# Between Classical and Quantum Monte Carlo Methods: "Variational" QMC

DARIO BRESSANINI[1]
*Istituto di Scienze Matematiche Fisiche e Chimiche*
*Universita' di Milano, sede di Como*
*Via Lucini 3, I-22100 Como (Italy)*

and

PETER J. REYNOLDS[1]
*Physical Sciences Division*
*Office of Naval Research*
*Arlington, VA 22217 USA*

## ABSTRACT

The variational Monte Carlo method is reviewed here. It is in essence a classical statistical mechanics approach, yet allows the calculation of quantum expectation values. We give an introductory exposition of the theoretical basis of the approach, including sampling methods and acceleration techniques; its connection with trial wavefunctions; and how in practice it is used to obtain high quality quantum expectation values through correlated wavefunctions, correlated sampling, and optimization. A thorough discussion is given of the different methods available for wavefunction optimization. Finally, a small sample of recent works is reviewed, giving results and indicating new techniques employing variational Monte Carlo.

## I. INTRODUCTION

Variational Monte Carlo (or VMC as it is now commonly called) is a method which allows one to calculate quantum expectation values given a trial wavefunction [1,2]. The actual Monte Carlo methodology used for this is almost identical to the usual classical

---

[1] Also Department of Physics, Georgetown University, Washington, D.C. USA.

Monte Carlo methods, particularly those of statistical mechanics. Nevertheless, quantum behavior can be studied with this technique. The key idea, as in classical statistical mechanics, is the ability to write the desired property $\langle O \rangle$ of a system as an average over an ensemble

$$\langle O \rangle = \frac{\int P(\mathbf{R}) O(\mathbf{R}) d\mathbf{R}}{\int P(\mathbf{R}) d\mathbf{R}} \tag{1.2}$$

for some specific probability distribution $P(\mathbf{R})$. In classical equilibrium statistical mechanics this would be the Boltzmann distribution. If $\langle O \rangle$ is to be a quantum expectation value, $P(\mathbf{R})$ must be the square of the wavefunction $\Psi(\mathbf{R})$. True quantum Monte Carlo methods (see, e.g., the following chapter) allow one to actually sample $\Psi(\mathbf{R})$. Nevertheless, classical Monte Carlo is sufficient (though approximate) through the artifact of sampling from a *trial* wavefunction. How to obtain such a wavefunction is *not* directly addressed by VMC. However, optimization procedures, which will be discussed below, and possibly feedback algorithms, enable one to modify an existing wavefunction choice once made.

A great advantage of Monte Carlo for obtaining quantum expectation values is that wavefunctions of great functional complexity are amenable to this treatment, since analytical integration is not being done. This greater complexity, including for example explicit two-body and higher-order correlation terms, in turn allows for a far more compact description of a many-body system than possible with most non-Monte Carlo methods, with the benefit of high absolute accuracy being possible. The primary disadvantage of using a Monte Carlo approach is that the calculated quantities contain a statistical uncertainty, which needs to be made small. This can always be done in VMC, but at the cost of CPU time, since statistical uncertainty decreases as $N^{1/2}$ with increasing number of samples $N$.

A quantity often sought with these methods is the expectation value of the Hamiltonian, i.e., the total energy. As with all total energy methods, whether Monte Carlo or not, one needs to consider scaling. That is, as the systems being treated increase in size, how does the computational cost rise? Large-power polynomial scaling offers a severe roadblock to the treatment of many physically interesting systems. With such scaling, even significantly faster computers would leave large classes of interesting problems untouchable. This is the motivation behind the so-called order-$N$ methods in, e.g., density functional theory, where in that case $N$ is the number of electrons in the system. While density functional theory is useful in many contexts, often an exact treatment of electron correlation, or at least a systematically improveable treatment, is necessary or desirable. Quantum chemical approaches of the latter variety are unfortunately among the class of methods that scale with large powers of system size. This is another advantage of Monte Carlo methods, which scale reasonably well, generally between $N^2$ and $N^3$; moreover, algorithms with lower powers are possible to implement (e.g. using fast multipole methods to evaluate the Coulomb potential, and the use of localized orbitals together with sparse matrix techniques for the wavefunction computation).

The term "variational" Monte Carlo derives from the use of this type of Monte Carlo in conjunction with the variational principle; this provides a bounded estimate of the total energy together with a means of improving the wavefunction and energy estimate. Despite the inherent statistical uncertainty, a number of very good algorithms have been created that allow one to optimize trial wavefunctions in this way[3,4,5], and we discuss this at some length below. The best of these approaches go beyond simply minimizing the energy, and exploit the minimization of the energy variance as well, this latter quantity vanishing for energy eigenfunctions.

Before getting into details, let us begin with a word about notation. The position vector $\mathbf{R}$ which we use, lives in the 3-$M$ dimensional coordinate space of the $M$ (quantum) particles comprising the system. This vector is, e.g., the argument of the trial wavefunction $\Psi_T(\mathbf{R})$; however, sometimes we will omit the explicit dependence on $\mathbf{R}$ to avoid cluttering the equations, and simply write $\Psi_T$. Similarly, if the trial wavefunction depends on some parameter $\alpha$ (this may be the exponent of a Gaussian or Slater orbital for example) we may write $\Psi_T(\mathbf{R};\alpha)$, or simply $\Psi_T(\alpha)$, again omitting the explicit $\mathbf{R}$ dependence.

The essence of VMC is the creation and subsequent sampling of a distribution $P(\mathbf{R})$ proportional to $\left|\Psi_T(\mathbf{R})\right|^2$. Once such a distribution is established, expectation values of various quantities may be sampled. Expectation values of non-differential operators may be obtained simply as

$$\left\langle \hat{O} \right\rangle = \frac{\int \left|\Psi_T(\mathbf{R})\right|^2 \hat{O}(\mathbf{R}) d\mathbf{R}}{\int \left|\Psi_T(\mathbf{R})\right|^2 d\mathbf{R}} \cong \frac{1}{N}\sum_{i=1}^{N} \hat{O}(\mathbf{R}_i) \quad . \tag{1.4}$$

Differential operators are only slightly more difficult to sample, since we can write

$$\left\langle \hat{O} \right\rangle = \frac{\int \left|\Psi_T(\mathbf{R})\right|^2 \frac{\hat{O}\Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})} d\mathbf{R}}{\int \left|\Psi_T(\mathbf{R})\right|^2 d\mathbf{R}} \cong \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{O}\Psi_T(\mathbf{R}_i)}{\Psi_T(\mathbf{R}_i)} \quad . \tag{1.6}$$

## II. MONTE CARLO SAMPLING A TRIAL WAVEFUNCTION

### A. Metropolis sampling

The key problem is how to create and sample the distribution $\Psi_T^2(\mathbf{R})$ (from now on, for simplicity, we consider only real trial wavefunctions). This is readily done in a number of ways, possibly familiar from statistical mechanics. Probably the most common method is simple Metropolis sampling[6]. Specifically, this involves generating a Markov chain of steps by "box sampling" $\mathbf{R}' = \mathbf{R} + \vec{\varsigma}\,\Delta$, with $\Delta$ the box size, and $\vec{\varsigma}$ a $3M$-dimensional vector of uniformly distributed random numbers $\varsigma \in [-1,+1]$. This is followed by the classic Metropolis accept/reject step, in which $\left(\Psi_T(\mathbf{R}')/\Psi_T(\mathbf{R})\right)^2$ is compared to a uniformly distributed random number between zero and unity. The new coordinate $\mathbf{R}'$ is accepted only if this ratio of trial wavefunctions squared exceeds the random number. Otherwise the new coordinate remains at $\mathbf{R}$. This completes one step of the Markov chain (or random walk). Under very general conditions, such a Markov chain results in an asymptotic equilibrium distribution proportional to $\Psi_T^2(\mathbf{R})$. Once established, the properties of interest can be "measured" at each point $\mathbf{R}$ in the Markov chain (which we refer to as a *configuration*) using Eqs. 1.4 and 1.6, and averaged to obtain the desired estimate. The more configurations that are generated, the more accurate the estimate one gets. As is normally done in standard applications of the Metropolis method, proper care must be taken when estimating the statistical error, since the configurations generated by a Markov chain are not statistically independent; they are *serially correlated* [7]. The device of dividing the simulation into *blocks* of sufficient length, and computing the statistical error only over the block averages, is usually sufficient to eliminate this problem.

### B. Langevin simulation

The sampling efficiency of the simple Metropolis algorithm can be improved when one switches to the Langevin simulation scheme [8]. The Langevin approach may be thought of as providing a kind of importance sampling which is missing from the standard Metropolis approach. One may begin by writing a Fokker-Planck equation whose steady state solution is $\Psi_T^2(\mathbf{R})$. Explicitly this is

$$\frac{\partial f(\mathbf{R},t)}{\partial t} = D\nabla^2 f(\mathbf{R},t) - D\nabla \cdot (f(\mathbf{R},t)\mathbf{F}(\mathbf{R})) \ , \tag{2.2}$$

where

$$\mathbf{F}(\mathbf{R}) = \nabla \ln \Psi_T^2(\mathbf{R}) \tag{2.4}$$

is an explicit function of $\Psi_T$ generally known as either the *quantum velocity* or the *quantum force*. By direct substitution it is easy to check that $\Psi_T^2(\mathbf{R})$ is the exact steady state solution.

The (time) discretized evolution of the Fokker-Planck equation may be written in terms of **R,** and this gives the following Langevin-type equation

$$\mathbf{R}' = \mathbf{R} + D\tau\mathbf{F}(\mathbf{R}) + \sqrt{2D\tau}\chi \ , \tag{2.6}$$

where $\tau$ is the step size of the time integration and $\chi$ is a Gaussian random variable with zero mean and unit width. Numerically one can use the Langevin equation to generate the path of a configuration, or random *walker* (more generally, an ensemble of such walkers), through position space. As with Metropolis, this path is also a Markov chain. One can see that the function $\mathbf{F}(\mathbf{R})$ acts as a drift, pushing the walkers towards regions of configuration space where the trial wavefunction is large. This increases the efficiency of the simulation, in contrast to the standard Metropolis move where the walker has the same probability of moving in every direction.

There is, however, a minor point that needs to be addressed: the time discretization of the Langevin equation, exact only for $\tau \to 0$, has introduced a time step *bias* absent in

Metropolis sampling. This can be eliminated by performing different simulations at different time steps and extrapolating to $\tau \rightarrow 0$. However, a more effective procedure can be obtained by adding a Metropolis-like acceptance/rejection step after the Langevin move. The net result is a generalization of the standard Metropolis algorithm in which a Langevin equation, containing drift and diffusion (i.e., a quantum force term depending on the positions of all the electrons, plus white noise), is employed for the transition matrix carrying us from $\mathbf{R}$ to $\mathbf{R}'$. This is a specific generalization of Metropolis. We discuss the generic generalization next.

## C. Generalized Metropolis

In the Metropolis algorithm, a single move of a walker starting at $\mathbf{R}$ can be split into two steps as follows: first a possible final point $\mathbf{R}'$ is selected; then an acceptance/rejection step is executed. If the first step is taken with a transition probability $T(\mathbf{R} \rightarrow \mathbf{R}')$, and if we denote for the acceptance/rejection step the probability $A(\mathbf{R} \rightarrow \mathbf{R}')$ that the attempted move from $\mathbf{R}$ to $\mathbf{R}'$ is accepted, then the total probability that a walker moves from $\mathbf{R}$ to $\mathbf{R}'$ is $T(\mathbf{R} \rightarrow \mathbf{R}')\ A(\mathbf{R} \rightarrow \mathbf{R}')$. Since we seek the distribution $P(\mathbf{R})$ using such a Markov process, we note that at equilibrium (and in an infinite ensemble), the fraction of walkers going from $\mathbf{R}$ to $\mathbf{R}'$, $P(\mathbf{R})T(\mathbf{R} \rightarrow \mathbf{R}')A(\mathbf{R} \rightarrow \mathbf{R}')$, must be equal to the fraction of walkers going from $\mathbf{R}'$ to $\mathbf{R}$, namely $P(\mathbf{R}')T(\mathbf{R}' \rightarrow \mathbf{R})A(\mathbf{R}' \rightarrow \mathbf{R})$. This condition, called *detailed balance*, is a *sufficient* condition to reach the desired steady state, and provides a constraint on the possible forms for $T$ and $A$. For a given $P(\mathbf{R})$

$$P(\mathbf{R})T(\mathbf{R} \rightarrow \mathbf{R}')A(\mathbf{R} \rightarrow \mathbf{R}') = P(\mathbf{R}')T(\mathbf{R}' \rightarrow \mathbf{R})A(\mathbf{R}' \rightarrow \mathbf{R}) \ ; \tag{2.8}$$

thus the acceptance probability must satisfy

$$\frac{A(\mathbf{R} \rightarrow \mathbf{R}')}{A(\mathbf{R}' \rightarrow \mathbf{R})} = \frac{P(\mathbf{R}')}{P(\mathbf{R})} \frac{T(\mathbf{R}' \rightarrow \mathbf{R})}{T(\mathbf{R} \rightarrow \mathbf{R}')} \quad . \tag{2.10}$$

Since for our situation $P(\mathbf{R}) = \Psi_T^2(\mathbf{R})$, a Metropolis solution for $A$ is

$$A(\mathbf{R} \to \mathbf{R}') = \min\left(1, \left(\frac{\Psi_T(\mathbf{R}')}{\Psi_T(\mathbf{R})}\right)^2 \frac{T(\mathbf{R}' \to \mathbf{R})}{T(\mathbf{R} \to \mathbf{R}')}\right) \quad . \tag{2.12}$$

The original Metropolis scheme moves walkers in a rectangular box centered at the initial position; in this case the ratio of the $T$ 's is simply equal to unity, and the standard Metropolis algorithm is recovered. This is readily seen to be less than optimal if the distribution to be sampled is very different from uniform, e.g., rapidly varying in regions of space. It makes sense to use a transition probability for which the motion towards a region of increasing $\Psi_T^2(\mathbf{R})$ is enhanced. Toward this goal there are many possible choices for $T$; the Langevin choice presented above is a particular, very efficient, choice of the transition probability:

$$T(\mathbf{R} \to \mathbf{R}') = (4\pi D\tau)^{-3N/2} e^{-(\mathbf{R}'-\mathbf{R}-D\tau\mathbf{F}(\mathbf{R}))^2/4D\tau} \quad . \tag{2.14}$$

This transition probability is a displaced Gaussian, and may be sampled using the Langevin equation (Eq. 2.6) by moving the walker first with a drift and then with a random Gaussian step representing diffusion.

## III. TRIAL WAVEFUNCTIONS

The exact wavefunction is a solution to the Schrodinger equation. For any but the simplest systems the form of the wavefunction is unknown. However, it can be approximated in a number of ways. Generally this can be done systematically through series expansions of some sort, such as basis set expansions or perturbation theory. The convergence of these series depends upon the types of terms included. Most variational electronic structure methods rely on a double basis-set expansion for the wavefunction: one in single electron orbitals, and the other in $M$-electron Slater determinants. This is in no way the most general form of expansion possible. At a minimum, it omits explicit two-body (and many-body) terms. This omission results in generally slow convergence of the resultant series. An important characteristic of Monte Carlo methods is their ability to use arbitrary wavefunction forms, including ones having explicit interelectronic distance and other many-body dependencies. This enables greater flexibility and hence more compact representation than is possible with forms constructed solely with one-electron functions.

The one-electron form, however, provides a useful starting point for constructing the more general forms we desire. The one-electron form comes from the widely used methods of traditional *ab initio* electronic structure theory, based on molecular orbital (MO) expansions and the Hartree-Fock approximation. As a first approximation, the $M$-electron wavefunction is represented by a single Slater determinant of spin orbitals. This independent-particle approximation completely ignores the many-body nature of the wavefunction, incorporating quantum exchange, but no correlation; within this approach correlation is later built in through a series expansion of Slater determinants (see below).

The MOs are themselves expressed as linear combinations of atomic orbitals (AOs), the latter usually a basis set of known functions. With a given basis set, the problem of variationally optimizing the energy transforms into that of finding the coefficients of the orbitals. Expressed in matrix form in an AO basis, and in the independent particle approximation of Hartree-Fock theory, this leads to the well-known self-consistent field (SCF) equations.

There are two broad categories of methods that go beyond Hartree-Fock in constructing wavefunctions: configuration interaction (CI), and many-body perturbation theory. In CI one begins by noting that the exact $M$-electron wavefunction can be expanded as a linear combination of an infinite set of Slater determinants which span the Hilbert space of electrons. These can be any complete set of $M$-electron antisymmetric functions. One such choice is obtained from the Hartree-Fock method by substituting all excited states for each MO in the determinant. This, of course, requires an infinite number of determinants, derived from an infinite AO basis set, possibly including continuum functions. Like Hartree-Fock, there are no many-body terms explicitly included in CI expansions either. This failure results in an extremely slow convergence of CI expansions[9]. Nevertheless, CI is widely used, and has sparked numerous related schemes that may be used, in principle, to construct trial wavefunctions.

What is the physical nature of the many-body correlations which are needed to accurately describe the many-body system? Insight into this question might provide us with a more compact representation of the wavefunction. There are essentially two kinds of correlation: *dynamical* and *non-dynamical*. An example of the former is angular correlation. Consider He, where the Hartree-Fock determinant places both electrons uniformly in spherical symmetry around the nucleus: the two electrons are thus uncorrelated. One could add a small contribution of a determinant of S symmetry, built using 2*p* orbitals, to increase the wavefunction when the electrons are on opposite sides of the nucleus and decreases it when they are on the same side. Likewise, radial correlation can be achieved by adding a 2*s* term. Both of these dynamical correlation terms describe (in part) the instantaneous positions taken by the two electrons. On the other hand, non-dynamic correlation results from geometry changes and near degeneracies. An example is encountered in the dissociation of a molecule. It also occurs when, e.g., a Hartree-Fock excited state is close enough in energy to mix with the ground state. These non-dynamical correlations result in the well-known deficiency of the Hartree-Fock method that dissociation is not into two neutral fragments, but rather into ionic configurations. Thus, for a proper description of reaction pathways, a multi-determinant wavefunction is required: one containing a determinant or a linear combination of determinants corresponding to all fragment states.

Hartree-Fock and post Hartree-Fock wavefunctions, which do not *explicitly* contain many-body correlation terms lead to molecular integrals that are substantially more convenient for numerical integration. For this reason, the vast majority of (non-Monte Carlo) work is done with such independent-particle-type functions. However, given the flexibility of Monte Carlo integration, it is very worthwhile in VMC to incorporate many-body correlation *explicitly*, as well as incorporating other properties a wavefunction ideally should possess. For example, we know that because the true wavefunction is a solution of the Schrodinger equation, the local energy must be a constant for an eigenstate. (Thus, for approximate wavefunctions the variance of the local energy becomes an important measure of wavefunction quality.) Because the local energy should be a constant everywhere in space, each singularity of the Coulomb potential must be canceled by a corresponding term in the local kinetic energy. This condition results in a *cusp*, i.e. a discontinuity in the first derivative of $\Psi_T$, where two charged particles meet [10]. Satisfying this leads, in large measure, to more rapidly convergent expansions. With a sufficiently flexible trial wavefunction one can include appropriate parameters, which can then be determined by the cusp condition. For the electron-nuclear cusp this condition is

$$\frac{1}{\Psi}\frac{\partial\Psi}{\partial r}\bigg|_{r=0} = -Z \ , \tag{3.2}$$

where $r$ is any single electron-nucleus coordinate. If we solve for $\Psi$ we find that, locally, it must be exponential in $r$. The extension to the many-electron case is straightforward. As any single electron (with all others fixed) approaches the nucleus, the exact wavefunction behaves asymptotically as in the one-electron case, for each electron individually. An extension of this argument to the electron-electron cusp is also readily done. In this case, as electron $i$ approaches electron $j$, one has a two-body problem essentially equivalent to the hydrogenic atom. Therefore, in analogy to the above electron-nucleus case, one obtains the cusp conditions

$$\begin{aligned} \frac{1}{\Psi}\frac{\partial\Psi}{\partial r_{ij}}\bigg|_{r_{ij}=0} &= \frac{1}{2} \qquad \textit{unlike spin} \\ \frac{1}{\Psi}\frac{\partial\Psi}{\partial r_{ij}}\bigg|_{r_{ij}=0} &= \frac{1}{4} \qquad \textit{like spin} \end{aligned} \quad . \tag{3.4}$$

For like spins, the vanishing of the Slater determinant at $r_{ij} = 0$ contributes partially to satisfying the cusp condition. (Another factor of two results from the indistinguishability of the electrons.) From these equations we see the need for *explicit* two-body terms in the wavefunction, for with a flexible enough form of $\Psi$ we can then satisfy the cusp conditions, thereby matching the Coulomb singularity for any particle pair with terms from the kinetic energy. Note also that while Slater-type (exponential) orbitals (STO's) have the proper hydrogenic cusp behavior, Gaussian-type orbitals (GTO's) do not. Thus, basis sets consisting of GTO's, although computationally expedient for non-Monte Carlo integral evaluation, cannot directly satisfy the electron-nucleus cusp condition, and are therefore less desirable as Monte Carlo trial wavefunctions.

Three-particle coalescence conditions also have been studied. These singularities are not a result of the divergence of the potential, but are entirely due to the kinetic energy (i.e., to the form of the wavefunction). To provide a feel for the nature of these terms, we note that Fock [11] showed by an examination of the helium atom in hyperspherical coordinates that terms of the form $(r_1^2 + r_2^2)\ln(r_1^2 + r_2^2)$ are important when $r_1$ and $r_2$ $\rightarrow 0$ simultaneously. Additional higher-order terms, describing correlation effects and higher *n*-body coalescences, also have been suggested.

Since *explicit* many-body terms are critical for a compact description of the wavefunction, let us review some early work along these lines. Hylleraas and Pekeris had great success for He with wavefunctions of the form

$$\Psi_{Hylleraas} = \sum_{k=1}^{L} d_k r_{12}^{a_k} s^{b_k} t^{e_k} e^{-s/2} \quad , \tag{3.6}$$

where $s = r_1 + r_2$ , and $t = r_1 - r_2$ . Here $r_1$ and $r_2$ are the scalar distances of the electrons from the nucleus. The electron-nucleus and the electron-electron cusp conditions can be satisfied by choosing the proper values for the coefficients. Because *all* the interparticle distances (for this simple two-electron case) are represented, very accurate descriptions of the He wavefunction may be obtained with relatively few terms. Moreover, this form may be readily generalized to larger systems, as has been done by Umrigar [4]. Despite a great improvement over single-particle expansions, a 1078-term Hylleraas function, which yields an energy of –2.903724375 hartrees, is surprisingly not all that much superior to a nine-term function which already yields –2.903464 hartrees.[12,13] On the other hand, by adding terms with powers of ln *s* and negative powers of *s*, one can obtain an energy of –2.9037243770326 hartrees with only 246 terms. The functional form clearly is very important. Recently, terms containing cosh *t* and sinh *t* have been added as well, [14] to model “in-out” correlation (the tendency for one electron to move away from the nucleus as the other approaches the nucleus).

We can distinguish between two broad classes of explicitly correlated wavefunctions: polynomials in $r_{ij}$ and other inter-body distances, and exponential or Jastrow forms [15, 16]

$$\Psi_{Corr} = e^{-U} \quad . \tag{3.8}$$

In this latter form, $U$ contains all the $r_{ij}$ and many-body dependencies, and the full wavefunction is given by

$$\Psi = \Psi_{Corr}\Psi_D \ , \tag{3.10}$$

the second factor being the determinant(s) discussed earlier. The Jastrow forms contain one or more parameters which can be used to represent the cusps. As an example, consider a commonly used form, the Pade-Jastrow function,

$$-U = \sum_{i=1}^{N-1}\sum_{j>i}^{N} \frac{a_1 r_{ij} + a_2 r_{ij}^2 + \ldots}{1 + b_1 r_{ij} + b_2 r_{ij}^2 + \ldots} \quad . \tag{3.12}$$

The general behavior of $e^{-U}$ is that it begins at unity (for $r_{ij}$ = 0) and asymptotically approaches a constant value for large $r_{ij}$. One can verify that the electron-electron cusp condition simply requires $a_1$ to be 1/2 for unlike spins and 1/4 for like spins. The linear Pade-Jastrow form has only one free parameter, namely $b_1$, with which to optimize the wavefunction. Creating a correlated trial wavefunction as above, by combining $\Psi_{Corr}$ with an SCF determinant, causes a global expansion of the electron density [17]. If we assume that the SCF density is relatively accurate to begin with, then one needs to re-scale this combined trial wavefunction to re-adjust the density. This can be accomplished simply by multiplying by an additional term, an *electron-nucleus* Jastrow function. If this Jastrow function is written with

$$-U = \sum_{i=1}^{N}\sum_{a=1}^{N_{nuc}} \frac{\lambda_1 r_{ia} + \lambda_2 r_{ia}^2 + \ldots}{1 + \nu_1 r_{ia} + \nu_2 r_{ia}^2 + \ldots} \quad , \tag{3.14}$$

then, in analogy to the electron-electron function, $\lambda_1$ is determined by the cusp condition. More general forms of $U$ have been explored in the literature, including ones with electron-electron-nucleus terms, and powers of $s$ and $t$.[4,18] These lead to greatly improved functional forms as judged by their rates of convergence and VMC variances.

A different approach towards building explicitly correlated wavefunctions is to abandon the distinction between the correlation part and the determinantal part. In such an approach one might try to approximate the exact wavefunction using a linear combination of many-body terms [19]. A powerful approach is to use explicitly correlated Gaussians [20, 21]. Such a functional form, although very accurate for few electron systems, is difficult to use with more electrons (due to the rapidly expanding number of terms needed), and does not exploit the more general (integral) capabilities of VMC. A form amenable to VMC computation is a linear expansion in correlated exponentials [22], which shows very good convergence properties.

For such correlated wavefunctions, one can optimize the molecular orbital coefficients, the atomic orbital exponents, the Jastrow parameters, and any other non-linear parameters. Clearly, practical limitations will be reached for very large systems; but such optimization is generally practical for moderate-sized systems, and has been done for several. The next section discusses the means by which such optimizations may be performed by Monte Carlo.

## IV. OPTIMIZATION OF A TRIAL WAVEFUNCTION USING VMC

In the previous sections we have seen how VMC can be used to estimate expectation values of an operator given a trial wavefunction $\Psi_T(\mathbf{R})$. Despite the "logic" used to select trial wavefunction forms, as described in the previous section, for a realistic system it is extremely difficult to know *a priori* the proper analytical form. Thus it is a challenge to generate a good trial wavefunction "out of thin air." Of course, one can choose a trial form which depends on a number of parameters; then, within this "family" one would like to be able to choose the "best" wavefunction. Moreover, if possible one would like to be able to arbitrarily improve the "goodness" of the description, in order to approach the true wavefunction. It is clear that we need first to clarify what we mean by a "good" $\Psi_T(\mathbf{R})$, otherwise the problem is ill-posed.

If for the moment we restrict our attention to trial wavefunctions that approximate the lowest state of a given symmetry, a possible (and surely the most-used in practice) criterion of goodness is provided by the beloved variational principle

$$E[\Psi_T] = \langle H \rangle = \frac{\int \Psi_T^*(\mathbf{R}) H \Psi_T(\mathbf{R}) d\mathbf{R}}{\int \Psi_T^*(\mathbf{R}) \Psi_T(\mathbf{R}) d\mathbf{R}} \geq E_0 \quad , \tag{4.2}$$

which provides an *upper* bound to the exact ground state energy $E_0$. According to this principle, $\Psi_{T1}(\mathbf{R})$ is better than $\Psi_{T2}(\mathbf{R})$ if $E[\Psi_{T1}] < E[\Psi_{T2}]$. This is a reasonable criterion since, in this sense, the exact wavefunction $\phi_0(\mathbf{R})$ is "better" than any other function. However, before discussing the practical application of Eq. 4.2, we point out that the variational principle is not the only available criterion of goodness for trial wavefunctions. We will see more about this later on; for now we simply point out that other criteria may well have advantages. A second remark worth making here is that the condition $\langle H \rangle \rightarrow E_0$ is not sufficient to guarantee that $\Psi_T \rightarrow \phi_0$ (i.e., for all points). This implies, for example, that although the energy is improving, expectation values of some other operators might appear to converge to an incorrect value, or may not converge at all (although this latter is a rather pathological case, and not very common for reasonably-well behaved trial wavefunctions).

Let us chose a family of trial wavefunctions $\Psi_T(\mathbf{R};\mathbf{c})$, where $\mathbf{c}$ is a vector of parameters $\mathbf{c} \equiv \{c_1, c_2, \ldots, c_n\}$ on which the wavefunction depends parametrically. The best function in the family is selected by solving the problem

$$\min_{\mathbf{c}} \langle H(\mathbf{c}) \rangle. \tag{4.4}$$

In most standard *ab initio* approaches, the parameters to minimize are the linear coefficients of the expansion of the wavefunction in some basis set. To make the problem tractable one is usually forced to choose a basis set for which the integrals of Eq. 4.2 are analytically computable. However, as we have seen, it is practical to use very accurate explicitly correlated wavefunctions with VMC.

A typical VMC computation to estimate the energy or other expectation values for a given $\Psi_T(\mathbf{R})$ might involve the calculation of the wavefunction value, gradient and Laplacian at several millions points distributed in configuration space. Computationally this is the most expensive part. So a desirable feature of $\Psi_T(\mathbf{R})$, from the point of view of Monte Carlo, is its compactness. It would be highly impractical to use a trial wavefunction represented, for example, as a CI expansion of thousands (or more) of Slater determinants.

For this reason, optimization of the wavefunction is a crucial point in VMC (and likewise in QMC as well; see the next chapter). To do this optimization well, and to allow a compact representation of the wavefunction, it is absolutely necessary to optimize the nonlinear parameters, most notably the orbital exponents and the parameters in the correlation factor(s), in addition to the linear coefficients.

### A. Energy optimization

The naive application of the variational principle to the optimization problem is limited by the statistical uncertainty inherent in every Monte Carlo calculation. The magnitude of this statistical error has a great impact on the convergence of the optimization, and on the ability to find the optimal parameter set as well.

Consider the following algorithm to optimize a one parameter function $\Psi_T(\mathbf{R},\alpha)$.

0) Choose the initial parameter $\alpha^{old}$<br>
1) Do a VMC run to estimate $E^{old} = \langle H(\alpha^{old})\rangle$<br>
2) Repeat:<br>
3)   Somehow select a new parameter $\alpha^{new}$ (maybe $\alpha^{new} = \alpha^{old} + \zeta$)<br>
4)   Do a VMC run to estimate $E^{new} = \langle H(\alpha^{new})\rangle$<br>
5)   If ($E^{new} < E^{old}$) $\alpha^{old} = \alpha^{new}$; $E^{old} = E^{new}$;<br>
6) Until the energy no longer diminishes.

Because the energy $E^{new}$ is only an *estimate* of the true expectation value $\langle H(\alpha^{new})\rangle$, step 5 is a weak point of this algorithm. Using a 95% confidence interval we can only say that

$$\begin{aligned} E^{old} - 2\sigma^{old} \le \left\langle H(\alpha^{old})\right\rangle \le E^{old} + 2\sigma^{old} \\ E^{new} - 2\sigma^{new} \le \left\langle H(\alpha^{new})\right\rangle \le E^{new} + 2\sigma^{new} \end{aligned} \quad . \tag{4.6}$$

Thus it is clear that sometimes we cannot decide which of $E^{old}$ and $E^{new}$ is the lower, and thus which parameter is the better.

### B. Correlated sampling

One way to treat this problem begins by noting that during the optimization we are not really interested in the energies associated with the different wavefunctions, but only in the energy differences. A key observation is that one can use the *same* random walk to estimate the energies of different wavefunctions, and that such energy estimates will be statistically correlated. Thus, their difference can be estimated with much greater precision than can be the energies themselves. (Loosely speaking, a part of the fluctuations that result from the particular walk chosen will cancel when computing the difference.)

Let us then proceed to estimate the energies of a whole set of wavefunctions with a single VMC run. For simplicity, let us consider a family of $K$ functions which depend on a single parameter:

$$\left\{\Psi(\mathbf{R},\alpha_k)\right\}_{k=1}^{K} \equiv \left\{\Psi(\mathbf{R},\alpha_1),\Psi(\mathbf{R},\alpha_2),\ldots,\Psi(\mathbf{R},\alpha_K)\right\} \ . \tag{4.8}$$

One can estimate the energy of *all* the $\Psi_k$=$\Psi(\mathbf{R},\alpha_K)$ by VMC simulation, simply by sampling from the function $\Psi_0$= $\Psi(\mathbf{R},\alpha_0)$ and noting that

$$\left\langle H(\alpha_k)\right\rangle = \frac{\int \Psi_k H \Psi_k}{\int \Psi_k \Psi_k} = \frac{\int \Psi_0^2 \left(\frac{\Psi_k}{\Psi_0}\right)^2 \frac{H\Psi_k}{\Psi_k} d\mathbf{R}}{\int \Psi_0^2 \left(\frac{\Psi_k}{\Psi_0}\right)^2 d\mathbf{R}} \cong \frac{\sum_i \omega_k^2(\mathbf{R}_i) E_{L,k}(\mathbf{R}_i)}{\sum_i \omega_k^2(\mathbf{R}_i)} = E_k \ , \tag{4.10}$$

where the points $\mathbf{R}_i$ are distributed according to $\Psi_0^2$, $E_{L,k}$ is the local energy of the function $\Psi_k$, and the reweighting factor $\omega_k^2$ is defined as the square of the ratio of the two wavefunctions. After only a single run we can choose the best parameter. This method is effective only if the functions $\Psi_k$ are not too different from $\Psi_0$; that is, if the reweighting factors are not too small or too large, but rather are of the order of magnitude of unity. Otherwise a few points will dominate the sum, reducing the statistical quality of the sample and resulting in a larger error.

This method, called sometimes "differential Monte Carlo" [23] has been used in the past to compute energy differences coming from small perturbations of the Hamiltonian, or from small changes of the geometry of a molecule (i.e. for computing potential energy surfaces). The main disadvantage of this method is that it becomes more and more cumbersome to apply as the number of parameters increases. One can see this by noting that we are, in essence, evaluating the energy on a grid in *parameter space*. Using a function that depends on say three parameters, and choosing 10 possible values for each parameter, we must evaluate $10^3$=1000 energy points.

## C. Optimization using derivatives

Although the direct application of correlated sampling can be inefficient in a high dimensional parameter space, it can be fruitfully combined with a standard minimization technique such as the steepest descent method [24]. Consider again the trial wavefunction $\Psi_T(\mathbf{R};c)$ containing a parametric dependence on $\mathbf{c}$. Again we construct $K$ "perturbed" functions, each one now differing from the reference function by the value of a single parameter,

$$\Psi_k = \Psi_T(\mathbf{R}, \mathbf{c}_k), \qquad k = 1,2,\ldots K \tag{4.12}$$

where

$$\mathbf{c}_k = \{c_1, c_2, \ldots, c_k + \delta c_k, \ldots, c_K\} \ . \tag{4.14}$$

Once again defining the weights

$$\omega_k^2 = \left(\frac{\Psi_k}{\Psi_0}\right)^2 \ , \tag{4.16}$$

the variational energies of the various functions $\Psi_k$ can be evaluated as in Eq. 4.10 using only the unperturbed wavefunction $\Psi_0$. If the variations are small, we can numerically estimates the partial derivatives of the energy with respect to the parameters using a finite difference formula

$$\frac{\partial E}{\partial c_k} \cong \frac{\langle H(\mathbf{c}_k)\rangle - \langle H(\mathbf{c}_0)\rangle}{\delta c_k} = \frac{E_k - E_0}{\delta c_k} \ . \tag{4.18}$$

Equipped with such a method to estimate the gradient of the energy with respect to the parameters, we can describe a simple steepest descent algorithm that tries to optimize the parameters in a single VMC run.

```
0) Choose the initial parameters c and the magnitude of the
     perturbations δc_k
1) Repeat:
2)   Estimate E0 = <H(c)> and Ek = <H(ck)> (averaging over a block)
3)   Estimate the gradient ∇c = {∂E/∂c1,∂E/∂c2,...,∂E/∂cK}
4)   Update the parameter vector c ← c − s·∇c where s is a step size
     vector;
5) Until the energy no longer diminishes.
```

It is also possible to estimate the derivative of the wavefunction with respect to a parameter *without* resorting to the finite difference approximation. Consider again our one-parameter wavefunction $\Psi(\alpha)\equiv\Psi_T(\mathbf{R};\alpha)$. We would like to express the expectation value of the derivative of the energy in a form amenable to computation with VMC. By differentiating the expression for the energy we get

$$\frac{\partial\langle H(\alpha)\rangle}{\partial\alpha}=\frac{2\int\frac{\partial\Psi(\alpha)}{\partial\alpha}H\Psi(\alpha)d\mathbf{R}}{\int\Psi^2(\alpha)d\mathbf{R}}-\frac{2\int\Psi(\alpha)H\Psi(\alpha)d\mathbf{R}}{\int\Psi^2(\alpha)d\mathbf{R}}\frac{\int\frac{\partial\Psi(\alpha)}{\partial\alpha}\Psi(\alpha)d\mathbf{R}}{\int\Psi^2(\alpha)d\mathbf{R}}\quad,\qquad(4.20)$$

and using the trick of multiplying and dividing inside the integral by our probability distribution $\Psi^2(\alpha)$ we obtain

$$\frac{\partial\langle H(\alpha)\rangle}{\partial\alpha}=\left\langle\frac{\partial}{\partial\alpha}\ln|\Psi(\alpha)|^2E_L(\alpha)\right\rangle-\left\langle\frac{\partial}{\partial\alpha}\ln|\Psi(\alpha)|^2\right\rangle\langle E_L(\alpha)\rangle\quad.\qquad(4.22)$$

This means that we can estimate the derivative of the energy with respect to $\alpha$ exactly, without using finite difference methods. Still, the computational power needed to optimize more than a few parameters remains large, even including the savings from correlated sampling, because the fluctuations on this quantity can be large.

### D. Optimization of the variance of the local energy

The statistical error in any VMC estimate of the energy obtained using $N$ sample points is related to the fluctuation of the local energy function. Specifically,

$$E=\langle E_L\rangle\cong\frac{1}{N}\sum_{i=1}^{N}E_L(\mathbf{R}_i)\pm\frac{\sigma(E_L)}{\sqrt{N}}\quad,\qquad(4.24)$$

where

$$\sigma^2(E_L) = \left\langle E_L^2 \right\rangle - \left\langle E_L \right\rangle^2 \quad . \tag{4.26}$$

Suppose we knew the exact wavefunction: in that case the local energy would be a constant over all configuration space, and the error in its estimate would be rigorously zero. Similarly, if we could obtain very good wavefunctions the fluctuations of the local energy would be very small. (This is an amusing property of VMC: the more accurate the trial wavefunction, the "easier" the simulations. However, if accuracy of the wavefunction entails evaluation of massive expressions, such an approach makes life "harder" and is counterproductive.) The error thus provides a measure of the quality of the trial wavefunction. This leads naturally to the idea of optimizing the wavefunction to *minimize the error* of the estimate.

We have previously shown that $\left\langle E_L \right\rangle = \left\langle H \right\rangle$; it is not difficult to show also, using the hermiticity of the Hamiltonian operator, that $\left\langle E_L^2 \right\rangle = \left\langle H^2 \right\rangle$, so that

$$\left\langle H^2 \right\rangle = \left\langle E_L^2 \right\rangle \cong \frac{1}{N}\sum_{i=1}^{N} E_L^2(\mathbf{R}_i) \quad . \tag{4.28}$$

We can now rewrite Eq. 4.26 as

$$\left\langle H^2 \right\rangle = \left\langle E_L^2 \right\rangle \cong \frac{1}{N}\sum_{i=1}^{N} E_L^2(\mathbf{R}_i) \quad , \tag{4.30}$$

and noting that $\sigma^2(H)$ is positive we can formulate the following variational principle:
*For a stationary state, the quantity* $\sigma^2(H)$ *is at a local minimum, and is equal to zero.*
Optimizing a wavefunction using this principle means solving the problem

$$\min_{\mathbf{c}} \sigma^2(H(\mathbf{c})) \quad . \tag{4.32}$$

This version of the variational principle, although very old, is quite possibly not known to readers more familiar with standard *ab-initio* methods; so before applying it to Monte Carlo optimization, let us discuss its properties. The reason it is little known and rarely applied in standard computational methods, is that Eq. 4.30 involves the expectation value of the square of the Hamiltonian, a very difficult quantity to compute analytically. A second problem is that for some (albeit, very poor) trial wavefunctions this quantity might diverge. A third problem is that sometimes, again for very poor wavefunctions, a local minimum can not be found (for example, the reader might easily check that using a single Gaussian function to describe the ground state of the hydrogen atom, a minimum of $\sigma^2(H)$ does not exist). The main point, i.e. the difficulty in its computation, is not an issue in VMC, while the other two problems are not encountered in practical calculations. Moreover, there are advantages in using the "sigma variational principle" instead of the more common "energy variational principle." The first advantage is that we know exactly the minimum possible value, namely zero; and this value is obtained only for the exact wavefunction. The minimum value of the energy, on the other hand, is not known. A second quality is that this variational principal can also be used for excited states, and not just for the first state of a given symmetry: $\sigma^2(H)$ is a minimum for *any* state. This is in contrast with the energy variational principle, where for an arbitrary stationary state the energy is not necessarily required to be a local minimum.

As a byproduct of the computation of $\sigma^2(H)$, after a VMC simulation we have both an *upper* bound to the energy (the usual $\langle H\rangle \geq E_0$), and a *lower* bound. The latter is a quantity that common techniques do not give. Under reasonable assumptions [25, 26]

$$\langle H\rangle \geq E_0 \geq \langle H\rangle - \sqrt{\langle H^2\rangle - \langle H\rangle^2} = \langle H\rangle - \sigma(H) \quad . \tag{4.34}$$

However, because the upper bound depends quadratically on the error of the trial wavefunction, while the lower bound depends only linearly on it, the lower bound is usually a significantly worse approximation to the exact energy than is the upper bound.

One might thus consider optimizing the wavefunction to *maximize the lower bound*. Due to the presence of $\langle H\rangle$ in Eq. 4.34, however, this usually leads to the same problems we have seen when trying to directly optimize the energy. If one instead decides to minimize the distance between the upper and the lower bounds, the sigma variational principle is recovered. (In practice, one optimizes $\sigma^2(H)$, not $\sigma(H)$, since they have the same minima, and one avoids computing a square root.)

Instead of minimizing $\sigma^2(H)$, sometimes it is preferable to optimize a related quantity, namely the second moment of the local energy with respect to a fixed (or reference) energy $E_R$,

$$\mu_{E_R}^2(H) = \frac{\int \Psi(\mathbf{R})\left(E_L(\mathbf{R}) - E_R\right)^2 \Psi(\mathbf{R}) d\mathbf{R}}{\int \Psi(\mathbf{R})\Psi(\mathbf{R}) d\mathbf{R}} \ . \tag{4.36}$$

The constant $E_R$ should be close to the energy of the state being sought, although the optimization does not depend strongly on its value. Minimizing this quantity (which many authors call $\sigma^2(H)$ without making the distinction) is *almost* equivalent to minimizing the variance. A little algebra shows that

$$\mu_{E_R}^2(H) = \sigma^2(H) + \left(\langle H \rangle - E_R\right)^2 \ . \tag{4.38}$$

The second term on the right hand side can be made small. In fact, having the second term present allows one to balance minimization of the variance with minimization of the energy, the value of the reference energy dictating the nature of the balance. This last form of minimization procedure is preferable when looking for an excited state, since by a careful choice of the reference energy one can prevent falling to the ground state.

If it were only for the above properties, optimization using the variance would not be as widespread as it is today in VMC calculations. The property that improves the optimization by orders of magnitude is what we might call *local boundedness*. In practice, all estimates in VMC are done using a finite set of configuration points (the *walkers* $\mathbf{R}$); however, for a finite number of such points, the estimate of the *energy* can be unbounded from below *with respect to variations of the parameters*, while the *variance* is always bounded. One can demonstrate this with a toy problem in which there are only two (fixed) walkers being used for the optimization. It is relatively easy to find such a pair of walkers for which the local energy surface in parameter space has singularities, i.e., values for the parameters at which the local energy goes to minus infinity; the variance, of course, must always be positive.

### E. Optimization with a fixed ensemble

Since $\sigma^2(H)$ is a bounded quantity, even when estimated with a finite number of points, it becomes possible to use a *fixed*, (relatively) small number of walkers to estimate the necessary integrals during the optimization.[27,28,29] Suppose then that we have an ensemble of $N$ walkers distributed according to $\Psi(\mathbf{c_0})$, where $\mathbf{c_0}$ is some initial set of parameters. The optimization algorithm for a fixed ensemble will use these $N$ points to estimate the integrals even when, later during the minimization, the *parameter vector* is modified. This is achieved by the same reweighting method we previously discussed for correlated sampling; in fact, we are using correlated sampling to compute expectation values for different values of **c**, only now with a fixed ensemble. Because the ensemble is fixed, the optimization process has become completely deterministic (apart from the input of random configuration points). One can thus choose whatever optimization procedure one is fond of (e.g., steepest descent, Powell's method, simulated annealing, etc...), and not have to worry about the uncertainties for now. We repeat here the relevant reweighting formulas

$$\left\langle H(\mathbf{c}^{new})\right\rangle \cong \frac{\sum_{i=1}^{N} \omega_{new}^2 E_L(\mathbf{R}_i)}{\sum_{i=1}^{N} \omega_{new}^2},$$
$$\left\langle H^2(\mathbf{c}^{new})\right\rangle \cong \frac{\sum_{i=1}^{N} \omega_{new}^2 E_L^2(\mathbf{R}_i)}{\sum_{i=1}^{N} \omega_{new}^2} \qquad (4.40)$$

where the $\omega$'s are the ratios of new to original trial wavefunctions, and the $E_L$'s are evaluated with respect to the original wavefunction (parameters).

After the optimization procedure has found the best set of parameters (for the given set of walkers) it is a good idea to run a new VMC simulation, to produce an ensemble of walkers distributed according to the *new* wavefunction; then one should reoptimize again, until self consistency is attained—i.e. until $\sigma^2(H)$, or $\mu_{E_R}^2(\mathbf{R})$, does not change significantly anymore. Another practical point: it is sometimes found that a more numerically stable optimization can be achieved by initially setting all the weights to unity (see Eq. 4.10). This is clearly an approximation, with the added *dis*advantage that it can prevent movements of the nodes during the optimization. (This can be seen by noting that, on approaching a node, the local energy diverges; but, in the weighted optimization, this is counterbalanced by a vanishing weight.) Nevertheless, the approximation is only to the *direction* of motion in parameter space, while the approach can frequently have the advantage of faster convergence, particularly when one is starting reasonably far from the optimum set of parameters.

With ensembles consisting of a few thousand points, this algorithm can reasonably be expected to optimize a hundred or so nonlinear parameters. Usually only two or three reoptimization iterations are necessary to reach convergence. Attempting to directly optimize the *energy* using this algorithm, however, might require one to two orders of magnitude more walkers, and a correspondingly increased computational cost.

Because ensuring that cusp conditions are met helps to minimize fluctuations, at least in the energy, optimization will be aided by satisfying these conditions. However, note that optimization also can be used to *obtain* good cusps; this can be done by including a *penalty* factor into the minimization procedure to force the various parameters of the trial wavefunction to build the exact or nearly exact cusps.

It is also worth mention that other functionals can (and have) been used for the minimization process. One can consider modifications of the variance functional [30]; or even using a completely different approach to solving the problem [31]. Consider, for example, the following minimization,

$$\min_{\mathbf{c}}(\varepsilon(\mathbf{c})) = \min_{\mathbf{c}}\left(\int \left(\Psi_T(\mathbf{R}) - \phi_0(\mathbf{R})\right)^2 d\mathbf{R}\right) \quad , \tag{4.42}$$

i.e., minimizing the least square error between the trial and the exact wavefunction. Although we do not know the exact wavefunction, one *can* sample it using quantum Monte Carlo techniques (such as those discussed in the next chapter).

## V. ACCELERATION TECHNIQUES

Having obtained the best possible trial wavefunction we are capable of, one is perhaps now ready to compute chemical properties. A good trial wavefunction will provide us with not only good expectation values but, as discussed earlier, also a low statistical error. However, the efficiency of the VMC method depends not only on the quality of the trial wavefunction used, but also on the *procedure* for generating the distribution to be sampled (i.e., for generating the trial wavefunction squared). We have seen that there are a number of possible algorithms, ranging from simple Metropolis, to Langevin, to generalized

Metropolis schemes, that can all be used to obtain the same distribution. The latter improve efficiency over the simple Metropolis scheme. The reason, in part, is that in all these approaches, the Markov chain (or random walk) nature of the generation procedure leads to correlations between sample points. Such correlations, depending on their magnitude, can greatly affect the rate of convergence.

Although we have identified the (mathematical) source of inefficiency to be the correlations present in the Markov chain, there is a deeper underlying physical cause. This is a fundamental problem of many simulations, namely that of multiple scales. Specifically it is this: Although VMC scales well with $M$, it scales much more poorly with atomic number $Z$. A common estimate is that computational time $T$ rises as $Z^{5.5}$ [32]. Upon reflection it is clear that the problem is the differing time (as well as distance and energy) scales for core and valence electrons. As $Z$ increases, this range of time scales increases as well. In fact, $Z \to \infty$ is in many ways analogous to a critical point.[33,34] As in critical slowing down, an unending hierarchy of time scales ensues as the critical point is approached. This is the problem that must be addressed. In critical phenomena the problem has been effectively treated through a class of acceleration methods, particularly so-called cluster acceleration methods. These take advantage of the self-similarity that occurs in the vicinity of a critical point. In electronic structure problems there do exist analogous critical points.[34, 35] However, typical systems are not near the regime of those critical points. Instead, a common way to address the large-$Z$ problem has been through the use of effective-core potentials or pseudopotentials which eliminate the large $Z$ at the outset. This is the standard approach in quantum chemistry and solid-state physics. It is also becoming widely (and effectively) used in quantum Monte Carlo simulations.[7]

However, instead of (or in addition to) effective core potentials, we can cope with the differing time scales at least in part by purely Monte Carlo means. The Markov chain correlations can be vastly different depending on the explicit sampling algorithm employed. Several methods have been explored to change the sampling (or, again, the details of the Markov chain) in conjunction with modifications of the Metropolis scheme. These have met with differing degrees of success.

For example, one can render the attempted moves position-dependent. This, for example, might allow one to move valence electrons further than core electrons, corresponding to their great available "phase space." However, naively doing so does not recognize that exchange symmetry does not allow us to know which electrons are core or which are valence; any *a priori* assumptions are likely to lead to a situation in which core electrons end up in the valence region after some time, and *vice versa*. There are a number of correct way to implement this idea, however. One allows step sizes to change dynamically, depending on relative positions, but entails the need for a modified coordinate system to maintain detailed balance.[36]. Another implementation of this idea does it straightforwardly, but explicitly prevents exchange by carrying out the VMC integration only over a subset of the full space. Because the subset is permutationally equivalent to the full space, the integrals are equal (up to a constant factor). Another very different approach, borrowed from high-energy theory, has been to modify the VMC dynamics at the level of the Fokker-Planck equation. This allows more rapid approach to time independent behavior while keeping the steady-state unchanged.[37] A very nice variant of this has recently been proposed by Mella et al. [38] In this algorithm the time step size of the Langevin transition matrix is allowed to be an explicit function of position, shrinking as one approaches either nodes of the wavefunction or two-body cusps. Since the functional dependence of the time step on position is put explicitly into a Gaussian transition matrix (rather than a standard Metropolis "box"), the forward and reverse transition matrices overlap, enabling one to preserve detailed balance. By having the time step shrink to a minimum at the electron-nuclear cusp, one effectively treats the core region differently from the valence region. Moreover, the time step behavior obtained near the nodes is *also* advantageous. Similar improvements to the Metropolis algorithm in the vicinity of cusps and nodes have been proposed previously by Umrigar and colleagues.[39] Yet other schemes exist for "accelerating" convergence. For example, one can radically change the algorithm by mixing a molecular dynamics approach with VMC [40].

To get a sense for this type of approach, we will discuss one of these in a little bit greater detail. This is the method of dividing the space. As a result of the antisymmetry of the electronic wavefunction there are multiple regions of ($3M$-dimensional) space which are equivalent. Specifically, up to a sign the value of the wavefunction is the same when any two coordinates representing like-spin electrons are interchanged. This results in $N_{\mathrm{up}}!N_{\mathrm{down}}!$ equivalent volumes or domains. We are *not* talking about *nodal volumes* here. The volumes we are distinguishing are the following: given a point $\mathbf{R}$ in configuration space there are another $N_{\mathrm{up}}!N_{\mathrm{down}}! - 1$ points generated by permutations of the indices. We can think of these points as belonging to different regions, or subspaces, of the full space. If we can explicitly construct such subspaces, then the integration over the entire $3M$-dimensional space is redundant, since for any operator $\hat{O}$ which is totally symmetric with respect to the exchange of two identical particles

$$\begin{aligned}\langle\hat{O}\rangle &= \int_{allspace} \Psi_T^*(\mathbf{R})\hat{O}\Psi_T(\mathbf{R})d\mathbf{R} \Big/ \int_{allspace} \Psi_T^*(\mathbf{R})\Psi_T(\mathbf{R})d\mathbf{R} \\ &= \int_{any\ subspace} \Psi_T^*(\mathbf{R})\hat{O}\Psi_T(\mathbf{R})d\mathbf{R} \Big/ \int_{any\ subspace} \Psi_T^*(\mathbf{R})\Psi_T(\mathbf{R})d\mathbf{R}\end{aligned}, \qquad (5.2)$$

meaning we only need to integrate over a single subspace. Note that such subspaces are *not* uniquely defined.

Now we will see how this simple fact can be used to our advantage, to help alleviate the multiple time scales problem. Let us concentrate on an atomic system; molecules can be treated by a simple extension of this approach [41]. Subdividing the space allows us to treat the electrons as "distinguishable" within the simulation. Unlike the standard algorithm in which electrons exchange, i.e. they can always cross the subspaces boundaries, in this approach—by integrating only over a subspace—we in effect enforce the boundaries, and constrain particles to stay within subspaces. Any starting configuration, e.g. random walker, is a single point in the $3M$-dimensional space, and thus resides in a single subspace. Subsequent moves need only enforce the boundaries by rejecting any attempts to cross them. Thus, by constructing subspaces such that electrons in the outer region of 3-space, away from the nucleus, stay far from the nucleus, and likewise electrons close-in, near the nucleus, stay close-in, we can assign different time steps to these different electrons. This allows them to explore their respective regions of configuration space with the most appropriate step sizes.

As the number of electrons increase, so does the number of equivalent subspaces and our freedom in choosing them. However, we can also combine some of the electrons (say into shells) and further increase efficiency [41]. All the electrons within the same shell can explore their entire "shell" space. This is more efficient because we avoid unnecessary rejections resulting from crossings of electrons having the same time scale.

## VI. RECENT RESULTS AND NEW DIRECTIONS

It is difficult, in such a short paper, to review all the published results and all the theoretical developments in the VMC field. Although the method has only recently become anything resembling widespread, it already does have a substantial literature, with more and more papers published every year. Another difficulty in reviewing "results" is that in many instances VMC is used only as a first step, that of generating a highly optimized trial wavefunction, to be used in a "diffusion" or "Green's function" quantum Monte Carlo simulation. This is, in fact, the major use of VMC, and the variational expectation values so obtained are not the point, and sometimes not even reported. To the extent that there are VMC "results" in their own right, these have been recently reported in QMC reviews. [42,7,43]. Here, in the spirit of Monte Carlo, we only *sample* the recent literature. We hope this will convey to the reader a feeling of how VMC is being used in actual works, and of the new directions.

Any method, particularly one just coming into the mainstream, needs benchmarks to be established: a series of calculations on "standard" systems, to be a reference when confronting untried systems. One paper that might serve this purpose is the work of Lüchow and Anderson [44], in which they calculate the ground state energies of atoms from Li to F, and the ground state energies and dissociation energies of the first row hydrides, using both VMC and QMC. They use simple standard trial wavefunctions (in the VMC sense) composed of a Slater determinant of near Hartree-Fock quality multiplied by a correlation factor of the kind used by Schmidt and Moskowitz [18], a generalized Jastrow factor. The correlation energies recovered with VMC range from 65% to 89% for atoms, while for the hydrides they range from 51% to 78%. Subsequent diffusion Monte Carlo simulations with these functions recovered up to 99% of the correlation energy. Along the same lines, Filippi and Umrigar [45] studied the first row homonuclear diatomics in a systematic way, using correlated wavefunctions composed of a determinantal part multiplied by a generalized Jastrow factor. They used both single and multi-configuration functions, recovering up to 94% of the correlation energy by VMC and up to 99% when the trial wavefunctions were used in a subsequent quantum Monte Carlo simulation.

As mentioned earlier, the treatment of heavy atoms can be approximated using a pseudopotential approach, both in standard *ab-initio* methods and in VMC as well. With this approach Flad and Dolg [46] studied the Hg atom and the $Hg_2$ molecule. They used a 20-valence-electron pseudopotential for the atom, while for the molecule they used a 2-valence-electron pseudopotential together with a core polarization potential, recovering 84% of the correlation energy. Also using a pseudopotential approach, Mitas and coworkers studied heavy atoms [47,48] with very good results, obtaining an electron affinity for the Sc atom in agreement with the experimental value, and values of the ionization potential and excitation energies for the Fe atom with an average error, with respect to the experimental value, of less that 0.15 eV. For still larger systems Grossman et al. studied the structure and stability of silicon [49] and carbon [50] clusters, ranging from a few to 20 atoms. Compared with standard quantum chemical approaches like HF, LDA and CCSD(T), Monte Carlo proved superior and efficient in predicting the most stable isomer. In the case of $Si_n$ for $n < 8$, the binding energies agree within 4% with experiment.

The VMC and QMC methods can be successfully applied to vibrational problems as well. For example, Blume et al. [51] calculated the vibrational frequency shift of HF molecules embedded in helium clusters as a function of cluster size, for up to 198 helium atoms. They obtained good agreement with experimental results. Quantum clusters have also been studied by Rick et al. [52] and Barnett and Whaley [53], with good agreement between the calculated energies and the exact values.

On the theoretical/algorithmic development side, Bueckert et al. [54] showed how to estimate the relativistic energy of atoms and molecules in VMC, without having to resort to perturbation theoretic approximations. Alexander et al. [55] used VMC to compute cross sections for the elastic and inelastic scattering of fast electrons and X-rays by $H_2$. Novel trial wavefunction optimization schemes has been proposed by Huang and Cao [56] in which they sample $\Psi_T^2(E_L - E_R)^2$, and by Tanaka [57] who introduces a fictitious Lagrangian, resembling that in the Car-Parrinello method, to be used in the simultaneous optimization of the wavefunction parameters and in the geometry optimization. This brief list is, as mentioned, only a sample.

## VII. CONCLUSIONS

We hope to have convinced the reader that the VMC approach to obtaining quantum expectation values of interest in both chemical and physical problems is a very powerful one. We believe that it, in combination with fully quantum Monte Carlo procedures, will be the preferred choice in the near future, for many of the calculations performed these days by more traditional non-stochastic means. VMC is the first, and a necessary step, toward a complete quantum simulation of a system. It has the very desirable feature (often a rare one) that it can be learned, implemented, and tested in a short period of time. (It is now part of the folklore in the quantum Monte Carlo community, that the original code, written by J. B. Anderson and used for his first QMC paper, was only 77 lines of FORTRAN code.) We hope our readers will be inspired to write their own (toy or otherwise) VMC code; possibly thereby contributing to and enlarging the growing Monte Carlo community.

## VIII. REFERENCES

[1] W. L. McMillan, *Phys. Rev.* **138**, A442 (1965).

[2] D. Ceperley, G. V. Chester, and M. H. Kalos, *Phys. Rev. B* **16**, 3081 (1977).

[3] R. L. Coldwell, *Int. J. Quantum Chem. Symp.* **11**, 215 (1977).

[4] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, *Phys. Rev. Lett.* **60**, 1719 (1988).

[5] S. Huang, Z. Sun and W. A. Lester, Jr., *J. Chem. Phys.* **92**, 597 (1990).

[6] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).

[7] B. L. Hammond, W. A. Lester, Jr., and P. J. Reynolds, *Monte Carlo methods in Ab initio Quantum Chemistry*, World Scientific, Singapore, 1994.

[8] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, Jr., *J. Chem. Phys.* **77**, 5593 (1982).

[9] J. D. Morgan III, *Numerical Determination of the Electronic Structure of Atoms, Diatomic and Polyatomic Molecules*, M. Defranceschi and J. Delhalle Eds.. NATO ASI Series C: Mathematical and Physical Sciences, Vol 271 (Kluwer, Dordrecht 1989).

[10] T. Kato, *Comm. Pure Appl. Math.* **10,** 151 (1957).

[11] V. Fock, *Izv. Akad. Nauk. SSSR, Ser. Fiz.* **18**, 161 (1954).

[12] E. A. Hylleraas, *Z. Physik* **54**, 347 (1929).

[13] C. L. Pekeris, *Phys. Rev.* **112**, 1649 (1958).

[14] J. D. Baker, J. D. Morgan, D. E. Freund, and R. N. Hill, *Phys. Rev. A* **43**, 1247 (1990).

[15] R. Jastrow, *Phys. Rev.* **98**, 1479 (1955).

[16] Z. Sun, P. J. Reynolds, R. K. Owen, and W. A. Lester, Jr., *Theor. Chim. Acta*, **75**, 353 (1989).

[17] R. N. Barnett, P. J. Reynolds, and W. A. Lester, Jr., *J. Chem. Phys.* **82**, 2700 (1985).

[18] K. E. Schmidt and J. Moskowitz, *J. Chem. Phys.* **93**, 4178 (1990).

[19] J. Rychlewski, *int. J. Quantum Chem.* **49**, 477 (1994).

[20] S. A. Alexander, H. J. Monkhorst, R. Roeland and K. Szalewicz, *J. Chem. Phys.* **93**, 4230 (1990)

[21] W. Cencek and J. Rychlewski, *J. Chem. Phys.* **98**, 1252 (1993)

[22] D. Bressanini, M. Mella, and G. Morosi, *Chem. Phys. Lett.* **240**, 566 (1995).

[23] B. H. Wells, *Chem. Phys. Lett.* **115**, 89 (1985).

[24] S. Huang, Z. Sun, and W. A. Lester, Jr., J. Chem. Phys. **92**, 597 (1990).

[25] D. H. Weinstein, *Proc. Natl. Acad. Sci. U.S.A.* **20**, 529 (1934).

[26] L. R. Pratt, *Phys. Rev. A* **40**, 6077 (1989).

[27] R. L. Coldwell and R. E. Lowther, *Int. J. Quantum Chem. Symp.* **12**, 329 (1978).

[28] H. Conroy, *J. Chem. Phys.* **47**, 5307 (1967).

[29] A. A. Frost, *J. Chem. Phys.* **10**, 242 (1942).

[30] S. A. Alexander, R. L. Coldwell, H. J. Monkhorst, and J. D. Morgan III, *J. Chem. Phys.* **95**, 6622 (1991).

[31] R. Bianchi, D. Bressanini, P. Cremaschi, M. Mella and G. Morosi, *Int. J. Quantum Chem.* **57**, 321 (1996).

[32] D. M. Ceperley, *J. Stat. Phys.* **43**, 815 (1986).

[33] P. J. Reynolds, in *Computational Physics and Cellular Automata*, A. Pires, D. P. Landau, and H. J. Herrmann, eds. p.144 (World Scientific, Singapore, 1990)

[34] P. Serra and S. Kais, *Phys Rev. Lett*. **77**, 466 (1996); *Phys. Rev. A* **55**, 238 (1997)

[35] P. Serra and S. Kais, *Chem. Phys. Lett*. **260**, 302 (1996); *J. Phys A: Math. Gen*. **30**, 1483 (1997)

[36] C. J. Umrigar, *Phys. Rev. Lett.* **71**, 408 (1993).

[37] P. J. Reynolds, *Int. J. Quantum Chem. Symp.* **24**, 679 (1990).

[38] M. Mella, A. Lüchow and J. B. Anderson, *Chem. Phys. Lett.* **265**, 467 (1996).

[39] C. J. Umrigar, M. P. Nightingale and K. J. Runge, *J. Chem. Phys.* **99**, 2865 (1993).

[40] D. Bressanini, M. Mella and G. Morosi, *in preparation*.

[41] D. Bressanini and P. J. Reynolds, *in preparation.*

[42] D. M. Ceperley and L. Mitas, *Advances in Chemical Physics XCIII*, I. Prigogine and S. A. Rice Eds, (Wiley, New York, 1996).

[43] W. A. Lester and B. L. Hammond, *Annual Review of Physical Chemistry* **41**, 283 (1990).

[44] A. Lüchow and J. B. Anderson, *J. Chem. Phys.* **105**, 7573 (1996).

[45] C. Filippi and C. J. Umrigar, *J. Chem. Phys.* **105**, 213 (1996).

[46] H.-J. Flad and M. Dolg, *J. Phys. Chem.* **100**, 6152 (1996).

[47] L. Mitas, *Phys. Rev. A* **49,** 4411 (1994).

[48] L. Mitas, *Comp. Phys. Comm.* **96**, 107 (1996).

[49] J. C. Grossman and L. Mitas, *Phys. Rev. Lett.* **74**, 1323 (1995).

[50] J. C. Grossman, L. Mitas and K. Raghavachari, *Phys. Rev. Lett.* **75**, 3870 (1995).

[51] D. Blume, M. Lewerenz, F. Huisken and M. Kaloudis, *J. Chem. Phys.* **105**, 8666 (1996).

[52] S. W. Rick, D. L. Lynch and J. D. Doll, *J. Chem. Phys.* **95**, 2506 (1991).

[53] R. N. Barnett and K. B. Whaley, *J. Chem. Phys.* **96**, 2953 (1992).

[54] H. Bueckert, S, Rothstein and J. Vrbik, *Chem. Phys. Lett.* **190**, 413 (1992).

[55] S. A. Alexander, R. L. Coldwell, R. E. Hoffmeyer and A. J. Thakkar, *Int. J. Quantum Chem. Symp.* **29**, 627 (1995).

[56] H. Huang and Z. Cao, *J. Chem. Phys.* **104**, 200 (1996).

[57] S. Tanaka, *J. Chem. Phys.* **100**, 7416 (1994).