and Chemistry*, Nato Science Series C: Mathematical and Physical Sciences, Vol. 525, (Kluwer Academic, Dordrecht, The Netherlands).
- Ortiz, G., and P. Ballone, 1994, Phys. Rev. B **50**, 1391.
- Ortiz, G., D. M. Ceperley, and R. M. Martin, 1993, Phys. Rev. Lett. **71**, 2777.
- Ortiz, G., M. Harris, and P. Ballone, 1999, Phys. Rev. Lett. **82**, 5317.
- Pack, R. T., and W. B. Brown, 1966, J. Chem. Phys. **45**, 556.
- Pandey, K. C., 1986, Phys. Rev. Lett. **57**, 2287.
- Panoff, R. M., and J. Carlson, 1989, Phys. Rev. Lett. **62**, 1130.
- Parr, R. G., and W. Yang, 1989, *Density Functional Theory of Atoms and Molecules* (Oxford University, New York).
- Perdew, J. P., J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, 1992, Phys. Rev. B **46**, 6671; **48**, 4978(E).
- Perdew, J. P., and A. Zunger, 1981, Phys. Rev. B **23**, 5048.
- Pierleoni, C., B. Bernu, D. M. Ceperley, and W. R. Magro, 1994, Phys. Rev. Lett. **73**, 2145.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992, *Numerical Recipes: The Art of Scientific Computing*, 2nd ed. (Cambridge University, Cambridge, England).
- Raghavachari, K., B. Zhang, J. A. Pople, B. G. Johnson, and P. M. W. Gill, 1993, Chem. Phys. Rev. Lett. **214**, 357.
- Rajagopal, G., and R. J. Needs, 1994, J. Comp. Phys. **115**, 399.
- Rajagopal, G., R. J. Needs, A. James, S. D. Kenny, and W. M. C. Foulkes, 1995, Phys. Rev. B **51**, 10591.
- Rajagopal, G., R. J. Needs, S. Kenny, W. M. C. Foulkes, and A. James, 1994, Phys. Rev. Lett. **73**, 1959.
- Reynolds, P. J., D. M. Ceperley, B. J. Alder, and W. A. Lester, Jr., 1982, J. Chem. Phys. **77**, 5593.
- Rohlfing, M., P. Krüger, and J. Pollmann, 1993, Phys. Rev. B **48**, 17791.
- Sakurai, Y., Y. Tanaka, A. Bansil, S. Kaprzyk, A. T. Stewart, Y. Nagashima, T. Hydo, S. Nanao, H. Kawata, and N. Shiotani, 1995, Phys. Rev. Lett. **74**, 2252.
- Schmidt, K. E., M. A. Lee, M. H. Kalos, and G. V. Chester, 1981, Phys. Rev. Lett. **47**, 807.
- Schmidt, K. E., and J. W. Moskowitz, 1990, J. Chem. Phys. **93**, 4172.
- Schmidt, K. E., and J. W. Moskowitz, 1992, J. Chem. Phys. **97**, 3382.
- Schmidt, M. W., *et al.*, 1993, J. Comput. Chem. **14**, 1347.
- Senatore, G., and N. H. March, 1994, Rev. Mod. Phys. **66**, 445.
- Senatore, G., S. Moroni, and D. M. Ceperley, 1999, in *Quantum Monte Carlo Methods in Physics and Chemistry*, Nato Science Series C: Mathematical and Physical Sciences, Vol. 525, edited by M. P. Nightingale and C. J. Umrigar (Kluwer Academic, Dordrecht, The Netherlands), p. 183.
- Shirley, E. L., and R. M. Martin, 1993, Phys. Rev. B **47**, 15413.
- Shirley, E. L., L. Mitas, and R. M. Martin, 1991, Phys. Rev. B **44**, 3395.
- Singwi, K. S., and M. P. Tosi, 1981, in *Solid State Physics*, Vol. 36, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York), p. 177.
- Slater, J. C., 1930, Phys. Rev. **35**, 210.
- Snir, M., S. Otto, S. Huss-Lederman, D. Walker, and J. Dongarra, 1998, *MPI—The Complete Reference*, 2nd ed. (MIT Press, Cambridge).
- Sommerfeld, A., and H. Bethe, 1933, *Handbuch der Physik* **24** (Springer, Berlin).
- Stedman, M. L., W. M. C. Foulkes, and M. Nekovee, 1998, J. Chem. Phys. **109**, 2630.
- Szabo, A., and N. S. Ostlund, 1989, *Modern Quantum Chemistry* (McGraw-Hill, New York).
- Tanaka, S., 1993, J. Phys. Soc. Japan **62**, 2112.
- Tanaka, S., 1995, J. Phys. Soc. Japan **64**, 4270.
- Tanatar, B., and D. M. Ceperley, 1989, Phys. Rev. B **39**, 5005.
- ten Haaf, D. F. B., H. J. M. van Bommel, J. M. J. van Leeuwen, W. van Saarloos, and D. M. Ceperley, 1994, Phys. Rev. B **51**, 13039.
- Tosi, M. P., 1964, in *Solid State Physics*, Vol. 16, edited by H. Ehrenreich and D. Turnbull (Academic, New York), p. 1.
- Towler, M. D., R. Q. Hood, R. J. Needs, and G. Rajagopal, 2000, unpublished.
- Umrigar, C. J., 1993, Phys. Rev. Lett. **71**, 408.
- Umrigar, C. J., and X. Gonze, 1994, Phys. Rev. A **50**, 3827.
- Umrigar, C. J., M. P. Nightingale, and K. J. Runge, 1993, J.

Chem. Phys. **99**, 2865.

Umrigar, C. J., K. G. Wilson, and J. W. Wilkins, 1988, Phys. Rev. Lett. **60**, 1719.

Valentí, R., C. Gros, P. J. Hirschfeld, and W. Stephan, 1991, Phys. Rev. B **44**, 13203.

Vosko, S. H., L. Wilk, and M. Nusair, 1980, Can. J. Phys. **58**, 1200.

Wigner, E., 1934, Phys. Rev. **46**, 1002.

Williamson, A. J., S. D. Kenny, G. Rajagopal, A. J. James, R. J. Needs, L. M. Fraser, W. M. C. Foulkes, and P. MacCallum, 1996, Phys. Rev. B **53**, 9640.

Williamson, A. J., G. Rajagopal, R. J. Needs, L. M. Fraser, W. M. C. Foulkes, Y. Wang, and M.-Y. Chou, 1997, Phys. Rev. B **55**, R4851.

Williamson, A. J., R. Q. Hood, R. J. Needs, and G. Rajagopal, 1998, Phys. Rev. B **57**, 12140.

Wu, Y. S. M., A. Kuppermann, and J. B. Anderson, 1999, PCCP Phys. Chem., Chem. Phys. **1**, 929.

Yan, Z., J. P. Perdew, S. Kurth, C. Fiolhais, and L. Almeida, 2000, Phys. Rev. B **61**, 2595.

Young, D. P., D. Hall, M. E. Torelli, Z. Fisk, J. L. Sarrao, J. D. Thompson, H. R. Ott, S. B. Oseroff, R. G. Goodrich, and R. Zysler, 1999, Nature **397**, 412.

Zong, F. H., and D. M. Ceperley, 1998, Phys. Rev. E **58**, 5123.