

Atomic Scale Simulations

Variational Quantum Monte Carlo

Variational Quantum Monte Carlo

In this lecture, we consider the simplest generalization of classical simulation techniques to quantum systems namely the Variational Monte Carlo (VMC). The VMC method was first used by McMillan[6] to calculate the ground state properties of liquid ^4He and then generalized to fermion systems by Ceperley *et al.*[7]. It is a relatively simple generalization from a classical Monte Carlo simulation to VMC.

The variational theorem which gives its name to VMC states that for Ψ a proper trial function, the expectation of the Hamiltonian with respect to the trial function is an upper bound to the exact ground state energy:

$$E_V = \frac{\int \Psi^*(R) \mathcal{H} \Psi(R)}{\int \Psi^*(R) \Psi(R)} \geq E_0. \quad (1)$$

In VMC we use the Metropolis algorithm to sample the distribution:

$$\pi(R) = \frac{|\Psi(R)|^2}{\int |\Psi(R)|^2}. \quad (2)$$

One can easily see that the variational energy is simply the average value of the *local residual energy* over this distribution,

$$E_V = \int \pi(R) E_L(R) = \langle E_L(R) \rangle_\pi, \quad (3)$$

where the local residual energy of Ψ is defined as:

$$E_L(R) = \Psi^{-1} \mathcal{H} \Psi(R). \quad (4)$$

Note that as the trial function approaches an exact eigenfunction, $\Psi \rightarrow \phi_\alpha$, the local residual energy approaches a constant, the energy eigenvalue, E_i . This leads to a very important and general property of VMC, the *zero variance property*: as we improve the trial function, the Monte Carlo estimate of the variational energy converges more rapidly with the number of steps in the random walk. Of course, in this limit the upper bound is also becoming closer to the true energy. (In fact, the variance of the local energy is usually proportional to $E_V - E_0$.) It is because of the zero variance property that Quantum Monte Carlo calculations of energies can be much more precise than Monte Carlo calculations of classical systems. Fluctuations are only due to inaccuracies in the trial function.

Let me summarize the conditions that Ψ must satisfy to be useful for QMC. These conditions are for a continuum system, not a lattice model.

1. $\mathcal{H}\Psi$ must be well defined everywhere. Hence both Ψ and $\nabla\Psi$ must be continuous wherever the potential is finite otherwise differentiating will give singular terms. One must be particularly careful at the edges of the periodic box and when two particles approach each other. Otherwise E_V could lie above or below the true energy.

2. The integrals $\int |\Psi|^2$, $\int \Psi^* \mathcal{H} \Psi$, and $\int |\Psi \mathcal{H}|^2$ should exist. The existence of these integrals should be demonstrated analytically. If the last integral is infinite the central limit theorem may not hold and hence error bars will not be meaningful. Again, examine the limit as two particles approach each other and at infinity.
3. Ψ must have the proper symmetry: $\Psi(R) = (-1)^P \Psi(PR)$ for fermions and the right behavior at the periodic boundaries. Since the boson ground state is the lowest of all symmetries, the variational principle will be satisfied even for an unsymmetrical trial function. As an example, one can tie atoms to crystal lattice sites. Although this wavefunction does not have Bose symmetry, it still gives an upper bound since the boson energy is lowest of all.

For a lattice spin model, only item 3 is applicable.

The Pair Product Trial Function

What should we use for a trial wave function? First note that the ground state of a real Hamiltonian (*i.e.* no magnetic fields) can always be made real and non-negative. This implies that the ground state has Bose symmetry. Consider a system interacting with a one-body (*e. g.* an external potential) and two-body potentials and suppose the potential is composed of repulsive interactions at short range, like between helium atoms. Then the wavefunction should vanish when any pair of atoms approaches each other. This is the motivation which led Bijl[9] to propose using a pair product wavefunction. Each factor should be similar to the solution of the two-body wavefunction.

The pair product trial function is the simplest generalization of the Slater determinant and the ubiquitous form for the trial function in QMC:

$$\Psi(R, \sigma) = \exp\left[-\sum_{i < j} u(r_{ij})\right] \det[\theta_k(r_i, \sigma_i)], \quad (5)$$

where $\theta_k(r, \sigma)$ is the k th spin-orbital and $u(r)$ is the *pseudopotential* or pair-correlation factor. This function also goes by the name of a Jastrow[8] wave function. Closely related forms are the Gutzwiller function for a lattice, or the Laughlin function in the fractional quantum hall effect. Both $u(r)$ and $\theta_k(r, \sigma)$ are to be chosen by minimizing the variational energy or other quantity.

Computational Details of VMC

First, how do the particles move in VMC? On a lattice one can make a random hop of a particle or a spin flip. In the *classic* Metropolis procedure for a continuum system, one moves the particles one at a time by adding a random vector to a particle's coordinate, where the vector is either uniform inside of a cube or is a normally distributed random vector centered around the old position. The move for the i th particle is accepted with probability:

$$q(R \rightarrow R') = |\Psi(R')/\Psi(R)|^2 = \exp[-2 \sum_{j \neq i} (u(\mathbf{r}'_i - \mathbf{r}_j) - u(\mathbf{r}_i - \mathbf{r}_j))] \left| \sum_k \theta_k(\mathbf{r}'_i) C_{ki} \right|^2, \quad (6)$$

C is equal to the cofactor matrix divided by the determinant. Remembering our linear algebra, the matrix, C , is also the transposed inverse to the Slater matrix defined as:

$$\sum_k \theta_k(r_i) C_{kj} = \delta_{jk}. \quad (7)$$

Now the evaluation of a general determinant takes $O(N^3)$ operations. The evaluation of the fermion part of the acceptance ratio will take only $O(N)$ operations if C is already calculated. So it pays to keep C current as particles are being moved. If a move is accepted, C needs to be updated[7] using the formula:

$$C'_{jk} = C_{jk} + [\delta_{ji} - b_j]C_{ik}/b_i \quad (8)$$

where $b_j = \sum_k \theta_k(\mathbf{r}')C_{ki}$. (Remember it is particle i which is being moved.) This takes $O(N^2)$ operations. Hence to attempt moves for all N particles (a pass) takes $O(N^3)$ operations.

The local energy, needed to evaluate the variational energy is calculated by applying the Hamiltonian to the trial function. We get:

$$E_L(R) = V(R) + \lambda \sum_i [\nabla_i^2 U - \sum_k \nabla_i^2 \theta_k(\mathbf{r}_i) C_{ki} - G_i^2], \quad (9)$$

where $G_i = -\nabla_i U + \sum_k \nabla_i \theta_k(\mathbf{r}_i) C_{ki}$, and $U = \sum u(\mathbf{r}_{ij})$. Thus the inverse matrix is also needed to determine the local energy. Very often the orbitals are taken to be exact solutions to an external one-body potential:

$$-\lambda \nabla^2 \theta_k(\mathbf{r}) = (e_k - v(\mathbf{r}))\theta_k(\mathbf{r}) \quad (10)$$

Then the term $\sum_i \nabla_i^2 \theta_k(\mathbf{r}_i) C_{ki} = (e_k - v(\mathbf{r}_i))\theta_k(\mathbf{r}_i)$ simplifies. Finally note that using Green's identity allows several alternative ways[7] of calculating the variational energy. While some of them are simpler and do not involve so many terms, for a sufficiently good trial function, the local energy estimator of Eq. (9) has the lowest variance and is the one that should be used. The transformed energy expressions give useful tests of the computer program and the convergence of the random walk. It is very important to test that the first and second derivatives of the trial function are computed correctly. This is a very common source of error. It is best to do this test automatically whenever the trial function form is changed. Calculation of the derivatives by numerical differentiation is not recommended because of the loss of numerical precision and the slowness of this approach.

Optimization of Trial Functions

Optimization of the parameters in a trial function is crucial for the success of VMC. Bad upper bounds do not give much physical information. Good trial functions will be needed in the Projector Monte Carlo method. First we must decide on what to optimize and then how to perform the optimization. There are several possibilities of the quantity to optimize and depending on the physical system, one or other of the criteria may be best.

- The variational energy: E_V . If the object of the calculation is to find the least upper bound one should minimize E_V . There is a general argument suggesting that the trial function with the lowest variational energy will maximize the efficiency of Projector Monte Carlo[10].
- The variance of the local energy: $\sigma^2 = \int |\mathcal{H}\Psi|^2 - E_V^2$. If we assume that every step on a QMC calculation is statistically uncorrelated with the others, then the variance of the average energy will equal σ^2/p where p is the number of steps. The minimization of σ^2 is statistically more robust than the variational energy because it is a positive definite quantity with zero as a minimum value. One can also minimize a linear combination of the variance and the variational energy.
- The overlap with the exact wave function: $\int \Psi \phi$. If we maximize the overlap we find the trial function closest to the exact wave function in the least squares sense. This is the preferred quantity to optimize if you want to calculate correlation functions, not just ground state energies since then

the VMC correlation functions will be closest to the true correlation functions. Optimization of the overlap will involve a Projector Monte Carlo calculation to determine the change of the overlap with respect to the trial function so it is more complicated and rarely used.

Let us now consider what properties the optimal pseudopotential, u^* , has. If we suppose we assume that the spin-orbits come from an exact solution of a one-body potential, then the local energy expression simplifies. In particular, examine the dominant terms in the local energy Eq. (9) as 2 particles are brought together. We get

$$E_L(R) = v(r) + 2\lambda\nabla^2 u(r) - 2\lambda(\nabla u(r))^2 + \dots, \quad (11)$$

where r is the distance separating the particles. To keep the local residual energy finite the singularities in the kinetic energy must cancel the singularities of the potential energy. We see that $e^{-u(r)}$ will equal the solution to the 2-body Schroedinger equation. If two particles are sufficiently close together, other particles are irrelevant.

For He atoms interacting with a short range Lennard-Jones potential $4\epsilon(\sigma/r)^{12}$, the small distance behavior will be:

$$u(r) = ((2\epsilon\sigma^2)/(25\lambda))^{1/2}(\sigma/r)^5. \quad (12)$$

Charged particles have a different rule, they obey the *cusp condition*:

$$e_i e_j + 2(\lambda_i + \lambda_j) \left. \frac{u_{ij}}{dr_{ij}} \right|_0 = 0 \quad (13)$$

The effect of fluctuations caused by the kinetic energy is always to make the wavefunction smoother and less singular than the potential energy. Thus an r^{-12} potential becomes in the wave function r^{-5} . The r^{-1} Coulomb potential becomes a constant.

To describe the long wavelength behavior of the optimal $u(r)$ one uses a description in terms of collective coordinates such as phonons, or plasmons. We can write the variational energy in Fourier space as:

$$E_V = E_F + \sum_k (S_k - 1)(v_k - \lambda k^2 u_k) \quad (14)$$

where E_F is the fermion energy in the absence of correlation, v_k and u_k are the fourier transforms of $v(r)$ and $u(r)$, and S_k is the static structure factor for a given $u(r)$. Minimizing E_V with respect to u_k and making the RPA assumption of how S_k depends on u_k : $S_k^{-1} = S_{0k}^{-1} + 2\rho u_k$ where ρ is the particle density and S_{0k} is the structure factor for uncorrelated fermions, we obtain[11] the optimal wavefunction at long wavelengths:

$$2\rho u_k = -\frac{1}{S_{0k}} + \left[\frac{1}{S_{0k}} + \frac{2\rho v_k}{\lambda k^2} \right]^{1/2}. \quad (15)$$

For a short-ranged potential, (e.g. liquid helium), v_k can be replaced by a constant and we find the Reatto-Chester[12] form: $u(r) \propto r^{-2}$. But for a charged system, where $v_k \propto k^{-2}$, then $u(r) \propto r^{-1}$. Careful studies[11] have shown that this zero-parameter wavefunction is excellent for both the 2 and 3 dimensional electron gas where it satisfies both the optimal large r -behavior and the short-distance cusp condition.

The optimal wavefunction is long-ranged so that correlation extends beyond the edge of the simulation box. The ground state energy is little affected by this tail in the wave function because it is screened. On the other hand, response functions, such as the dielectric function or the static structure factor are crucially dependent on using the correct long-range properties. In order to maintain the upper bound property, the correlation function must be properly periodic in the simulation cell. For

high accuracy results and physically correct properties in the long wavelength limit, the Ewald image method[11] is needed to represent the correct long-range behavior of the optimal trial function.

For more complex systems, a purely Monte Carlo optimization method is needed to find a good trial function. The most direct method consists of running independent VMC runs using different variational parameters. One can fit the energies to a polynomial, performing more calculations near the predicted minimum and iterating until convergence in parameter space is attained. The difficulty with this direct approach is that close to the minimum the independent statistical errors will mask the variation with respect to the trial function parameters. This is because the derivative of the energy with respect to trial function parameters is very poorly calculated. Also, it is difficult to optimize in this way functions involving more than 3 variational parameters because so many independent runs are needed to cover the parameter space.

A correlated sampling method, known as reweighting[1, 7] is much more efficient. One samples a set of configurations $\{R_j\}$ (usually several thousand points at least) according to some distribution function, usually taken to be the square of the wavefunction for some initial trial function: $|\Psi_t(R; a_0)|^2$. Then the variational energy (or variance) for trial function nearby in parameter space can be calculated by using the same set of points:

$$E_v(a) = \frac{\sum_j w(R_j, a) E_L(R_j, a)}{\sum_j w(R_j, a)}, \quad (16)$$

where the weight factor $w(R) = |\Psi_T(R; a)/\Psi_T(R; a_0)|^2$ and the local energy is $E_L(R, a)$. The weight factors take into account that the distribution function changes as the variational parameters change. One then can use a minimizer to find the lowest variational energy or variance as a function of a keeping the points fixed. However there is an instability: if the parameters move too far away, the weights span too large of a range and the error bars of the energy become large. The number of effective points of a weighted sum is:

$$N_{eff} = (\sum w_j)^2 / \sum w_j^2. \quad (17)$$

If this becomes much smaller than the number of points, one must resample and generate some new points. When minimizing the variance, one can also simply neglect the weight factors. Using the reweighting method one can find the optimal value of wavefunction containing tens of parameters.

Problems with Variational Methods

The variational method is very powerful, and intuitively pleasing. One posits a form of the trial function and then obtains an upper bound. In contrast to other theoretical methods, no further approximations are made. The only restriction on the trial function is that the trial function be computable in a reasonable amount of time. There is no sign problem associated with fermi statistics in VMC. To be sure, the numerical work has to be done very carefully which means that convergence of the random walk has to be tested and dependence on system size needs to be understood.

One of the problems with VMC is that it favors simple states over more complicated states. As an example, consider the liquid-solid transition in helium at zero temperature. The solid wave function is simpler than the liquid wave function because in the solid the particles are localized so that the phase space that the atoms explore is much reduced. This biases the difference between the liquid and solid variational energies for the same type of trial function, (*e.g.* a pair product form) since the solid energy will be closer to the exact result than the liquid. Hence the transition density will be systematically lower than the experimental value. Another illustration is the calculation of the polarization energy of liquid ^3He . The wave function for fully polarized helium is simpler than for unpolarized helium because

antisymmetry requirements are higher in the polarized phase so that the spin susceptibility computed at the pair product level has the wrong sign!

The optimization of trial functions for many-body systems is time consuming, particularly for complex trial functions. In a one component system (say the electron gas) one only has to optimize a single $u(r)$ function since the orbitals are determined by symmetry. By contrast in the H_2O molecule, one has 5 different 3-dimensional orbitals (some related to each other by symmetry) and a 6-dimensional correlation function ($u(\mathbf{r}_i, \mathbf{r}_j)$). Clearly it is quite painful to fully optimize all these functions! Here I am not speaking of the computer time, but of the human time to decide which terms to add, to program them and their derivatives in the VMC code. This allows an element of human bias into VMC; the VMC optimization is more likely to be stopped when the expected result is obtained. The basis set problem is still plaguing quantum chemistry even at the SCF level where one only has 1-body orbitals. VMC shares this difficulty with basis sets as the problems get more complex.

Finally, the variational energy is insensitive to long range order. The energy is dominated by the local order (nearest neighbor correlation functions). If one is trying to compare the variational energy of a trial function with and without long range order, it is extremely important that both functions have the same short-range flexibility and both trial functions are equally optimized locally. Only if this is done, can one have any hope of saying anything about the long range order. The error in the variational energy is second order in the trial function, while any other property will be first order. Thus variational energies can be quite accurate while correlation functions are not very accurate.

As a consequence, the results typically reflect what was put into the trial function. Consider calculating the momentum distribution. Suppose the spin-orbitals have a Fermi surface. Then the momentum distribution of the pair product trial function will also have a Fermi surface although it will be renormalized. This does not imply that the true wave function has a sharp Fermi surface.

1 Beyond the pair-product trial function

Relatively little has been done to take the variational results beyond the two-body level. The possibilities for improving the pair-product trial function in a homogeneous one-component system are relatively limited. I will describe several of the recent directions.

The dominant term missing in the trial function for a bosonic system is a three-body (or polarization) term with the functional form of a squared force:

$$U_3(R) = - \sum_i \left[\sum_j \xi(r_{ij}) \vec{r}_{ij} \right]^2. \quad (18)$$

The new function $\xi(r)$ can be shown to be roughly given by $\xi(r) \approx du(r)/dr$. Because the polarization has the form of a squared force it is rapid to compute. The computational time is roughly the same as the product trial wave function.

For a fermion system, the interaction can shift the antisymmetric part away from the non-interacting Slater determinant. The simplest correction in a homogeneous system is known as “backflow”. The particle coordinates in the Slater determinants become “quasi-particle” coordinates:

$$\det[\theta_k(\vec{s}_i, \sigma_i)], \quad (19)$$

where the ‘quasi-particle’ coordinates are defined by:

$$\vec{s}_i = \vec{r}_i + \sum_j \eta(r_{ij}) \vec{r}_{ij}. \quad (20)$$

<i>terms</i>	E_V	$[E_V - E_0]/[2T]$
$u(McMillan)$	-5.702(5)	5.1%
$u(optimized)$	-6.001(16)	4.1%
u, ξ	-6.901(4)	0.86%
<i>DMC</i>	-7.143(4)	0.0%

Table 1: The energies of liquid ^4He in Kelvin/atom at zero pressure and zero temperature with various forms of trial functions[15]. In the first column u refers to pair correlations, ξ implies that three body terms were included. The second column shows the variational energies and the third column the percentage of the energy missed by the trial function. The numbers in parenthesis are the statistical error in units of 0.001K.

<i>terms</i>	E_V	$[E_V - E_0]/[2T]$	E_{FN}	<i>ref.</i>
u	-1.15(4)	5.7%	-2.20(3)	[13]
$u(optimized)$	-1.30(3)	5.7%	-2.20(3)	[15]
u, ξ	-1.780(17)	4.6%	-2.20(3)	[13]
u, η	-1.730(4)	3.7%	-2.37(1)	[13]
u, ξ, η	-2.163(6)	1.3%	-2.37(5)	[14]
<i>exp.</i>	-2.47	0.0%	-2.47	

Table 2: The energies of liquid ^3He in Kelvin/atom at zero pressure and zero temperature with various forms of trial functions. In the first column u refers to pair correlations, ξ implies that three body terms were included and η means backflow terms were included. The second column shows the variational energies and the third column the percentage of the energy missed by the trial function. The fourth column shows the results obtained with FN-DMC. The numbers in parenthesis are the statistical error in units of 0.01K.

Backflow is needed to satisfy local current conservation. However the computation of the determinant and energy become much more complex, because each element of the Slater matrix now depends on all the electron coordinates.

Table 1 gives VMC energies for ^4He and Table 2 for ^3He , for a variety of trial functions. It is important to realize that the kinetic and potential energies are almost completely cancelling out, liquid helium is very weakly bound. The third column $(E_V - E_0)/(2T)$ is a measure of the accuracy of the trial function, where $T = 12.3K$ is the kinetic energy and $E_0 = -2.47K$ is the ground state energy. This ratio is independent of how the zero of potential energy is defined and is equal to the percentage error in the upper bound for a harmonic potential. The chief motivation for the simulation of ^3He is that the results can rather directly be compared with experiment, assuming of course that the assumed inter-atomic potential is known accurately enough. There is a gratifying convergence toward experiment as more terms are added to the trial function. The most important terms beyond the pair-product level are the backflow terms. Similar results have been obtained using backflow and polarization terms on the electron gas[16].

References

- [1] D. M. Ceperley and M. H. Kalos, *Monte Carlo Methods in Statistical Physics*, ed. K. Binder, Springer-Verlag (1979).
- [2] K. E. Schmidt and M. H. Kalos, *Monte Carlo Methods in Statistical Physics II, Topics in Current Physics*, ed. K. Binder, Springer-Verlag (1984).
- [3] K. E. Schmidt and D. M. Ceperley, *Monte Carlo Methods III*, ed. K. Binder, Springer-Verlag,(1991).
- [4] B.L. Hammond, W.A. Lester, Jr., and P.J. Reynolds, *Monte Carlo Methods in ab initio quantum chemistry*, World Scientific, Singapore, 1994
- [5] Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953, J. Chem. Phys. **21**, 1087.
- [6] W. L. McMillan, Phys. Rev. A **138**, 442 (1965).
- [7] D. M. Ceperley, G. V. Chester and M. H. Kalos Phys. Rev. B **16**, 3081 (1977).
- [8] R. Jastrow, Phys. Rev. **98**, 1479 (1955).
- [9] A. Bijl, Physica **7**, 869 (1940).
- [10] D. Ceperley, J. of Stat. Phys. **43**, 815 (1986).
- [11] D. M. Ceperley, Phys. Rev. **B18**, 3126 (1978).
- [12] L. Reatto , G. V. Chester, Phys, Lett. **22**,276 (1966).
- [13] K. E. Schmidt,M. A. Lee, M. H. Kalos, and G.V. Chester, Phys. Rev. Lett. **47** , 807 (1981).
- [14] R.M. Panoff and J. Carlson, Phys. Rev. Letts. **62** , 1130 (1989).
- [15] S. Moroni, S. Fantoni and G. Senatore, Europhys. Lett. **30**, 93 (1995); Phys. Rev. B **52**, 13547 (1995).
- [16] Kwon, Y., D. M. Ceperley, and R. M. Martin, Phys. Rev. **B48**, 12037 (1993); ibid. (1998).

©D. M. Ceperley, 2000.