

A 11 Quantum Monte Carlo methods for electronic structure problems

C. Filippi

Instituut-Lorentz, Universiteit Leiden

Leiden, The Netherlands

Contents

1	Introduction	2
2	Variational Monte Carlo	3
2.1	Generalized Metropolis algorithm	4
3	Quantum Monte Carlo wave functions	6
3.1	Jastrow-Slater wave functions	7
3.2	Spin-assigned wave functions	7
3.3	Wave function and divergences in the potential	9
3.4	Jastrow factor for atoms and molecules	11
3.5	Dynamic and static correlation	12
4	Wave function optimization	14
5	Diffusion Monte Carlo	16
5.1	Importance sampling	20
5.2	Expectation values and mixed estimators	21
5.3	Fixed-node approximation	22
5.4	Non-local pseudopotentials	23

1 Introduction

The many-electron Schrödinger equation gives an accurate description of materials at the quantum mechanical level but is an intractable $3N + 1$ dimensional partial differential equation, where the number of electrons N may be very large. To circumvent the problem of the high dimensionality, most computational quantum mechanical studies of large systems are based on simpler one-electron theories such as Kohn-Sham density-functional theory (DFT) and the Hartree-Fock (HF) approximation. These replace the electron-electron interactions by an effective potential, thereby reducing the problem to a set of one-electron equations.

Despite the successes of DFT in describing the electronic structure of complex molecules and solids, the treatment of electronic correlation within DFT is only approximate, sometimes leading to incorrect results for both strongly and weakly correlated systems. An important area of research within electronic structure theory is therefore the development of alternatives to density functional theory. Quantum Monte Carlo (QMC) methods are among the most successful of the post-DFT approaches and have yielded very accurate calculations of correlated properties for large molecules or solid systems, where conventional quantum chemistry methods are extremely difficult to apply. For these reasons, QMC is establishing itself as a unique tool for exploring electronic correlation in systems of interest to materials science and for obtaining conclusive answers in cases where density functional theory is shown to be inadequate.

One of the early and best-known examples of the important role played by QMC in the field of electronic structure theory is the computation of the correlation contribution to the local density approximation (the basic ingredient of DFT) through a correlated calculation for the electron gas [1]. Recently, QMC has been of further help in gaining a better understanding of DFT, either by establishing the performance of approximate exchange-correlation functionals or through the construction of accurate density functional quantities [2, 3]. Moreover, significant progress has been made in applying QMC techniques to problems of material science interest, such as binding properties of clusters [4], defects in semiconductors [6], structural energies and reactions at semiconductor surfaces [5], and excited states of molecules [7, 8] and solids [9] (for an exhaustive review of recent applications of QMC, see Ref. [10]).

In these notes, we focus on two particular quantum Monte Carlo approaches, the variational Monte Carlo and the projector Monte Carlo methods. They are both zero temperature methods and are used to calculate the properties of a single state (usually, the ground state) of a given Hamiltonian. Variational Monte Carlo is a simple application of the Metropolis Monte Carlo method with the non trivial complication that we do not have the analytical expression of the distribution to be sampled: the distribution is equal to the square of the wave function which must be constructed and optimized to be a good approximation of the state of interest. In projector Monte Carlo, the transition rule is chosen so that the asymptotic distribution is the exact ground state wave function. The diffusion Monte Carlo method is a particular way of projecting the ground state from an initial distribution and corresponds to a random walk given by a diffusion and a branching process. Fermi statistics and the collapse of the solution to the bosonic ground state pose a severe problem for projector Monte Carlo and can be addressed in an approximate but stable manner through the constraints introduced by the fixed-node algorithm.

While the features of QMC we discuss hold in general, we restrict the examples to atomic and molecular systems and do not cover the specifics of QMC calculations for solids which are given in Ref. [10]. Other useful sources for QMC are the simple introduction to Monte Carlo methods by Kalos and Whitlock [11], the book by Hammond, Lester and Reynolds [12] on the use of QMC in quantum chemistry, the book edited by Nightingale and Umrigar [13] and the

recent review on QMC methods and their applications [10]. We omit the description of finite temperature path integral Monte Carlo which can be found in Ref. [14].

Let us define the notation we adopt in these notes. We will assume that we have a non-relativistic system of N interacting electrons described by the Hamiltonian:

$$\mathcal{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 + \sum_{i=1}^N v_{\text{ext}}(\mathbf{r}_i) + \sum_{i<j}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad (1)$$

where we used atomic units ($\hbar = m = e = 1$). The external potential $v_{\text{ext}}(\mathbf{r})$ is given either by the bare electron-ion Coulomb potential $-Z/r$ where Z is the charge of the ion, or by a pseudopotential describing the ion plus the core electrons which have been eliminated from the calculation. We denote with \mathbf{R} the $3N$ particle coordinates, and with $\mathbf{x} = (\mathbf{r}, \sigma)$ the 3 spacial and 1 spin coordinates of one electron where $\sigma = \pm 1$.

2 Variational Monte Carlo

Variational Monte Carlo (VMC) is the simplest quantum Monte Carlo method and represents a generalization of classical Monte Carlo. VMC is used to compute expectation values of quantum mechanical operators on a given trial wave function and allows the use of any “computable” trial wave function without severe restrictions on its functional form. This must be contrasted to other traditional quantum chemical methods which express the wave functions in terms of products of single particle orbitals in order to perform the high-dimensional integrals in the various expectation values. VMC was first used by McMillan [15] to calculate the ground state properties of liquid ^4He and then generalized to fermionic systems by Ceperley *et al.* [16]. VMC derives the name from the variational theorem which states that, given a trial wave function $\Psi(\mathbf{R})$, the expectation value of the Hamiltonian with respect the trial wave function is an upper bound to the exact ground state energy E_0 :

$$E_V = \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\int d\mathbf{R} \Psi^*(\mathbf{R}) \mathcal{H} \Psi(\mathbf{R})}{\int d\mathbf{R} \Psi^*(\mathbf{R}) \Psi(\mathbf{R})} \geq E_0. \quad (2)$$

The variational energy E_V can be rewritten as an average value of the local energy $E_L(\mathbf{R})$ over the distribution $\rho(\mathbf{R})$

$$E_V = \int d\mathbf{R} \frac{\mathcal{H} \Psi(\mathbf{R})}{\Psi(\mathbf{R})} \frac{|\Psi(\mathbf{R})|^2}{\int d\mathbf{R} |\Psi(\mathbf{R})|^2} = \int d\mathbf{R} E_L(\mathbf{R}) \rho(\mathbf{R}) = \langle E_L(\mathbf{R}) \rangle_\rho, \quad (3)$$

where the local energy is defined as

$$E_L(\mathbf{R}) = \frac{\mathcal{H} \Psi(\mathbf{R})}{\Psi(\mathbf{R})}, \quad (4)$$

and the distribution ρ as

$$\rho(\mathbf{R}) = \frac{|\Psi(\mathbf{R})|^2}{\int d\mathbf{R} |\Psi(\mathbf{R})|^2}. \quad (5)$$

In VMC, we use the Metropolis method to sample the distribution $\rho(\mathbf{R})$ and compute the average of the local energy over M configurations sampled with the probability $\rho(\mathbf{R})$ as

$$E_V = \langle E_L(\mathbf{R}) \rangle_\rho \approx \frac{1}{M} \sum_{i=1}^M E_L(\mathbf{R}_i), \quad (6)$$

where the average is an approximation to the expectation value and becomes an equality in the limit of $M \rightarrow \infty$.

Which conditions should Ψ satisfy to be used within VMC? The wave function should be continuous, normalizable and have the proper symmetry. Very importantly, for the statistical error on the energy E_V to be meaningful, Ψ should yield a finite variance of the energy σ_V^2 ,

$$\sigma_V^2 = \frac{\langle \Psi | (\mathcal{H} - E_V)^2 | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \langle (E_L(\mathbf{R}) - E_V)^2 \rangle_\rho, \quad (7)$$

since the error as a function of the Monte Carlo steps behaves as

$$\text{err}(E_V) \sim \frac{\sigma_V}{\sqrt{M}}. \quad (8)$$

Note that, as the trial wave function approaches an exact eigenstate of the Hamiltonian \mathcal{H} , the local energy becomes constant and equal to the eigenvalue of \mathcal{H} on that state. Therefore, in this limit, the expectation value becomes the true energy and the variance goes to zero. This important property of VMC is known as the *zero variance principle*: as the trial wave function improves, the Monte Carlo estimate of the variational energy converges more rapidly as a function of the number of Monte Carlo configurations (in the limit of Ψ being the exact eigenstate, only one configuration in Eq. 6 is necessary to compute the energy). Because of the zero variance property, Monte Carlo computations of energies can be much more efficient than classical Monte Carlo calculations.

Even though we focused on the energy, VMC can be used to compute the expectation value of any operator \mathcal{X} which is diagonal in the \mathbf{R} representation. In the previous expressions, simply replace the Hamiltonian \mathcal{H} with \mathcal{X} and the local energy with $X(\mathbf{R}) = \mathcal{X}\Psi(\mathbf{R})/\Psi(\mathbf{R})$.

2.1 Generalized Metropolis algorithm

We will briefly review the Metropolis algorithm in the context of VMC. The Metropolis algorithm is a general method to sample an arbitrary probability distribution $\rho(\mathbf{R})$ and is an application of a Markov chain. In a Markov chain, one changes the state of the system randomly from an initial state \mathbf{R}_i to a final state \mathbf{R}_f according to the stochastic transition matrix $M(\mathbf{R}_f|\mathbf{R}_i)$ which satisfies

$$M(\mathbf{R}_f|\mathbf{R}_i) \geq 0 \quad \text{and} \quad \sum_f M(\mathbf{R}_f|\mathbf{R}_i) = 1. \quad (9)$$

To sample the desired distribution $\rho(\mathbf{R})$, one evolves the system by repeated application of a Markov matrix M which satisfies the *stationarity condition*

$$\sum_i M(\mathbf{R}_f|\mathbf{R}_i) \rho(\mathbf{R}_i) = \rho(\mathbf{R}_f),$$

for any state \mathbf{R}_f . The stationarity condition tells us that if we start from the desired distribution ρ , we will continue to sample ρ . Moreover, if the stochastic matrix M is ergodic, this condition ensures that any initial distribution will evolve to ρ under repeated applications of M . Therefore, ρ is the right-eigenvector of M with eigenvalue 1 and it is also the dominant eigenvector.

Note that both the Metropolis method and the projector Monte Carlo methods discussed later in section 5 are stochastic implementation of the power method for projecting out the dominant eigenvector of a matrix. In the Metropolis method, one constructs a matrix M whose dominant eigenstate is the desired known distribution while, in the projector Monte Carlo methods, one projects out the unknown eigenstate of a known matrix.

In practice, one imposes the more stringent *detailed balance* condition

$$M(\mathbf{R}_f|\mathbf{R}_i) \rho(\mathbf{R}_i) = M(\mathbf{R}_i|\mathbf{R}_f) \rho(\mathbf{R}_f) \quad (10)$$

which is a sufficient but not necessary condition to satisfy the stationarity condition as can be easily seen by summing both sides of the equation over \mathbf{R}_i and using Eq. 9. The transition M is then rewritten as the product of a proposal matrix T and the acceptance A :

$$M(\mathbf{R}_f|\mathbf{R}_i) = A(\mathbf{R}_f|\mathbf{R}_i) T(\mathbf{R}_f|\mathbf{R}_i), \quad (11)$$

where M and T are stochastic matrices but A is not. The detailed balance condition finally becomes

$$A(\mathbf{R}_f|\mathbf{R}_i) T(\mathbf{R}_f|\mathbf{R}_i) \rho(\mathbf{R}_i) = A(\mathbf{R}_i|\mathbf{R}_f) T(\mathbf{R}_i|\mathbf{R}_f) \rho(\mathbf{R}_f). \quad (12)$$

or equivalently

$$\frac{A(\mathbf{R}_f|\mathbf{R}_i)}{A(\mathbf{R}_i|\mathbf{R}_f)} = \frac{T(\mathbf{R}_i|\mathbf{R}_f) \rho(\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i) \rho(\mathbf{R}_i)}. \quad (13)$$

Clearly, for a given choice of T , there are infinitely many choices of A which satisfy this equation since any function

$$A(\mathbf{R}_f|\mathbf{R}_i) = F \left(\frac{T(\mathbf{R}_i|\mathbf{R}_f) \rho(\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i) \rho(\mathbf{R}_i)} \right) \quad (14)$$

with $F(x)/F(1/x) = x$ is appropriate. The choice originally made by Metropolis *et al.* [17] is

$$A(\mathbf{R}_f|\mathbf{R}_i) = \min \left\{ 1, \frac{T(\mathbf{R}_i|\mathbf{R}_f) \rho(\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i) \rho(\mathbf{R}_i)} \right\}, \quad (15)$$

and is the one which maximizes the acceptance.

In choosing the proposal matrix T , we observe that the Metropolis algorithm generates points which are sequentially correlated so that the effective number of independent observations in a Monte Carlo run of M steps is M/T_{corr} , where T_{corr} is the autocorrelation time of the observable of interest. Therefore, to achieve a fast evolution and reduce T_{corr} , the optimal T should yield a high acceptance and at the same time allow large proposed moves. The choice of T will of course be limited by the fact that we need to be able to sample T directly.

In the original Metropolis method, T is chosen to be a constant inside a cube and zero outside, and consequently drops out from the expression of the acceptance (Eq. 15). However, T does

not have to be symmetric and a generalized Metropolis procedure with a non-symmetric choice can actually be more efficient. If we limit the moves to a neighborhood of \mathbf{R}_i of linear dimension Δ , the acceptance will behave as

$$\frac{A(\mathbf{R}_f|\mathbf{R}_i)}{A(\mathbf{R}_i|\mathbf{R}_f)} = \frac{T(\mathbf{R}_i|\mathbf{R}_f) \rho(\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i) \rho(\mathbf{R}_i)} \approx 1 - \mathcal{O}(\Delta^m). \quad (16)$$

If T is symmetric as in the original Metropolis algorithm, $m = 1$ while other choices may give higher values of m . For example, a better choice (with $m = 2$) motivated by diffusion Monte Carlo is

$$T(\mathbf{R}_f|\mathbf{R}_i) = \mathcal{N} \exp \left[-\frac{(\mathbf{R}_f - \mathbf{R}_i - \mathbf{V}(\mathbf{R}_i)\tau)^2}{2\tau} \right] \quad \text{with} \quad \mathbf{V}(\mathbf{R}_i) = \frac{\nabla \Psi(\mathbf{R}_i)}{\Psi(\mathbf{R}_i)}. \quad (17)$$

A different choice of transition matrix particularly efficient for atomic and molecular systems is discussed in Ref. [18].

In short, the generalized Metropolis algorithm will consist of the following steps:

1. Choose the distribution $\rho(\mathbf{R})$ and the transition probability $T(\mathbf{R}_f|\mathbf{R}_i)$.
2. Initialize the configuration \mathbf{R}_i .
3. Advance the configuration from \mathbf{R}_i to \mathbf{R}_f :
 - a) Sample \mathbf{R}' from $T(\mathbf{R}'|\mathbf{R}_i)$.
 - b) Calculate the ratio

$$q = \frac{T(\mathbf{R}_i|\mathbf{R}') \rho(\mathbf{R}')}{T(\mathbf{R}'|\mathbf{R}_i) \rho(\mathbf{R}_i)}. \quad (18)$$

- c) Accept or reject: If $q > 1$ or $q > r$ where r is a uniformly distributed random number in $(0,1)$, set the new configuration $\mathbf{R}_f = \mathbf{R}'$. Otherwise, set $\mathbf{R}_f = \mathbf{R}_i$.
4. Throw away the first κ configurations corresponding to the equilibration time.
5. Collect the averages and block them to obtain the error bars.

Two final comments on the Metropolis algorithm. First, the distribution $\rho(\mathbf{R})$ does not have to be normalized since only ratios enter in the acceptance. Therefore, it is possible to sample the square of complex wave functions (Eq. 5) whose normalization we do not know (see section 3.1). Second, if M_1, M_2, \dots, M_n are matrices which satisfy the stationarity condition, the matrix $M = \prod_{i=1}^n M_i$ also satisfies the stationarity condition. Consequently, particles can be moved one at the time, a necessary feature as the system size grows since the size of the move would need to be decreased to have a reasonable acceptance of a move of all particles.

3 Quantum Monte Carlo wave functions

The use of VMC to compute the expectation values of quantum mechanical operators allows great freedom in the choice of the trial wave function which on the other hand determines the accuracy as well as the efficiency of the calculation. Therefore, the form of wave function should yield accurate results while being compact and easy to evaluate.

The ingredients entering in the wave function most commonly used in quantum Monte Carlo can be understood by inspecting the advantages and limitations of traditional quantum chemistry approaches. Methods such as configuration interaction (CI) expand the many body wave function in a linear combination of Slater determinants of single-particle spin-orbitals. This form allows the evaluation of the high-dimensional integrals in all expectation values but the convergence of the expansion is very slow, in part because of the difficulty in describing the cusps which occur as two electrons approach each other (see section 3.3). Quantum Monte Carlo uses a much more compact representation of the wave function which is usually given by a sum of few determinants (tens and not millions like in a CI calculation) multiplied by a component which can exactly impose the cusps at the inter-particle coalescence points.

3.1 Jastrow-Slater wave functions

The trial wave functions commonly used in quantum Monte Carlo methods for electronic structure problems are of the Jastrow-Slater form, that is the product of a sum of determinants of single-particle orbitals and a Jastrow correlation factor

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \mathcal{J}(\mathbf{r}_1, \dots, \mathbf{r}_N) \sum_k d_k D_k^\uparrow(\mathbf{r}_1, \dots, \mathbf{r}_{N_\uparrow}) D_k^\downarrow(\mathbf{r}_{N_\uparrow+1}, \dots, \mathbf{r}_N), \quad (19)$$

where D_k^\uparrow and D_k^\downarrow are Slater determinants of single-particle orbitals for the up- and down-spin electrons, respectively. The orbitals are a linear combination of Slater functions centered on the atoms for all-electron calculations while they are expanded on a Gaussian basis when pseudopotentials are employed. The Jastrow correlation function is a positive function of the interparticle distances and explicitly depends on the electron-electron separations.

Two questions should immediately arise when inspecting the wave function of Eq. 19. Why is the wave function only dependent on the space variables \mathbf{r} and not on the spin variables σ ? Why is the wave function not antisymmetric with respect to the interchange of any two particles?

3.2 Spin-assigned wave functions

Given a system of N electrons with $N = N_\uparrow + N_\downarrow$ and $S_z = (N_\uparrow - N_\downarrow)/2$, we define a spin function ζ_1

$$\zeta_1(\sigma_1, \dots, \sigma_N) = \chi_\uparrow(\sigma_1) \dots \chi_\uparrow(\sigma_{N_\uparrow}) \chi_\downarrow(\sigma_{N_\uparrow+1}) \dots \chi_\downarrow(\sigma_N). \quad (20)$$

and construct a set of $K = N!/(N_\uparrow!N_\downarrow!)$ distinct spin functions ζ_i by permuting the indices in ζ_1 . Since the spin functions ζ_i form a complete orthonormal set in spin space,

$$\sum_{\sigma_1 \dots \sigma_N} \zeta_i(\sigma_1, \dots, \sigma_N) \zeta_j(\sigma_1, \dots, \sigma_N) = \delta_{ij}, \quad (21)$$

we can decompose the wave function Ψ in terms of its spin components as

$$\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sum_{i=1}^K F_i(\mathbf{r}_1, \dots, \mathbf{r}_N) \zeta_i(\sigma_1, \dots, \sigma_N). \quad (22)$$

It follows from the antisymmetry of Ψ under the interchange of particle indices that each function F_i is antisymmetric under the interchange of like-spin electrons and that the F_i are all the

same except for a relabeling of the particle indices and a change in sign for odd permutations. Therefore, we can rewrite the wave function as

$$\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N) = \mathcal{A} \{F_1(\mathbf{r}_1, \dots, \mathbf{r}_N) \zeta_1(\sigma_1, \dots, \sigma_N)\} \quad (23)$$

It is easy to show using orthonormality of the functions ζ_i that the expectation value of an operator \mathcal{O} which is spin-independent is the same if we use the fully antisymmetric wave function Ψ or just one spatial function, say F_1 :

$$\langle \Psi | \mathcal{O} | \Psi \rangle = \langle F_1 | \mathcal{O} | F_1 \rangle. \quad (24)$$

Since it is more convenient to use the function F_1 than the full wave function Ψ , in quantum Monte Carlo, we always work with spin-assigned wave functions. To obtain F_1 , we simply assign the spin-variables of the particles as

Particle	1	2	...	N_\uparrow	$N_{\uparrow+1}$...	N
σ	1	1	...	1	-1	...	-1

so that $F_1(\mathbf{r}_1, \dots, \mathbf{r}_N) = \Psi(\mathbf{r}_1, 1, \dots, \mathbf{r}_{N_\uparrow}, 1, \mathbf{r}_{N_\uparrow+1}, -1, \dots, \mathbf{r}_N, -1)$.

As a simple example, we consider the spin-assignment of a one-determinant wave function of the $1s^2 2s^2$ state of the Be atom. The system has $N_\uparrow = N_\downarrow = 2$ and $S_z = 0$, and we construct the determinant from the four spin-orbitals $\phi_{1s} \chi_\uparrow$, $\phi_{2s} \chi_\uparrow$, $\phi_{1s} \chi_\downarrow$, and $\phi_{2s} \chi_\downarrow$ as

$$D(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) = \frac{1}{\sqrt{4!}} \begin{vmatrix} \phi_{1s}(\mathbf{r}_1) \chi_\uparrow(\sigma_1) & \dots & \phi_{1s}(\mathbf{r}_4) \chi_\uparrow(\sigma_4) \\ \phi_{2s}(\mathbf{r}_1) \chi_\uparrow(\sigma_1) & \dots & \phi_{2s}(\mathbf{r}_4) \chi_\uparrow(\sigma_4) \\ \phi_{1s}(\mathbf{r}_1) \chi_\downarrow(\sigma_1) & \dots & \phi_{1s}(\mathbf{r}_4) \chi_\downarrow(\sigma_4) \\ \phi_{2s}(\mathbf{r}_1) \chi_\downarrow(\sigma_1) & \dots & \phi_{2s}(\mathbf{r}_4) \chi_\downarrow(\sigma_4) \end{vmatrix}. \quad (25)$$

The spin-assigned function F_1 is given by

$$F_1(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = D(\mathbf{r}_1, +1, \mathbf{r}_2, +1, \mathbf{r}_3, -1, \mathbf{r}_4, -1) \quad (26)$$

$$= \frac{1}{\sqrt{4!}} \begin{vmatrix} \phi_{1s}(\mathbf{r}_1) & \phi_{1s}(\mathbf{r}_2) & 0 & 0 \\ \phi_{2s}(\mathbf{r}_1) & \phi_{2s}(\mathbf{r}_2) & 0 & 0 \\ 0 & 0 & \phi_{1s}(\mathbf{r}_3) & \phi_{1s}(\mathbf{r}_4) \\ 0 & 0 & \phi_{2s}(\mathbf{r}_3) & \phi_{2s}(\mathbf{r}_4) \end{vmatrix}, \quad (27)$$

so that the original determinant factorizes in the determinants of up and down spin-orbitals:

$$D \rightarrow F_1 = \begin{vmatrix} \phi_{1s}(\mathbf{r}_1) & \phi_{1s}(\mathbf{r}_2) \\ \phi_{2s}(\mathbf{r}_1) & \phi_{2s}(\mathbf{r}_2) \end{vmatrix} \times \begin{vmatrix} \phi_{1s}(\mathbf{r}_3) & \phi_{1s}(\mathbf{r}_4) \\ \phi_{2s}(\mathbf{r}_3) & \phi_{2s}(\mathbf{r}_4) \end{vmatrix} = D^\uparrow(\mathbf{r}_1, \mathbf{r}_2) \times D^\downarrow(\mathbf{r}_3, \mathbf{r}_4). \quad (28)$$

When constructing a spin-assigned wave function from the sum of several determinants $\Psi = \sum_k d_k D_k$, we proceed for each determinant as above. Note that we must first order the spin orbitals in all the determinants in the same way, e.g. first all the up-spin orbitals, then all the down-spin orbitals.

Finally, the Jastrow-Slater spin-assigned wave function obtained by imposing $\sigma = +1$ for first N_\uparrow particles and $\sigma = -1$ for the others is given by

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = F_1(\mathbf{r}_1, \dots, \mathbf{r}_N) = \mathcal{J} \sum_k d_k D_k^\uparrow(\mathbf{r}_1, \dots, \mathbf{r}_{N_\uparrow}) D_k^\downarrow(\mathbf{r}_{N_\uparrow+1}, \dots, \mathbf{r}_N) \quad (29)$$

where $\mathcal{J} = \mathcal{J}(\mathbf{r}_1, \dots, \mathbf{r}_N)$ is the Jastrow factor.

How do we impose that the Jastrow-Slater wave function Ψ has a given space and spin symmetry? Usually, the determinantal component $\sum_k d_k D_k$ is constructed to have the proper spacial symmetry while the Jastrow factor is a function of only the interparticle distances, $\mathcal{J} = \mathcal{J}(\{r_{ij}\}, \{r_{i\alpha}\})$ where the indices i, j refer to the electrons and α to the nuclei. Therefore, \mathcal{J} is invariant under rotations and does not affect the spacial symmetry of Ψ . As for the spin symmetry, $\sum_k d_k D_k$ is constructed to be an eigenstate of S^2 and S_z . If \mathcal{J} is symmetric for interchange of like-spin particles, the full wave function Ψ is an eigenstate of S_z and, if it is symmetric under interchange of the spacial variables, Ψ is also an eigenstate of S^2 [19].

3.3 Wave function and divergences in the potential

When the inter-particle distance goes to zero, the electron-nucleus and electron-electron potentials diverge as $-Z/r$ and $1/r$, respectively. Since the local energy

$$E_L = \frac{\mathcal{H}\Psi}{\Psi} = -\frac{1}{2} \sum_{i=1}^N \frac{\nabla_i^2 \Psi}{\Psi} + \mathcal{V} \quad (30)$$

must remain finite, the kinetic energy must have an opposite divergence to the potential \mathcal{V} . It is possible to ensure this cancelation if the trial wave function satisfies a set of *cuspl conditions* and displays a proper discontinuity of the derivatives at the coalescence points. To derive these conditions, let us consider two particles of masses m_i and m_j and charges q_i and q_j approaching each other while all other particles remain well separated. If we only keep the relevant diverging terms in the local energy and rewrite them in relative coordinates, we obtain close to $\mathbf{r} = \mathbf{r}_{ij} = 0$

$$-\frac{1}{2\mu_{ij}} \frac{\nabla^2 \Psi}{\Psi} + \mathcal{V}(r) \sim -\frac{1}{2\mu_{ij}} \frac{\Psi''}{\Psi} - \frac{1}{\mu_{ij}} \frac{1}{r} \frac{\Psi'}{\Psi} + \mathcal{V}(r) \sim -\frac{1}{\mu_{ij}} \frac{1}{r} \frac{\Psi'}{\Psi} + \mathcal{V}(r), \quad (31)$$

where $\mu_{ij} = m_i m_j / (m_i + m_j)$ is the reduced mass of the two particles and the derivatives are with respect to the spherical coordinate r . Therefore, as originally derived by Kato [20], the local energy is finite if

$$\left. \frac{1}{\Psi} \frac{\partial \Psi}{\partial r} \right|_{r=0} = \mu_{ij} q_i q_j, \quad (32)$$

where we used that $\mathcal{V}(r) = q_i q_j / r$ and we wrote the condition more precisely in terms of the spherical average of the wave function over an infinitesimally small sphere centered at $\mathbf{r} = 0$. Note that this condition is obtained assuming that $\Psi(r = r_{ij} = 0) \neq 0$.

At the electron-nucleus and electron-electron coalescence points, we therefore have

$$\text{for the electron-nucleus cusp: } \mu = 1, q_i = 1, q_j = -Z \Rightarrow \left. \frac{\Psi'}{\Psi} \right|_{r=0} = -Z;$$

$$\text{for the electron-electron cusp: } \mu = \frac{1}{2}, q_i = 1, q_j = 1 \Rightarrow \left. \frac{\Psi'}{\Psi} \right|_{r=0} = \frac{1}{2}.$$

How do we proceed if $\Psi(r_{ij} = 0) = 0$? This will for instance occur for two electrons in a triplet state or more generally two like-spin electrons, since the determinantal part and therefore the

wave function goes to zero, or for a highly excited state such as the $2p^2$ state of the He atom. Pack and Byers-Brown generalized Kato's cusp conditions to the case $\Psi(r_{ij} = 0) = 0$ [21]. By rewriting the wave function near $\mathbf{r} = \mathbf{r}_{ij} = 0$ as

$$\Psi = \sum_{l=l_0}^{\infty} \sum_{m=-l}^l f_{lm}(r) r^l Y_{lm}(\theta, \phi), \quad (33)$$

and expanding $f_{lm}(r) = \sum_{k=0}^{\infty} f_{lm}^{(k)} r^k$, they obtained at the coalescence points that

$$f_{lm}(r) = f_{lm}^{(0)} \left[1 + \frac{\gamma}{(l+1)} r + O(r^2) \right], \quad (34)$$

where $\gamma = q_i q_j \mu_{ij}$. Therefore, while for two electrons in a singlet state, we recover the previous result, the behavior of the wave function is different for like-spin electrons:

for electrons in a singlet state: $l = 0 \Rightarrow \Psi \sim \left(1 + \frac{1}{2} r \right) \Rightarrow \frac{\Psi'}{\Psi} = \frac{1}{2}$;

for electrons in a triplet state: $l = 1 \Rightarrow \Psi \sim \left(1 + \frac{1}{4} r \right) r$.

In a spin-assigned QMC wave function, the electron-electron cusp conditions are imposed through the Jastrow factor. For anti-parallel spins ($i \leq N_{\uparrow}, j \geq N_{\uparrow} + 1$), usually the determinantal component is different from zero and the Jastrow factor is constructed so that

$$\mathcal{J}(r_{ij}) \sim \left(1 + \frac{1}{2} r_{ij} \right) \Leftrightarrow \left. \frac{\mathcal{J}'}{\mathcal{J}} \right|_{r_{ij}=0} = \frac{1}{2}. \quad (35)$$

For parallel-spin electrons ($i, j \leq N_{\uparrow}$ or $i, j \geq N_{\uparrow} + 1$), the determinantal part goes to zero and the Jastrow factor behaves as

$$\mathcal{J}(r_{ij}) \sim \left(1 + \frac{1}{4} r_{ij} \right) \Leftrightarrow \left. \frac{\mathcal{J}'}{\mathcal{J}} \right|_{r_{ij}=0} = \frac{1}{4}. \quad (36)$$

Note that, if the cusps conditions are imposed for both parallel- and anti-parallel electrons, the Jastrow factor is not symmetric with respect to the interchange of the spacial coordinates and the wave function is not an eigenstate of S^2 . However, for well optimized wave functions, the spin contamination is found to be very small [19].

The electron-nucleus cusp conditions are imposed through the determinantal part by requiring that each orbital separately satisfies the cusp conditions. Note that there are no electron-nucleus cusps when pseudopotentials are employed.

As an example, consider again the Be atom in its $1s^2 2s^2$ state and its spin-assigned wave function $\Psi(\mathbf{r}_1^+, \mathbf{r}_2^+, \mathbf{r}_3^-, \mathbf{r}_4^-) = \mathcal{J} D$. In section 3.2, we obtained the factorized determinantal component which we can now multiply by a simple Jastrow factor satisfying the electron-electron cusp conditions:

$$\begin{aligned} \Psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) &= \begin{vmatrix} \phi_{1s}(\mathbf{r}_1) & \phi_{1s}(\mathbf{r}_2) \\ \phi_{2s}(\mathbf{r}_1) & \phi_{2s}(\mathbf{r}_2) \end{vmatrix} \times \begin{vmatrix} \phi_{1s}(\mathbf{r}_3) & \phi_{1s}(\mathbf{r}_4) \\ \phi_{2s}(\mathbf{r}_3) & \phi_{2s}(\mathbf{r}_4) \end{vmatrix} \\ &\times \prod_{ij=13,14,23,24} \exp \left\{ \frac{1}{2} \frac{r_{ij}}{1 + b r_{ij}} \right\} \times \prod_{ij=12,34} \exp \left\{ \frac{1}{4} \frac{r_{ij}}{1 + b r_{ij}} \right\}. \quad (37) \end{aligned}$$

3.4 Jastrow factor for atoms and molecules

In the previous section, we already introduced one of the simplest forms of a Jastrow factor which only includes electron-electron correlations:

$$\mathcal{J}(\mathbf{r}_1, \dots, \mathbf{r}_N) = \prod_{i < j} \mathcal{J}(r_{ij}) = \prod_{i < j} \exp \left\{ b_0 \frac{r_{ij}}{1 + b r_{ij}} \right\}. \quad (38)$$

where the product is over all pairs of electrons and $b_0 = 1/2$ or $1/4$ for anti-parallel and parallel spin electrons, respectively. A more general form of Jastrow factor proposed by Boys and Handy [22] depends not only on electron-electron but also on electron-nucleus distances and describes electron-electron, electron-nucleus and electron-electron-nucleus correlations:

$$\mathcal{J}(\mathbf{r}_1, \dots, \mathbf{r}_N) = \prod_{i < j} \mathcal{J}(r_i, r_j, r_{ij}) = \prod_{\alpha, i < j} \exp \left\{ \sum_{mnk} c_{mnk}^{\alpha} (\bar{r}_{i\alpha}^m \bar{r}_{j\alpha}^n + \bar{r}_{i\alpha}^n \bar{r}_{j\alpha}^m) \bar{r}_{ij}^k \right\}, \quad (39)$$

where the index α runs over the nuclei and the scaled distances $\bar{r}_{i\alpha} = r_{i\alpha}/(1 + a r_{i\alpha})$ and $\bar{r}_{ij} = r_{ij}/(1 + d r_{ij})$ are introduced to ensure that the Jastrow factor is well behaved at large interparticle separations. The electron-electron cusp conditions are imposed by requiring that $m = n = 0$ if $k = 1$, while the electron-nucleus cusps can be satisfied by the determinant with no contribution from the Jastrow factor if $n = 1$ or $m = 1$. A more general form of Jastrow factor is obtained by lifting the constraints to allow all values of n , m , and k , and impose the cusp conditions via linear dependencies among the coefficients c_{mnk}^{α} .

In general, we note that the Jastrow factor is chosen to be a positive function of the interparticle distances and, therefore, does not affect the sign of the wave function, which is solely determined by the determinantal component. At large interparticle distances, it plays no role since it becomes constant, which is achieved by using either a ratio of polynomials or scaled variables. To discuss the role played by the various terms in the Jastrow factor (as well as for other practical purposes), it is preferable to separate the electron-nucleus, electron-electron and electron-electron-nucleus terms as

$$\mathcal{J}(\mathbf{r}_1, \dots, \mathbf{r}_N) = \prod_{\alpha, i} \exp \{A(r_{i\alpha})\} \prod_{i < j} \exp \{B(r_{ij})\} \prod_{\alpha, i < j} \exp \{C(r_{i\alpha}, r_{j\alpha}, r_{ij})\}. \quad (40)$$

The electron-electron terms B are introduced to impose the electron-electron cusp conditions and to keep the electrons apart since the electron-electron interaction is repulsive. As shown in Fig. 1, a simple Jastrow factor $\mathcal{J}(r_{ij})$ displays a cusp at the origin and decreases in magnitude as the electrons approach each other. Using electron-electron Jastrow factors more general than the simple form does not lead to significant improvements. To really see the difference, we need to include a dependence on the electron-nucleus distances.

The electron-nucleus terms A should be included if the determinantal part is obtained from a density functional theory or a Hartree-Fock calculation and was not reoptimized after the inclusion of the electron-electron Jastrow factor. Otherwise, the electron-electron terms alter the single-particle density by reducing/increasing it in high/low density regions and the resulting density will in general be worse than the original DFT or HF density which can be restored by the inclusion of the electron-nucleus terms.

Finally, we have the electron-electron-nucleus terms C . If the order of the polynomial in these terms is infinite, the wave function can exactly describe a two-electron atom or ion in an S state.

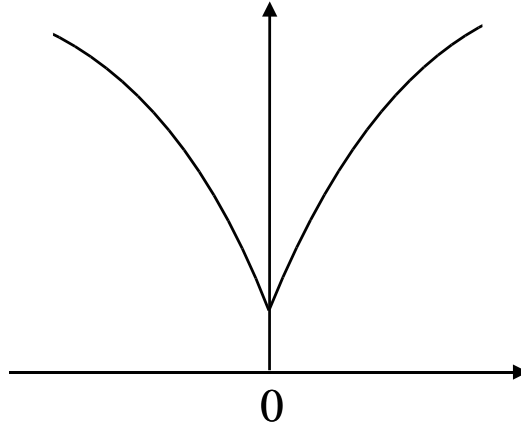


Fig. 1: Simple electron-electron Jastrow factor $\mathcal{J}(r_{ij})$ as a function of the inter-electron distance r_{ij} .

For these systems, a 5th-order polynomial recovers more than 99.99% of the correlation energy, which is defined as the difference between the exact and the HF energy, $E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}}$. A natural question is if this wave function is adequate for multi-electron systems. Clearly, the electron-electron-nucleus terms are the most important ones: due to the exclusion principle, it is rare for three or more electrons to be close since at least two electrons must necessarily have the same spin.

To demonstrate the effect of the different terms in the Jastrow factor, we report in Table 1 the results from Ref. [23] where the variational energy E_V and root mean square fluctuations of the local energy σ_V are computed for the Li, Be and Ne atoms with wave functions with Jastrow factors including up to electron-electron-electron-nucleus terms. While a large gain is obtained when adding the electron-electron-nucleus terms in both energy and σ_V , the improvement achieved with the inclusion of the electron-electron-electron-nucleus terms is significantly smaller. Note that as the number of electrons increases, one recovers a smaller percentage of correlation energy at the variational level.

3.5 Dynamic and static correlation

In the QMC wave functions, the Jastrow factor and the determinantal part account for two different types of correlations, so called *dynamic* and *static*, respectively. Dynamic correlation is described by the Jastrow factor and is due to the inter-electron repulsion and always present. Static correlation is due to the near-degeneracy of occupied and unoccupied orbitals, is not always present and is described by a linear combination of determinants.

As an example which ideally demonstrates the role of static correlation, let us consider the Be atom in its ground state. The HF ground state configuration is $1s^2 2s^2$ and, due to the $2s$ - $2p$ near-degeneracy, an additional important configuration is the $1s^2 2p^2$ obtained by promoting two electrons from the $2s$ to the $2p$ orbitals. Since the ground state has 1S symmetry, a good quality wave function of the proper symmetry can be obtained using a determinantal component of four determinants as

$$(1s^\uparrow, 2s^\uparrow, 1s^\downarrow, 2s^\downarrow) + c [(1s^\uparrow, 2p_x^\uparrow, 1s^\downarrow, 2p_x^\downarrow) + (1s^\uparrow, 2p_y^\uparrow, 1s^\downarrow, 2p_y^\downarrow) + (1s^\uparrow, 2p_z^\uparrow, 1s^\downarrow, 2p_z^\downarrow)] . \quad (41)$$

		\mathcal{J}	E_{VMC}	$E_{\text{VMC}}^{\text{corr}} (\%)$	σ_{VMC}
Li	E_{HF}		-7.43273	0	
	e-e		-7.47427(4)	91.6	0.24
	+ e-e-n		-7.47788(1)	99.6	0.037
	+ e-e-e-n		-7.47797(1)	99.8	0.028
	E_{exact}		-7.47806	100	
Be	E_{HF}		-14.57302	0	
	e-e		-14.66088(5)	93.1	0.35
	+ e-e-n		-14.66662(1)	99.2	0.089
	+ e-e-e-n		-14.66681(1)	99.4	0.078
	E_{exact}		-14.66736	100	
Ne	E_{HF}		-128.5471	0	
	e-e		-128.713(2)	42.5	1.9
	+ e-e-n		-128.9008(1)	90.6	0.90
	+ e-e-e-n		-128.9029(3)	91.1	0.88
	E_{exact}		-128.9376	100	

Table 1: Variational Monte Carlo energies (E_{VMC}) and root mean square fluctuations of the local energy (σ_{VMC}) in Hartree for the Li, Be and Ne atoms from Ref. [23]. $E_{\text{VMC}}^{\text{corr}}$ is the percentage of correlation energy gained, $E_{\text{VMC}}^{\text{corr}} = (E_{\text{VMC}} - E_{\text{HF}})/(E_{\text{exact}} - E_{\text{HF}})$. Different Jastrow factors are employed, including electron-electron (e-e), electron-electron-nucleus (e-e-n) and electron-electron-electron-nucleus (e-e-e-n) terms. All parameters are optimized in variance minimization (see section 4).

If we multiply the determinantal component by a simple Jastrow factor, we gain 61% of the correlation energy when we use only one determinant corresponding to the configuration $1s^2 2s^2$ while adding the $1s^2 2p^2$ configuration we obtain 93% of the correlation energy.

A natural question is whether we should bother with using a sophisticated Jastrow factor since the Jastrow factor is positive and the determinantal part yields the nodes of wave function and, therefore, determines the quality of the fixed-node diffusion Monte Carlo (DMC) solution (see section 5.3). There are however several reasons why we should use a good Jastrow factor also in a DMC calculation. The first consideration is efficiency: a better wave function with a smaller variance will result in a gain in CPU time and will also yield a smaller time-step error. Furthermore, when using the mixed estimator to compute expectation values other than the energy, we obtain a better estimate of the real expectation value if the trial wave function is accurate (see section 5.2). Finally, the Jastrow factor does affect the fixed-node energy when using pseudopotentials due to the localization error (see section 5.4).

To conclude, we briefly discuss why the Jastrow-Slater form of wave functions $\Psi = \mathcal{J}D$ should actually work. Let us assume that the solution Ψ of the eigenvalue problem $\mathcal{H}\Psi = E\Psi$ is known, and rewrite the exact wave function as $\Psi = \mathcal{J}\Phi$ where the Jastrow factor \mathcal{J} is given and Φ is defined as $\Phi = \Psi/\mathcal{J}$. Substituting the factorized form of Ψ in the eigenvalue equation, we obtain

$$\mathcal{H}\Psi = E\Psi \Leftrightarrow \mathcal{H}\mathcal{J}\Phi = E\mathcal{J}\Phi \Leftrightarrow \frac{\mathcal{H}\mathcal{J}}{\mathcal{J}}\Phi = E\Phi \Leftrightarrow \mathcal{H}_{\text{eff}}\Phi = E\Phi. \quad (42)$$

Therefore, the function Φ is the right eigenstate of the effective Hamiltonian $\mathcal{H}_{\text{eff}} = \mathcal{H}\mathcal{J}/\mathcal{J}$ with the same eigenvalue. Since the Jastrow factor cures the divergences of the potential in the Hamiltonian \mathcal{H} , we can view \mathcal{H}_{eff} as a weaker Hamiltonian than \mathcal{H} and therefore expect that approximating Φ with a non-interacting or weakly correlated wave function given by one or few determinants is a reasonable ansatz.

4 Wave function optimization

The ability of optimizing the parameters of the trial wave function is crucial for the success of quantum Monte Carlo methods. Good trial wave functions are needed to obtain accurate expectation values of the physical quantities of interest and reduced variance of the Monte Carlo estimators. If the trial wave function approaches an exact eigenstate of the Hamiltonian, the expectation value of the Hamiltonian and of any operator commuting with the Hamiltonian satisfies the zero variance principle: the expectation value approaches the exact eigenvalue and the Monte Carlo variance goes to zero.

Let us consider a wave function $\Psi(\mathbf{R}, \{\alpha\})$ depending on a set of parameters $\{\alpha\}$ and try to optimize the parameters using a fixed set of configurations N_{conf} sampled from the square of the wave function, $|\Psi(\mathbf{R}, \{\alpha_0\})|^2$, where $\{\alpha_0\}$ are the best values of the parameters available. The first thought one may have is to minimize the energy on the fixed set of configurations

$$E[\alpha] = \frac{1}{N_{\text{conf}}} \sum_{i=1}^{N_{\text{conf}}} \frac{\mathcal{H}\Psi(\mathbf{R}_i, \{\alpha\})}{\Psi(\mathbf{R}_i, \{\alpha\})} w_i, \quad (43)$$

where the weights w_i are given by

$$w_i = \left| \frac{\Psi(\mathbf{R}_i, \{\alpha\})}{\Psi(\mathbf{R}_i, \{\alpha_0\})} \right|^2 \bigg/ \sum_{i=1}^{N_{\text{conf}}} \left| \frac{\Psi(\mathbf{R}_i, \{\alpha\})}{\Psi(\mathbf{R}_i, \{\alpha_0\})} \right|^2, \quad (44)$$

and are introduced to allow the correct weighting as the wave function Ψ is changed since the configurations are sampled from the initial wave function with parameters $\{\alpha_0\}$. Unfortunately, the straightforward minimization of $E[\alpha]$ does not work: since N_{conf} is a relatively small number of configurations, the minimization often yields worse wave functions which give an arbitrary low value of $E[\alpha]$ on the given set of configurations but a higher energy than the one of the initial wave function in a long VMC run.

A second thought is to minimize the variance of the local energy [24]

$$\sigma^2[\alpha] = \sum_{i=1}^{N_{\text{conf}}} \left(\frac{\mathcal{H}\Psi(\mathbf{R}_i, \{\alpha\})}{\Psi(\mathbf{R}_i, \{\alpha\})} - \bar{E} \right)^2 w_i, \quad (45)$$

where \bar{E} is the average energy. In practice, \bar{E} is replaced by a guess energy E_{guess} which is chosen to be a bit less than the current estimate of the energy since this is equivalent to optimizing a combination of the variance and the energy.

Variance minimization is more stable than the minimization of $E[\alpha]$ since the variance has a known lower bound given by zero. The approach is found to be very robust even for a remarkably small number of configurations: 2000-3000 configurations are sufficient to optimize 50-100 parameters for spaces as large as 800 dimensions. Another advantage is that the variance

is zero for all eigenstates so that we can also use variance minimization to calculate true excited states, that is, excited states that are not the lowest lying state of that symmetry. Moreover, additional constraints on the wave functions can be easily included by minimizing the modified quantity $\chi^2 = \sigma^2 + \text{cost functions}$. Finally, very efficient algorithms such as the Levenberg-Marquard method are available to minimize the sum of squares.

Variance minimization is the most commonly used method in QMC to optimize the parameters in the wave function, in particular those in the Jastrow factor. An obvious drawback of the method is that we are minimizing the variance and not the energy while one is typically interested in obtaining the lowest energy in either a variational or a diffusion Monte Carlo calculation, rather than the lowest variance. Moreover, while variance minimization is very efficient in optimizing the parameters in the Jastrow factor, the optimization of the determinantal parameters is less stable and requires more iterations. So, most authors have used variance minimization for the Jastrow parameters only, where these problems are absent. Finally, for a given form of the trial wave function, energy-minimized wave functions on average yield more accurate values of other expectation values than variance minimized wave functions do.

If we abandon the idea of using a fixed set of configurations, a natural approach to the minimization of the energy with respect to the wave function parameters is to simply compute the gradient and the Hessian of the energy with respect to the parameters in a VMC run, and use them to improve the current wave function. However, a straightforward derivation of these quantities yields estimators with large statistical fluctuations as can be easily seen when deriving the gradient of the energy:

$$\partial_k \bar{E} = \left\langle \frac{\partial_k \Psi}{\Psi} E_L + \frac{\mathcal{H} \partial_k \Psi}{\Psi} - 2 \bar{E} \frac{\partial_k \Psi}{\Psi} \right\rangle_{\Psi^2} = 2 \left\langle \frac{\partial_k \Psi}{\Psi} (E_L - \bar{E}) \right\rangle_{\Psi^2}, \quad (46)$$

where, in the last step, we used the Hermiticity of the Hamiltonian. While the two estimators of the gradient yield the same expectation value, it is preferable to use the second expression not only for its greater simplicity but also because it has a significantly smaller Monte Carlo variance: as the wave function approaches the exact eigenstate, this expression has zero fluctuations since the local energy E_L becomes constant and equal to the energy. Similarly, one needs to rewrite the expression of the Hessian so that the estimator yields an unchanged expectation value with smaller fluctuations. The trick to doing this is to rewrite the expression of the Hessian in terms of covariances, $\langle ab \rangle - \langle a \rangle \langle b \rangle$, which have usually smaller fluctuations than $\langle ab \rangle$, by adding terms which are zero for infinite sampling but which cancel most of the fluctuations for finite sampling [25, 26].

The use of an estimate of the Hessian characterized by reduced statistical fluctuations yields a simple and robust optimization algorithm for the Jastrow parameters. The gain in efficiency is of at least three orders of magnitude with respect to the use of the original expression of the Hessian even for few-atom systems. However, the Hessian with respect to the orbital parameters in the determinant is affected by higher statistical noise so that devising a stable energy-minimization scheme is more difficult [26]. Other more complex approaches for the optimization of the wave function, in particular the determinantal component, have been recently proposed [27] but we will here omit their discussion due to their more advanced nature.

In summary, what is the customary practice for optimizing the wave function? Most authors optimize the parameters in the Jastrow factor using variance minimization on a fixed set of configurations as described above while they use methods such as Hartree-Fock, density functional theory or a small scale configuration interaction as a practical way of constructing the determinantal component, which is generally not reoptimized when the Jastrow factor is added. While

State	Wave function	E_{VMC}	E_{DMC}	ΔE (eV)
1^1A_g	HF	-38.684(1)	-38.7979(7)	–
	B3LYP	-38.691(1)	-38.7997(7)	–
	optimized	-38.691(1)	-38.7992(7)	–
1^1B_{1u}	CAS(2,2)	-38.472(1)	-38.5910(7)	5.63(3)
	B3LYP	-38.482(1)	-38.6030(7)	5.35(3)
	optimized	-38.493(1)	-38.6069(8)	5.23(3)
expt.				5.22

Table 2: Variational (E_{VMC}) and diffusion (E_{DMC}) Monte Carlo energies in Hartree for the 1^1A_g and the 1^1B_{1u} states of trans-hexatriene (C_6H_8) from Ref. [28]. The DMC excitation energies ΔE are with respect to the ground state obtained with the same wave function type. The errors on the last figure are given in parenthesis.

this procedure has proven to be very successful in many cases, it is not guaranteed to always be foolproof. In particular, the determinantal part of the wave function solely determines the fixed-node diffusion Monte Carlo energy (see section 5.3) and may need to be reoptimized to obtain accurate results. This is demonstrated in Table 2 where the use of HF or DFT orbitals in constructing the determinantal component of the wave function is not sufficient to give a reliable estimate of the excitation of trans-hexatriene [28]. Only using optimized orbitals are we able to obtain an excitation in very good agreement with the experimental results.

Whatever optimization technique is adopted to determine the wave function, it is fair to say that the construction of the trial wave function requires a good degree of knowledge and intuition about the physical problem being studied. As for quantum chemical methods, choices must be made about the functional form of the wave function to adopt. For example, for some systems, a single determinant wave function may be sufficient while there are situations where static correlation plays an important role and a multi-determinant wave function is necessary. In this last case, one needs to further decide how many and which determinants to include etc. Therefore, despite the significant advances in optimization techniques, quantum Monte Carlo retains a non-black box character which is instead so appealing for instance in DFT approaches.

5 Diffusion Monte Carlo

VMC is a very powerful method for a series of reasons. We can use any wave function as long as it is computable in a reasonable amount of computer time. We can learn a lot about the system while exploring which ingredients work in the wave function. Finally, there is no sign problem associated with Fermi statistics.

There are however several drawbacks to VMC. First, as already mentioned, there is no automatic method to construct the wave function and choices must be made about its functional form for each particular problem. Therefore, VMC tends to favor simple states over more complicated ones: for example, it is easier to construct a good wave function for a closed-shell than an open-shell system so that the energy of the former will be closer to the exact result than for the latter. Furthermore, properties other than the energy are significantly less accurate since they are first order in the error made on the wave function instead of second order as for the energy. In summary, one may say that whatever goes into the wave function is directly reflected in the

Method	Projector	Reference
Diffusion Monte Carlo	$\exp[-\tau(\mathcal{H} - E_T)]$	[29]
Green's Function Monte Carlo	$1/(\mathcal{H} - E_T)$	[30]
Power Monte Carlo	$E_T - \mathcal{H}$	[31]

Table 3: The projection operators for various projector Monte Carlo methods. τ is the time-step and E_T is the trial energy.

quality of the results.

Projector Monte Carlo is a more powerful method than VMC and removes (at least in part) the bias of the trial wave function from the results. It is a stochastic implementation of the power method for finding the dominant eigenstate of a matrix or integral kernel. In a projector Monte Carlo method, one uses an operator that inverts the spectrum of \mathcal{H} to project out the ground state of \mathcal{H} from a given trial state. Different operators have been used as projectors and are summarized in Table 3. Here, for simplicity, we only discuss diffusion Monte Carlo (DMC).

Let us consider an initial trial wave function $\Psi^{(0)}$ and repeatedly apply the projection operator to obtain the sequence of wave functions:

$$\Psi^{(n)} = e^{-\tau(\mathcal{H} - E_T)} \Psi^{(n-1)}. \quad (47)$$

If we expand the initial wave function $\Psi^{(0)}$ on the eigenstates Ψ_i with energies E_i of \mathcal{H} , we obtain for $\Psi^{(n)}$:

$$\Psi^{(n)} = \sum_i \Psi_i \langle \Psi^{(0)} | \Psi_i \rangle e^{-n\tau(E_i - E_T)}, \quad (48)$$

where $\langle \Psi^{(0)} | \Psi_i \rangle$ is the overlap between $\Psi^{(0)}$ and the eigenstate Ψ_i . Since the coefficients of the excited states die off exponentially fast relative to the coefficient of the ground state, we obtain

$$\lim_{n \rightarrow \infty} \Psi^{(n)} = \Psi_0 \langle \Psi^{(0)} | \Psi_0 \rangle e^{-n\tau(E_0 - E_T)}. \quad (49)$$

Therefore, if we choose the trial energy $E_T \approx E_0$ to keep the overall normalization of $\Psi^{(n)}$ fixed, the projection yields the ground state Ψ_0 of the Hamiltonian. Note that the starting wave function must have a non-zero overlap with the ground state.

How do we perform this projection? Let us first rewrite Eq. 47 in integral form and obtain

$$\Psi^{(n)}(\mathbf{R}', t + \tau) = \int d\mathbf{R} G(\mathbf{R}', \mathbf{R}, \tau) \Psi^{(n-1)}(\mathbf{R}, t), \quad (50)$$

where the coordinate Green's function is defined as

$$G(\mathbf{R}', \mathbf{R}, \tau) = \langle \mathbf{R}' | e^{-\tau(\mathcal{H} - E_T)} | \mathbf{R} \rangle. \quad (51)$$

If we can sample the trial wave function and the Green's function in Eq. 50, we can perform this high-dimensional integral by Monte Carlo integration. For fermions, since the wave function must be antisymmetric, it cannot be interpreted as a probability distribution. Therefore, for the moment, we will assume that we are dealing with bosons which are characterized by a positive ground state wave function.

Can we interpret the Green's function as a transition probability? Let us consider some properties of the Green's function. It is easy to show that $G(\mathbf{R}', \mathbf{R}, 0) = \delta(\mathbf{R}' - \mathbf{R})$ and that it satisfies the imaginary-time Schrödinger equation:

$$(\mathcal{H} - E_T)G(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial G(\mathbf{R}, \mathbf{R}_0, t)}{\partial t}, \quad (52)$$

If the Hamiltonian is only the kinetic energy operator, the Schrödinger equation in imaginary time is a diffusion equation in a $3N$ dimensional space:

$$-\frac{1}{2}\nabla^2 G(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial G(\mathbf{R}, \mathbf{R}_0, t)}{\partial t}, \quad (53)$$

and the Green's function is given by a Gaussian:

$$G(\mathbf{R}', \mathbf{R}, \tau) = (2\pi\tau)^{3N/2} \exp\left[-\frac{(\mathbf{R}' - \mathbf{R})^2}{2\tau}\right], \quad (54)$$

Therefore, it is positive, can be interpreted as a transition probability and easily sampled. On the other hand, if the Hamiltonian is only the potential operator, the Schrödinger equation is the rate equation:

$$(\mathcal{V}(\mathbf{R}) - E_T)G(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial G(\mathbf{R}, \mathbf{R}_0, t)}{\partial t}, \quad (55)$$

and the Green's function is given by

$$G(\mathbf{R}', \mathbf{R}, \tau) = \exp[-\tau(\mathcal{V}(\mathbf{R}) - E_T)] \delta(\mathbf{R} - \mathbf{R}'), \quad (56)$$

which is again positive but it does not preserve the normalization, and represents a factor by which, at each step, we multiply the current distribution function $\Psi(\mathbf{R}, t)$. We may therefore expect that the solution for a general Hamiltonian comprising both a kinetic and a potential term will describe a combination of a diffusion and a branching process. Formally, Trotter's theorem tells us that, in the limit of small time steps, we are allowed to consider the potential and kinetic energy contribution separately since, given two operators A and B , we have

$$e^{(A+B)\tau} = e^{A\tau}e^{B\tau} + \mathcal{O}(\tau^2). \quad (57)$$

Therefore, we can approximate the Green's function at short time steps as

$$\begin{aligned} \langle \mathbf{R}' | e^{-\mathcal{H}\tau} | \mathbf{R}_0 \rangle &\approx \langle \mathbf{R}' | e^{-T\tau} e^{-\mathcal{V}\tau} | \mathbf{R}_0 \rangle = \int d\mathbf{R}'' \langle \mathbf{R}' | e^{-T\tau} | \mathbf{R}'' \rangle \langle \mathbf{R}'' | e^{-\mathcal{V}\tau} | \mathbf{R}_0 \rangle \\ &= \langle \mathbf{R}' | e^{-T\tau} | \mathbf{R}_0 \rangle e^{-\mathcal{V}(\mathbf{R}_0)\tau}, \end{aligned} \quad (58)$$

where we used the result of Eq. 56, which combined with Eq. 54 gives the Green's function in the *short-time approximation*:

$$G(\mathbf{R}', \mathbf{R}, \tau) = (2\pi\tau)^{3N/2} \exp\left[-\frac{(\mathbf{R}' - \mathbf{R})^2}{2\tau}\right] \exp[-\tau(\mathcal{V}(\mathbf{R}) - E_T)] + \mathcal{O}(\tau^2). \quad (59)$$

Therefore, the iteration in Eq. 50 can be interpreted as a Markov process with the difference that the Green's function is not normalized and we obtain a branching random walk: the first factor in the short-time Green's function is the Green's function for diffusion while the second term multiplies the distribution by a positive scalar. Since the short-time expression of the Green's function is only valid in the limit of τ approaching zero, in practice, DMC calculations must be performed for different values of τ and the result extrapolated for τ which goes to zero.

The basic DMC algorithm is rather simple:

1. Generate M_0 walkers $\mathbf{R}_1, \dots, \mathbf{R}_{M_0}$ by sampling the trial wave function $\Psi^{(0)}(\mathbf{R})$ with the Metropolis algorithm. This is the zero-th *generation* and the number of configurations is the *population* of the zero-th generation.
2. Diffuse each walker as $\mathbf{R}' = \mathbf{R} + \xi$ where ξ is sampled from the Gaussian distribution $g(\xi) = (2\pi\tau)^{3N/2} \exp(-\xi^2/2\tau)$.
3. For each walker, compute the factor

$$p = \exp[-\tau((\mathcal{V}(\mathbf{R}) - E_T)] . \quad (60)$$

Branch the walker by treating p as the probability to survive at the next step: if $p < 1$, the walker survives with probability p while, if $p > 1$, the walker continues and new walkers with the same coordinates are created with probability $p - 1$. This is achieved by creating a number of copies of the current walker equal to the integer part of $p + \eta$ where η is a random number between (0,1).

4. Adjust E_T so that the overall population fluctuates around the target value M_0 .

As we can infer from the expression of p , in regions where the potential is less than the average energy, walkers will proliferate while, in regions of high potential energy, walkers will disappear. All walkers will propagate independently at future generations.

The role of the energy E_T is to keep the population stable and within computationally acceptable limits. If $M(t)$ is the current population and M_0 the desired population, we should adjust E_T by δE_T so that, at a later time T , we have

$$M(t+T) = M(t) \exp[-T(-\delta E_T)] = M_0 \Rightarrow \delta E_T = \frac{1}{T} \ln[M_0/M(t)] . \quad (61)$$

Consequently, the expression for E_T becomes

$$E_T(t+\tau) = E_{\text{est}}(t) + \frac{1}{g\tau} \ln[M_0/M(t)] , \quad (62)$$

where $E_{\text{est}}(t)$ is the current best estimate of the ground state energy and the second term attempts to reset the population to the target value M_0 some g generations later. Because of the use of a finite population, we must control the size of the population but this feedback of the number of walkers onto E_T introduces a systematic bias, the so-called *population control bias*. To understand the bias on the equilibrium distribution which will only approximately be equal to Ψ_0 , consider the branching with $E_T = E_0$. If fluctuations increase the population in regions where $\mathcal{V}(\mathbf{R}) < E_T$ the second term in Eq. 62 will have the opposite effect, decreasing the equilibrium distribution relative to Ψ_0 . A solution is to choose g large enough to stabilize the population around the target value without biasing too much the distribution. The drawback of working with a large g as well as a method to estimate the size of the population control error are discussed in Ref. [32].

Finally, note that a symmetric expression for the branching given by $\exp[-\tau(\mathcal{V}(\mathbf{R}) + \mathcal{V}(\mathbf{R}'))/2]$ can be obtained starting from the alternative expansion:

$$e^{(A+B)\tau} = e^{A\tau/2} e^{B\tau} e^{A\tau/2} + \mathcal{O}(\tau^3) . \quad (63)$$

This choice yields a better short-time approximation since it also preserves the symmetry property of the exact Green's function, $G(\mathbf{R}, \mathbf{R}', \tau) = G(\mathbf{R}', \mathbf{R}, \tau)$

5.1 Importance sampling

The simple algorithm described in the previous section is highly inefficient and unstable since the potential can vary significantly from configuration to configuration or also be unbounded like the Coulomb potential. For example, the electron-nucleus potential diverges to minus infinity as the two particles approach each other, and the branching factor will give rise to an unlimited number of walkers. Even if the potential is bounded, the approach becomes inefficient with increasing size of the system since the branching factor also grows with the number of particles.

These difficulties can be overcome by using *importance sampling* which was originally proposed by Kalos [30] for Green's function Monte Carlo and extended by Ceperley and Alder [1] to DMC. We start from Eq. 50, multiply each side by a trial wave function Ψ and define the probability distribution $f^{(n)}(\mathbf{R}) = \Psi(\mathbf{R})\Psi^{(n)}(\mathbf{R})$ which satisfies

$$f^{(n)}(\mathbf{R}', t + \tau) = \int d\mathbf{R} \tilde{G}(\mathbf{R}', \mathbf{R}, \tau) f^{(n-1)}(\mathbf{R}, t), \quad (64)$$

where the importance sampled Green's function is given by

$$\tilde{G}(\mathbf{R}', \mathbf{R}, \tau) = \Psi(\mathbf{R}') \langle \mathbf{R}' | e^{-\tau(\mathcal{H} - E_T)} | \mathbf{R} \rangle / \Psi(\mathbf{R}). \quad (65)$$

It is easy to show by differentiating $\tilde{G}(\mathbf{R}, \mathbf{R}_0, \tau)$ with respect to τ that it satisfies the equation:

$$-\frac{1}{2} \nabla^2 \tilde{G} + \nabla \cdot [\tilde{G} \mathbf{V}(\mathbf{R})] + [E_L(\mathbf{R}) - E_T] \tilde{G} = -\frac{\partial \tilde{G}}{\partial \tau} \quad (66)$$

where the local energy $E_L(\mathbf{R})$ was defined in Eq. 4 and the quantum velocity $\mathbf{V}(\mathbf{R})$ is given by

$$\mathbf{V}(\mathbf{R}) = \frac{\nabla \Psi(\mathbf{R})}{\Psi(\mathbf{R})}. \quad (67)$$

Now, we have three terms in the evolution equation corresponding to diffusion, drift and branching which can be considered separately for sufficiently short times as ensured by Trotter's theorem. We already discussed diffusion and branching, so we simply need to add the drift term to the Green's function in the short-time approximation. If we assume that $\mathbf{V}(\mathbf{R})$ remains essentially constant over the move (which is true as τ approaches zero), we can solve the evolution equation for the operator $\mathbf{V} \cdot \nabla + \nabla \cdot \mathbf{V} \approx \mathbf{V} \cdot \nabla$,

$$\mathbf{V} \cdot \nabla G(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial G(\mathbf{R}, \mathbf{R}_0, t)}{\partial t}, \quad (68)$$

which has as solution $\delta(\mathbf{R} - \mathbf{R}_0 - \mathbf{V}t)$. The resulting drift-diffusion-branching short-time Green's function is given by

$$\begin{aligned} \tilde{G}(\mathbf{R}', \mathbf{R}, \tau) &= (2\pi\tau)^{3N/2} \exp \left[-\frac{(\mathbf{R}' - \mathbf{R} - \tau\mathbf{V}(\mathbf{R}))^2}{2\tau} \right] \times \\ &\times \exp \{ -\tau [(E_L(\mathbf{R}) + E_L(\mathbf{R}'))/2 - E_T] \} + \mathcal{O}(\tau^2). \end{aligned} \quad (69)$$

There are two important new features of $\tilde{G}(\mathbf{R}', \mathbf{R}, \tau)$. First, the quantum velocity $\mathbf{V}(\mathbf{R})$ pushes the walkers to regions where $\Psi(\mathbf{R})$ is large. In addition, the local energy $E_L(\mathbf{R})$ instead of the potential $\mathcal{V}(\mathbf{R})$ appears in the branching factor. Since the local energy becomes constant and

equal to the eigenvalue as the trial wave function approaches the exact eigenstate, we expect that, for a good trial wave function, the fluctuations in the branching factor will be significantly smaller. In particular, imposing the cusp conditions on the wave function will remove the instabilities coming from the singular Coulomb potential.

The DMC algorithm will now be:

1. A set of configurations is sampled from $|\Psi(\mathbf{R})|^2$ using the Metropolis algorithm.
2. The walkers are advanced as $\mathbf{R}' = \mathbf{R} + \xi + \tau \mathbf{V}(\mathbf{R})$ where ξ is a normally distributed $3N$ dimensional random vector, and the last term is the drift.
3. The walkers are branched as in the simple algorithm but using the factor

$$p = \exp \left\{ -\tau [(E_L(\mathbf{R}) + E_L(\mathbf{R}'))/2 - E_T] \right\} . \quad (70)$$

4. The trial energy is adjusted to keep the population stable as in Eq. 62.

If the trial wave function equals the exact ground state wave function (perfect importance sampling), the local energy is constant and the branching step is no longer present. Then, the desired distribution to sample is $|\Psi|^2$ which is however not the equilibrium distribution of the drift-diffusion Green's function due to the short-time approximation: the repeated use of step 2 will produce $|\Psi|^2$ only in the limit of τ that goes to zero. As suggested by Reynolds *et al.* [29], it is possible to sample $|\Psi|^2$ with no time-step error by introducing an accept/reject step as in the generalized Metropolis method of section 2.1. The approximate drift-diffusion Green's function is used to propose a move which is then accepted with probability

$$p = \min \left\{ 1, \frac{|\Psi(\mathbf{R}')|^2 \tilde{G}(\mathbf{R}, \mathbf{R}', \tau)}{|\Psi(\mathbf{R})|^2 \tilde{G}(\mathbf{R}', \mathbf{R}, \tau)} \right\} . \quad (71)$$

This simple modification of the algorithm leads to a significant improvement on the basic algorithm. Further improvements are discussed in Ref. [32].

5.2 Expectation values and mixed estimators

The result of the DMC projection will now be the distribution given by $\rho(\mathbf{R}) = \Psi(\mathbf{R})\Psi_0(\mathbf{R})$. How do we compute the expectation values of the operators of interest? In a DMC run, we can easily collect the mixed estimator of an operator \mathcal{X} as

$$X_M = \frac{\langle \Psi_0 | \mathcal{X} | \Psi \rangle}{\langle \Psi_0 | \Psi \rangle} = \int d\mathbf{R} \frac{\mathcal{X} \Psi(\mathbf{R})}{\Psi(\mathbf{R})} \frac{\Psi(\mathbf{R}) \Psi_0(\mathbf{R})}{\int d\mathbf{R} \Psi(\mathbf{R}) \Psi_0(\mathbf{R})} = \langle X(\mathbf{R}) \rangle_\rho , \quad (72)$$

while we are interested in the expectation value of the operator on Ψ_0 :

$$X_0 = \frac{\langle \Psi_0 | \mathcal{X} | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle} . \quad (73)$$

For the energy or the expectation values of operators commuting with the Hamiltonian, we can use the mixed estimator since $X_M = X_0$. For instance, for the energy, we have

$$E_M = \frac{\langle \Psi_0 | \mathcal{H} | \Psi \rangle}{\langle \Psi_0 | \Psi \rangle} = E_0 \frac{\langle \Psi_0 | \Psi \rangle}{\langle \Psi_0 | \Psi \rangle} = E_0 , \quad (74)$$

since Ψ_0 is an eigenstate of \mathcal{H} with eigenvalue E_0 . However, the expectation values of operators that do not commute with the Hamiltonian will be biased. It is possible to partially correct this systematic error by combining the mixed estimator and the variational estimate on the wave function Ψ of the operator \mathcal{X} as

$$X_0 \approx 2X_M - X_V + \mathcal{O}[(\Psi - \Psi_0)^2]. \quad (75)$$

where the error is now quadratic in the trial wave function instead of linear as for the mixed estimator.

It is possible to get unbiased expectation values, at the cost of having a somewhat larger statistical error, using the *forward walking* technique [33]. Another possible solution is to use the reptation Monte Carlo approach to DMC [34].

5.3 Fixed-node approximation

In the discussion of DMC, we have not yet addressed the problem posed by the fact that electrons are fermions and that the trial wave function must be antisymmetric. First, let us consider how we may try to include Fermi statistics in the non-importance-sampled DMC method. Since we can represent the starting antisymmetric wave function $\Psi^{(0)}$ as the difference of two positive function $\Psi_+^{(0)}$ and $\Psi_-^{(0)}$, one could imagine assigning a sign to the walkers (positive walkers for $\Psi_+^{(0)}$ and negative walkers for $\Psi_-^{(0)}$) and separately evolving them using the simple DMC algorithm described at the beginning of section 5. Both distributions will eventually converge as

$$\Psi_{\pm}(\mathbf{R}, t) \rightarrow c_0^S \Psi_0^S \pm c_0 e^{-(E_0 - E_0^S)t} \Psi_0, \quad (76)$$

where the superscript S refers to the symmetric ground state. Since the bosonic ground state is lower in energy than the fermionic one, the difference $E_0 - E_0^S$ is positive. Therefore, the contribution along the fermionic ground state decays and both distributions converge to the bosonic state. It is still possible to obtain some information on the exact ground state by computing the mixed estimator with an antisymmetric trial wave function so that the bosonic component drops out of the averages. However, the signal to noise ratio decays exponentially.

Let us consider the DMC algorithm with importance sampling. The initial distribution is given by the square of the trial wave function $\Psi(\mathbf{R})^2$ and poses no problems since it can be interpreted as a probability distribution. However, the importance-sampled Green's function $\tilde{G}(\mathbf{R}', \mathbf{R}, \tau)$ (Eq. 65) is negative if a move changes the sign of Ψ so that $\Psi(\mathbf{R})/\Psi(\mathbf{R}') < 0$. Using the sign of the Green's function as a weight to assign to the walkers yields a growing statistical error on all expectation values.

To avoid this problem, we can simply forbid moves in which the sign of the trial wave function changes and the walker crosses the nodes which are defined as the set of points where the trial wave function is zero. This procedure is known as the *fixed-node approximation*. Forbidding node crossing is equivalent to finding the solution of the evolution equation with the boundary condition that it has the same nodes as the trial wave function. The Schrödinger equation is therefore solved exactly inside the nodal regions but not at the nodes where the solution will have a discontinuity of the derivatives. The fixed-node solution will be exact only if the nodes of the trial wave function are exact.

In general, the fixed-node energy will be an upper bound to the exact energy, in particular the best upper bound consistent with the boundary conditions given by the nodes of the trial wave

function. To prove this property, suppose that we have solved the Schrödinger equation in a subvolume Ω determined by the nodes of the trial wave function:

$$\mathcal{H}\Psi_{\text{FN}}(\mathbf{R}) = E_{\text{FN}}\Psi_{\text{FN}}(\mathbf{R}) \quad \mathbf{R} \in \Omega, \quad (77)$$

where $\Psi_{\text{FN}}(\mathbf{R}) = 0$ for $\mathbf{R} \notin \Omega$. We can extend the solution over all space by considering all permutations as

$$\tilde{\Psi}_{\text{FN}}(\mathbf{R}) = \frac{1}{N!} \sum_{\mathbf{P}} (-1)^{\mathbf{P}} \Psi_{\text{FN}}(\mathbf{P}\mathbf{R}), \quad (78)$$

which is not identically zero since permutations which maps a point in Ω back into Ω must be even, so that $\tilde{\Psi}_{\text{FN}}$ is at least non zero inside Ω . The variational energy is given by

$$\frac{\int d\mathbf{R} \tilde{\Psi}_{\text{FN}}^*(\mathbf{R}) \mathcal{H} \tilde{\Psi}_{\text{FN}}(\mathbf{R})}{\int d\mathbf{R} \tilde{\Psi}_{\text{FN}}^*(\mathbf{R}) \tilde{\Psi}_{\text{FN}}(\mathbf{R})} = E_{\text{FN}} \geq E_0. \quad (79)$$

We note that, if the nodes are not exact, the derivatives of $\tilde{\Psi}_{\text{FN}}$ are discontinuous and, consequently, the Laplacian in the Hamiltonian gives a delta function contribution. However, this delta function does not contribute to the integral in Eq. 79 since Ψ_{FN} is zero at the nodes.

In practice, how do we enforce the fixed-node condition that the walkers do not cross the nodes? We note that using a DMC algorithm with an accept/reject step (Eq. 71) ensures that the distribution of walkers $f(\mathbf{R})$ goes to zero quadratically at the nodes so that the asymptotic distribution corresponds to the fixed-node solution. Common practice is to reject the moves that cross the nodes. However, no walker will cross the nodes anyhow in the limit of τ going to zero since the drift term in the importance sampled Green's function in the short-time approximation (Eq. 69) diverges at the nodes, pushing the walkers away from the nodes.

5.4 Non-local pseudopotentials

The computational cost of a VMC or a DMC calculation scales in the number of electrons as N^4 since the computation of the determinant in the Jastrow-Slater wave function goes as N^3 while the root mean square fluctuations of the energy σ (and therefore the statistical error in Eq. 8) grows as \sqrt{N} as we homogeneously increase the size of the system. This must be compared to a scaling of N^3 for density functional theory and to a much less favourable N^7 behavior of a highly-correlated quantum chemical method such as coupled cluster theory.

While the QMC methods can be extended to large systems containing many electrons, the computational effort increases dramatically with the atomic number Z as the scaling is approximately Z^6 , rendering all-electron calculations quickly intractable. The problem is caused by the core electrons which yield large energies and large fluctuations of the energy. The most common way to overcome this difficulty is to replace the core electrons by pseudopotentials, an approximation which is usually rather good as the core is chemically inert. An electron-nucleus pseudopotential is usually non-local and the most commonly used form is a potential which is local in the radial coordinate and non-local in the angular part as

$$\langle \mathbf{r} | v^{\text{NL}} | \mathbf{r}' \rangle = \sum_{l=0}^{l_{\text{max}}} v^l(r) \delta(r - r') \sum_{m=-l}^l Y_{lm}(\Omega) Y_{lm}^*(\Omega'), \quad (80)$$

where l_{max} is the maximum angular momentum considered, and the function v^l is radial and vanishes outside a core radius r_c . The non-local potential acting on the trial wave functions gives

$$\langle \mathbf{R} | \mathcal{V}^{NL} | \Psi \rangle = \sum_{i=1}^N \sum_{l=0}^{l_{max}} v^l(r_i) \sum_{m=-l}^l Y_{lm}(\Omega_i) \int d\Omega'_i Y_{lm}^*(\Omega'_i) \Psi(\mathbf{r}_1, \dots, \mathbf{r}'_i, \dots, \mathbf{r}_N), \quad (81)$$

where the integral is over a sphere of radius $r'_i = r_i$ centered on the pseudoatom. This angular integration poses no particular problem in a VMC calculation and is done by a numerical quadrature on a regular polyhedron defined by a set of vertices whose number will depend on the value of l_{max} [35].

The use of non-local pseudopotentials is however a problem in DMC since the Green's function (Eq. 51) is no longer positive. This can be understood by analysing the behavior of the Green's function at short time steps:

$$\langle \mathbf{R}' | e^{-\tau \mathcal{H}} | \mathbf{R} \rangle \approx \langle \mathbf{R}' | \mathcal{I} - \tau \mathcal{H} | \mathbf{R} \rangle = \delta(\mathbf{R}' - \mathbf{R}) - \tau \langle \mathbf{R}' | \mathcal{H} | \mathbf{R} \rangle. \quad (82)$$

While the diagonal elements can always be made positive by choosing τ small enough, the off-diagonal elements are positive if and only if the off-diagonal elements of the Hamiltonian are non-positive. Therefore, if the potential has off-diagonal elements which are positive as for a non-local pseudopotential, the Green's function cannot have a probabilistic interpretation. This creates a new sign problem with difficulties similar to the fermionic sign problem discussed in the previous section.

A possible way to circumvent this problem is to introduce the so-called *locality approximation* and define a new effective core potential by localizing the non-local potential on the trial wave function [36] as

$$\mathcal{V}_{\text{eff}}(\mathbf{R}) = \frac{1}{\Psi(\mathbf{R})} \langle \mathbf{R} | \mathcal{V}^{NL} | \Psi \rangle. \quad (83)$$

This new effective potential is explicitly many-body but is local and can be easily incorporated in a DMC algorithm. However, the potential depends now on the trial wave function, and the DMC energy computed with the mixed estimator is no longer necessarily variational and depends on the quality of the trial wave function. As the trial wave function approaches the fixed-node solution obtained without the locality approximation, the DMC energy converges to the correct fixed-node energy quadratically fast in the error on the trial wave function. A different approach to handle non-local pseudopotential which is variational and improves the accuracy upon the DMC approach with the locality approximation was recently proposed in Ref. [37].

References

- [1] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
- [2] C. J. Umrigar and X. Gonze, Phys. Rev. A **50**, 3827 (1994); C. Filippi, X. Gonze, and C. J. Umrigar in *Theoretical and Computational Chemistry: Recent Developments in Density Functional Theory*, pgs. 295-326, edited by J. Seminario (Elsevier Publisher, 1996).
- [3] R. Q. Hood, M. Y. Chou, A. J. Williamson, G. Rajagopal, R. J. Needs, and W. M. C. Foulkes, Phys. Rev. Lett. **78**, 3350 (1997).
- [4] J. C. Grossman, L. Mitas, and K. Raghavachari, Phys. Rev. Lett. **75**, 3870 (1995).
- [5] S. B. Healy, C. Filippi, P. Kratzer, E. Penev and M. Scheffler, Phys. Rev. Lett. **87**, 016105 (2001); C. Filippi, S. B. Healy, P. Kratzer, E. Pehlke and M. Scheffler, Phys. Rev. Lett. **89**, 166102 (2002).
- [6] R. Q. Hood, P. R. C. Kent, R. J. Needs, and P. R. Briddon, Phys. Rev. Lett. **91**, 076403 (2004).
- [7] J. C. Grossman, M. Rohlfing, L. Mitas, S. G. Louie, and M. L. Cohen, Phys. Rev. Lett. **86**, 472 (2001).
- [8] A. Puzder, A. J. Williamson, J. C. Grossman, and G. Galli, Phys. Rev. Lett. **88**, 097401 (2002); A. J. Williamson, J. C. Grossman, R. Q. Hood, A. Puzder, and G. Galli, Phys. Rev. Lett. **89**, 196803 (2002).
- [9] M. D. Towler, R. Q. Hood and R. J. Needs, Phys. Rev. B **62**, 2330 (2000).
- [10] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
- [11] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods Volume 1: Basics* (Wiley, New York, 1986).
- [12] B.L. Hammond, W.A. Lester Jr. and P.J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry* (World Scientific, Singapore, 1994).
- [13] M. P. Nightingale and C. J. Umrigar Eds. *Quantum Monte Carlo Methods in Physics and Chemistry*, NATO ASI Series C. **525** (Kluwer Academic, Dordrecht, 1999).
- [14] D. M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995).
- [15] W. L. McMillan, Phys. Rev. **138**, A4422 (1965).
- [16] D. M. Ceperley, G. V. Chester, and M. H. Kalos, Phys. Rev. B **16**, 3081 (1977).
- [17] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
- [18] C. J. Umrigar, Phys. Rev. Lett. **71**, 408 (1993); C. J. Umrigar in Ref. [13].
- [19] C. J. Huang, C. Filippi and C. J. Umrigar, J. Chem. Phys. **108**, 8838 (1998).

- [20] T. Kato, *Comm. Pure Appl. Math.* **10**, 151 (1957).
- [21] R. T. Pack and W. Byers-Brown, *J. Chem. Phys.* **45**, 556 (1966).
- [22] S. F. Boys and N. C. Handy, *Proc. R. Soc. London Ser. A* **309**, 209 (1969).
- [23] C. Huang, C. J. Umrigar, and M. P. Nightingale, *J. Chem. Phys.* **107**, 3007 (1997)
- [24] C.J. Umrigar, K.G. Wilson and J.W. Wilkins, in *Computer Simulation Studies in Condensed Matter Physics: Recent Developments*, edited by D.P. Landau and H.B. Schüttler (Springer-Verlag, Berlin, 1988); *Phys. Rev. Lett.* **60**, 1719 (1988).
- [25] C. J. Umrigar and C. Filippi, *Phys. Rev. Lett.* **94**, 150201 (2005).
- [26] S. Sorella, *Phys. Rev. B* **71**, 241103 (2005).
- [27] F. Schautz and C. Filippi, *J. Chem. Phys.* **120**, 10931 (2004), and references therein.
- [28] A. Scemama and C. Filippi, *cond-mat/0511278* (2005).
- [29] J. B. Anderson, *J. Chem. Phys.* **63**, 1499 (1975); *ibid.* **65**, 4121 (1976); P. J. Reynolds, D. M. Ceperley, B. J. Alder and W. A. Lester, *ibid.* **77**, 5593 (1982); J. W. Moskowitz, K. E. Schmidt, M. A. Lee and M. H. Kalos, *ibid.* **77**, 349 (1982).
- [30] M. H. Kalos, *Phys. Rev.* **128**, 1791 (1962); *J. Comp. Phys.* **2**, 257 (1967); M. H. Kalos, D. Levesque, and L. Verlet, *Phys. Rev. A* **9**, 2178 (1974).
- [31] N. Trivedi and D. M. Ceperley, *Phys. Rev. B* **41**, 4552 (1990).
- [32] C. J. Umrigar, M. P. Nightingale and K. J. Runge, *J. Chem. Phys.* **99**, 2865 (1993).
- [33] M. H. Kalos, *Phys. Rev. A* **2**, 250 (1970).
- [34] S. Baroni and S. Moroni, *Phys. Rev. Lett.* **82**, 4745 (1999).
- [35] S. Fahy, X. W. Wang and S. G. Louie, *Phys. Rev. B* **42**, 3503 (1990).
- [36] B. L. Hammond, P. J. Reynolds, and W. A. Lester Jr., *J. Chem. Phys.* **87**, 1130 (1987); M. M. Hurley and P. A. Christiansen, *J. Chem. Phys.* **86**, 1069 (1986); P. A. Christiansen, *ibid.* **88**, 4867 (1988); L. Mitás, E. L. Shirley and D. M. Ceperley, *J. Chem. Phys.* **95**, 3467 (1991).
- [37] M. Casula, C. Filippi and S. Sorella, *Phys. Rev. Lett.* **95**, 100201 (2005).