A Bayesian analysis of Green's function Monte Carlo correlation functions

M. Caffarel[a] and D. M. Ceperley
Department of Physics and National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

(Received 20 April 1992; accepted 18 August 1992)

Imaginary-time correlation functions calculated by quantum Monte Carlo (QMC) are analyzed using the maximum entropy method (MaxEnt) to determine the ground-state energy and spectral overlap function. In contrast to earlier applications of MaxEnt, the data is obtained from importanced-sampled zero-temperature quantum Monte Carlo simulations. The analysis includes two steps. First, that spectral overlap function and ground state energy which maximizes the entropy and agrees with the QMC correlation functions is obtained. Then the errors in the energy are evaluated by averaging over all the possible images (average MaxEnt method), the multidimensional integrals being computed using the Metropolis algorithm. The central feature of this approach is that all the information present in the correlation functions is used in the only way consistent with fundamental probabilistic hypotheses. This allows us to fully exploit the information contained in the correlation functions at small imaginary times, thus avoiding large statistical fluctuations associated with large imaginary times. In addition, the computed errors include both the statistical errors and systematic extrapolation errors. The method is illustrated with a harmonic oscillator and the four-electron LiH molecule.

I. INTRODUCTION

It is known to be very difficult to solve the Schrödinger equation for fermion systems with a Monte Carlo procedure because of the "sign problem." This problem arises because Fermi averages decompose into a difference of contributions corresponding to even and odd permutations of the particle labels. At large imaginary times, these contributions nearly cancel and become exponentially smaller than the statistical error, but large times are needed to project out from the trial wave function excited state components. The increased variance coming from the cancellations at large times will be greatly reduced by exploiting more fully all the information present in data at small times. In any case, it is important to estimate not only the statistical errors, but the systematic errors resulting from stopping the calculation at a fixed projection time. These same issues are involved in computing the energy of any excited state, not just the Fermion ground state.

In a previous work,[1] we have shown that the usual methods of calculating the energy are not optimal. More specifically, using a variation of the Lanczós algorithm for QMC, we have shown that the ground-state energy could be recovered from fixed-node or released-node data computed at smaller imaginary times. However, there is no guarantee that the Lanczós method will use all the information present in the data and the stability of the algorithm with respect to statistical fluctuations is bad.

In this work, we present a general procedure for taking full advantage of the QMC data. In addition, this method allows one to evaluate error bars on the energies including systematic errors. The framework employed is Bayesian probability theory and more specifically the maximum entropy (MaxEnt) method. Maximum entropy has been used in a wide variety of image reconstruction problems encountered in such different fields as astronomy, magnetic resonance imaging, neutron scattering, etc.[2] MaxEnt is a general and powerful technique for reconstructing positive images from noisy and incomplete data. Recently, several groups have applied this technique to extract dynamical information from imaginary-time quantum Monte Carlo Green's functions of lattice models,[3–5] e.g., a single-impurity Anderson model.

Here, we propose to apply this method to data from electronic systems, calculated by zero-temperature quantum Monte Carlo. Although we shall also be concerned with extracting a spectral overlap, our main purpose is to estimate the ground-state energy. As explained before, our ultimate goal is to avoid the sign problem appearing at large times while still calculating converged results, both for the ground state of Fermion systems and for quantum excited states.

Let us briefly outline our approach. Quantum Monte Carlo can estimate imaginary-time correlation functions of the form

$$h(t)=\langle\Psi_T|e^{-tH}|\Psi_T\rangle, \tag{1}$$

where Ψ_T is a known antisymmetric trial function. This correlation function is related by a Laplace transform

$$h(t)=\int_{-\infty}^{+\infty} dE\, c(E)e^{-tE} \tag{2}$$

to the spectral overlap

$$c(E)=\sum_i \delta(E-E_i)\,|\langle\Psi_T|\Phi_i\rangle|^2, \tag{3}$$

[a]On leave from: Laboratoire Dynamique des Interactions Moléculaires, Université Paris VI, 75252 Paris Cedex 05, France.

Downloaded 28 Apr 2003 to 130.126.9.235. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/jcpo/jcpcr.jsp

where $\{E_i,\Phi_i\}$ are the ith exact eigenvalue and eigenfunction of H. In contrast to the usual spectral density, $c(E)$ includes the squared overlap between the trial function and the exact excited states. The use of a good trial function leads to very accurate estimates of the ground-state energy. Essentially, one is doing Monte Carlo only on the errors of the trial function.

The usual method of calculating the ground-state energy from $h(t)$ is to find its rate of decay at large time. Let the transient estimate energy be defined as

$$E_{\mathrm{TE}}(t)=-\frac{d\ln[h(t)]}{dt}. \tag{4}$$

Then

$$\lim_{t\to\infty} E_{\mathrm{TE}}(t)=E_0\,. \tag{5}$$

It is this transient estimate which has an exponentially increasing signal-to-noise ratio at large time for a Fermion ground state.

The approach considered here is to regard Eq. (2) as an ill-conditioned inversion problem and ask what spectral overlaps are consistent with the noisy estimates of $h(t)$. Using Bayes' theorem, we postulate that the probability distribution of a given spectral overlap is

$$P(c|h,c^*)\propto P(h|c)P(c|c^*), \tag{6}$$

where the likelihood function $P(h|c)$ is the distribution of Monte Carlo errors given a spectral overlap and $P(c|c^*)$ is the prior probability of a given spectral overlap and may depend on an assumed model spectral function c^*. We will use an entropic form for this prior probability.

The proposed approach includes two steps. First, we maximize $P(c|h,c^*)$ with respect to c and thus determine the most likely overlap. This we will call the MaxEnt step. Then, to obtain a reliable estimate of the statistical errors, we sample possible overlaps with probability $P(c|h,c^*)$. This we will call the AvEnt step.

The organization of this paper is as follows: In Sec. II, the quantum Monte Carlo methods used in this work are discussed briefly. Section III is concerned with the general presentation of the proposed approach. In Sec. IV, the main ideas of the method and details of implementation are presented for a harmonic oscillator and in Sec. V, for the LiH molecule. Finally, some concluding remarks are made in Sec. VI.

II. THE QUANTUM MONTE CARLO METHODS

In this paper, we are concerned with quantum systems described by a Hamiltonian of the form

$$H=-\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2+V(\mathbf{r}_1,\ldots,\mathbf{r}_N), \tag{7}$$

where $\mathbf{r}_i$ is the position of the ith particle and $V(R)$ is the potential energy. Our analysis is based on the following three imaginary-time correlation functions:

$$h^{(k)}(t)=\langle\Psi_T|H^kO(t)|\Psi_T\rangle \tag{8}$$

where $k=\{0,1,2\}$, Ψ_T is the trial function, generally the best available computable approximate wave function known for the system under consideration, and the operator $O(t)$ is an evolution operator. Two forms for $O(t)$ are used for continuous systems; they define the two following algorithms:

(i) Diffusion Monte Carlo (DMC)

$$O(n\tau)=\exp[-n\tau(H-E_T)]; \tag{9}$$

(ii) Green's function Monte Carlo (GFMC)

$$O(n\tau)=[1+\tau(H-E_T)]^{-n}. \tag{10}$$

Here E_T is the reference energy and τ is the time step.

In both approaches, time-correlation functions of Eq. (8) may be computed as averages over an ensemble of configurations (walkers) evolving in the $3N$-dimensional configuration space according to some appropriate probabilistic rules. Since both methods will be used to compute the time-correlation functions $h^{(k)}(t)$, we shall give a brief description of each of them. For the complete description of the basic aspects of QMC techniques, the reader is referred to the original works.

A. Diffusion Monte Carlo (DMC)

In DMC, configurations advance according to three elementary processes—diffusion, drift, and branching. In fact, branching is not a necessary step in DMC since it may be taken into account by introducing weights in the correlation functions defined along the stochastic trajectories generated by diffusion and drift. In this case, the number of walkers remains constant and the notion of trajectory is identical with that used in a classical molecular dynamics simulation. In this "pure" DMC, the configurations advance from t to $t+\tau$ according to

$$f(\mathbf{R}',t+\tau)=\int d\mathbf{R}\,p(\mathbf{R}\to\mathbf{R}',\tau)f(\mathbf{R},t), \tag{11}$$

where $f(\mathbf{R},t)$ represents the density of configurations (walkers) at time t and $p(\mathbf{R}\to\mathbf{R}',\tau)$ is the transition probability density describing the drifted diffusion. In the short-time approximation, it has the form

$$p(\mathbf{R}\to\mathbf{R}',\tau)=\left(\frac{1}{2\pi\tau}\right)^{3N/2}\exp\{-[\mathbf{R}'-\mathbf{R}-\mathbf{b}(\mathbf{R})\tau]^2/2\tau\}, \tag{12}$$

where the drift is

$$\mathbf{b}(\mathbf{R})=\nabla\ln(\Psi_G) \tag{13}$$

and Ψ_G a strictly positive function is a guiding (or importance) function used to increase the efficiency of the simulation by keeping the walk in important regions of phase space.

It is possible to show by iterating the relation (11), and by introducing at each iteration the weight factor $e^{-\tau E_L^G}$, where $E_L^G(R)\equiv H\Psi_G(R)/\Psi_G$ is the local energy associated with Ψ_G, $h^{(k)}(t)$ may be estimated as

$$h^{(0)}(t)=I\left\langle w(0)w(t)\exp\left[-\int_0^t E_L^G(s)ds\right]\right\rangle_{\mathrm{DRW}},$$

Downloaded 28 Apr 2003 to 130.126.9.235. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/jcpo/jcpcr.jsp

$$h^{(1)}(t)=I\left\langle \langle w(0)w(t)\tfrac{1}{2}[E_L^T(0)+E_L^T(t)] \times\exp\left[-\int_0^t E_L^G(s)ds\right]\right\rangle_{\mathrm{DRW}},$$

and

$$h^{(2)}(t)=I\langle w(0)w(t)E_L^T(0)E_L^T(t) \times\exp\left[-\int_0^t E_L^G(s)ds\right]\rangle_{\mathrm{DRW}}, \quad (14)$$

where $w(R)\equiv\Psi_T/\Psi_G$ is a weight, $E_L^T(R)\equiv H\Psi_T/\Psi_T$ is the local energy associated with the trial function, $\langle\cdots\rangle_{\mathrm{DRW}}$ refers to the average over drifting random walks generated by Eqs. (11)–(13), and $I\equiv\int\Psi_G^2$ is a normalization constant which will eventually drop out. The practical evaluation of time-correlation functions from Eq. (14) is rather simple; for more details, the interested reader is referred to Ref. 6.

Note that when Ψ_G is chosen to be $|\Psi_T|$, drifting random walks generated via Eqs. (11)–(13) are trapped in subdomains of the configuration space delimited by the $(3N-1)$-dimensional nodes of the trial wave function Ψ_T and no change of sign of the weight factors w occurs. This approach is called fixed-node approximation since the nodes are in general approximate. When Ψ_G is chosen to be strictly positive everywhere, no approximation is made, but the weights have no longer a definite sign for fermions. This exact, but unstable method will be referred to in the following as the transient method. More details about both approaches may be found elsewhere.[6–9]

B. Green's function Monte Carlo (GFMC)

The Green's function Monte Carlo method is similar to the DMC method, but it does not have the time-step error inherent to DMC approaches. We will see that it is important to remove all systematic errors before using the Bayesian statistical analysis since otherwise these errors could be amplified. There exist different variants of GFMC; we shall use the formalism developed by Ceperley[8] which is particularly convenient for particles interacting via a Coulombic force. Precise rules for diffusion, drift, and branching are described in the Ref. 8.

To compute the time correlation functions let us assume that $\mathbf{R}_0$ is a configuration distributed according to $\Psi_G^2(\mathbf{R}_0)/I$. Then assume that the set of configurations $\{\mathbf{R}_n^j\}$, with $1\leqslant j\leqslant M(\mathbf{R}_0,n)$, are produced after n generations (applications of the evolution operator). Then

$$h^{(0)}(n\tau)=I\left\langle w(\mathbf{R}_0)\sum_{j=1}^{M(\mathbf{R}_0,n)} w(\mathbf{R}_n^j)\right\rangle. \quad (15)$$

The other correlation functions $h^{(1)}(t)$ and $h^{(2)}(t)$ are computed in a similar way.

The situation with respect to Fermi statistics is identical in GFMC and DMC. The fixed-node approximation is obtained when Ψ_G is chosen to be $|\Psi_T|$, and a transient approach corresponds to the use of a positive guiding function.

Here, we shall consider an additional exact approach, the released-node method,[8] where the basic GFMC algorithm is used with an important difference. The initial configuration (denoted above as $\mathbf{R}_0$) comes from the output of a fixed-node calculation instead of a variational one. It should be emphasized that by so doing, slightly different matrix elements of the evolution operator are obtained, namely,

$$h^{(0)}(t)=\left\langle\Psi_T\left|\left[\frac{1}{1+\tau(H-E_T)}\right]^n\right|\Phi_{\mathrm{FN}}\right\rangle$$

and

$$h^{(1)}(t)=\left\langle\Psi_T\left|H\left[\frac{1}{1+\tau(H-E_T)}\right]^n\right|\Phi_{\mathrm{FN}}\right\rangle, \quad (16)$$

where Φ_{FN} stands for the "exact" fixed-node wave function. It is difficult to compute $h^{(2)}(t)$, since the analytical form for the result of the action of H on the fixed-node wave function is not known.

III. MAXIMUM ENTROPY ANALYSIS

In order not to repeat almost identical formulas for both DMC and GFMC data, we shall present the method only with DMC evolution [Eq. (9)], the extension to GFMC evolution [Eq. (10)] being straightforward.

The first step consists of realizing that the data $h^{(i)}(t)$ are related to the spectral overlap $c(E)$ via a linear transformation

$$h^{(k)}(t)=\int_{-\infty}^{+\infty} dE\,E^k c(E)e^{-t(E-E_T)}, \quad k=0,1,2, \quad (17)$$

where $c(E)$ has already been defined in Eq. (3). Now $c(E)$ exhibits a very sharp maximum at $E=E_0$ since Ψ_T is chosen as close as possible to the ground-state wave function Φ_0. The "zero-variance" principle of (zero-temperature) quantum Monte Carlo states that as the trial function approaches an exact eigenfunction, the statistical variance vanishes. It is important to preserve this property in the MaxEnt analysis. This we do by requiring that the reconstruction fit both $h^{(0)}(t)$ and $h^{(1)}(t)$.

Now, our problem is the following: Having computed with QMC a set of data $\{h^{(0)},h^{(1)},h^{(2)}\}$ at different times t_k: $1\leqslant k\leqslant M$ and estimated (via statistically independent calculations) the statistical errors of this data, we would like to find the "best" and the "average" spectral overlap $c(E)$ compatible—in a sense to be specified below—with our incomplete and noisy data. This problem is often encountered in image processing where noisy data are related to the quantity of interest—the image—by a linear transformation. A robust and coherent way of tackling this difficult problem is the maximum entropy method based on Bayesian logic. The essence of this approach is to look for the most probable function $c(E)$ compatible with the data and with any prior knowledge about $c(E)$.

For simplicity, we shall represent the data by using a single vector $h(t)$ as follows:

Downloaded 28 Apr 2003 to 130.126.9.235. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/jcpo/jcpcr.jsp

$$h(t_k)=h^{(0)}(t_k), \quad h(t_{k+M})=h^{(1)}(t_k),$$

$$h(t_{k+2M})=h^{(2)}(t_k) \tag{18}$$

and use a subscript C and F to distinguish between computed (QMC) data $h_C(t_k)$ and their corresponding fitted values $h_F(t_k)$ obtained via a representation of Eq. (17) in terms of a large (but finite) number of real exponentials

$$h_F^{(k)}(t)=\sum_{l=0}^{P} c_l E_l^k e^{-t(E_l-E_T)}, \tag{19}$$

where P is typically of the order of 200. The number of components and the spacing in energy are chosen to have a good representation of the integral in Eq. (17).

Now consider the first of the two probability functions in Bayes' theorem [Eq. (6)], the likelihood function. According to the central limit theorem [assuming that the variances of $w(R)$ and $w(R)E_L(R)$ exist] for sufficiently large simulation times, the probability of finding of a given QMC correlation function $h_C(t)$ will have a Gaussian distribution about its exact value $h_F(t)$;

$$P(h_C|c) \propto \exp(-\chi^2/2), \tag{20}$$

where

$$\chi^2=\sum_{i,j}^{3M} [h_F(t_i)-h_C(t_i)]C_{ij}^{-1}[h_F(t_j)-h_C(t_j)]. \tag{21}$$

Here C_{ij} is the covariance matrix defined by

$$C_{ij}=\langle h_C(t_i)h_C(t_j)\rangle-\langle h_C(t_i)\rangle\langle h_C(t_j)\rangle, \tag{22}$$

where the averages are over a set of statistically independent calculations. We have checked systematically that our QMC-calculated correlation functions obey Gaussian statistics.

Any image $c(E)$ [in practice, the finite number of coefficients c_l in Eq. (19)] with a χ^2 significantly greater than the number $3M$ of data points is improbable. The set of feasible images is defined as the set of images verifying

$$\chi^2 \sim \text{number of data points.} \tag{23}$$

Clearly, when the chi squared is significantly smaller than the number of data points, we are overfitting the data, and the resulting fit, which is essentially determined by the noise, should be excluded from the set of feasible images.

Now we need a criterion to pick up among all these feasible images. Many of the feasible images are physically impossible or improbable, e.g., those with a negative spectral overlap $C(E)<0$. The role of the prior probability $P(c|c^*)$ is to filter out from all feasible images those that are very different from a default model containing any prior knowledge we have about the exact solution. It can be argued by using very general probabilistic concepts that for positive and additive images, there is a natural measure for that probability—the entropic form[10]

$$P(c|c^*) \propto \exp[\alpha S(c|c^*)], \tag{24}$$

where the entropy is

$$S=\sum_{l=0}^{P} c(E_l)-c^*(E_l)-c(E_l)\ln\frac{c(E_l)}{c^*(E_l)} \tag{25}$$

and $c^*(E)$ is a default spectral overlap which encapsulates all exact knowledge about the spectral overlap before adding in the information from QMC. One may question whether this entropic function is appropriate for the spectral overlap function of a small molecule, where the discreteness of the energy levels may be important. Two of its features are significant—it only allows overlaps with $c(E)>0$ and it has a maximum at $c=c^*$.

The maximum entropy image is then defined as the image having a $\chi^2 \sim$ number of data and maximizing the entropy. Since both the χ^2 function and the entropy are convex functions, this image is defined uniquely. The parameter α controls the competition between S and χ^2.

Since our goal is to have an accurate evaluation of the ground state energy E_0, it is included in the set of fitting parameters. In practice, this means that we shall define our max entropy solution is the set $\{E_0,c_0,\ldots,c_P\}$ maximizing S with the constraint that $\chi^2 \sim 3M$. In the Bayesian framework, the parameter α becomes an additional variable with its own prior distribution. Following standard practice,[10] we use a probability distribution uniform in $\log(\alpha)$ and the MaxEnt optimization and AvEnt averages evaluated with $P(c|h,c^*)P(\alpha)$, where $P(\alpha) \propto 1/\alpha$ over some "sensible" range.

Once the MaxEnt spectral function has been determined, it is important to estimate the statistical and systematic error. For that, we look at a typical set of feasible images, sampled from $P(c|h,c^*)$ as defined from Eqs. (6), (21), and (24). Then the error on any component of the spectral density, say the ground-state energy, is

$$\delta E_0=\sqrt{\langle (E_0-\langle E_0\rangle)^2\rangle}, \tag{26}$$

where the averages are over $P(c|h,c^*)$. A similar formula may be written for the error of any coefficient c_l. This approach, which consists of integrating the fluctuations around the MaxEnt solution in the functional space of all possible images, is the average maximum entropy method.[11] The multidimensional integrals involved can be computed using the Metropolis algorithm.

The reader may be wondering why we have chosen to fit the three functions $h^{(k)}(t)$ rather than simply $h^{(0)}(t)$, since analytically $h^{(0)}(t)$ is equivalent. There are several reasons. In the pure DMC approach, if the trial function equals the guiding function, it is true that one could numerically differentiate $h^{(0)}(t)$ to obtain $h^{(1)}(t)$ and $h^{(2)}(t)$. In practice, this introduces systematic errors which cause the statistical analysis to become unreliable. In the more important case where the guiding function is not equal to the trial function, statistical fluctuations in the time derivative of $h^{(0)}(t)$ are different from those where $h^{(1)}(t)$ is computed directly. In fact, it is the very strong correlation between the fluctuations of $h^{(0)}(t)$ and $h^{(1)}(t)$ which lead to the zero variance property of QMC. It is clearly important to use both $h^{(0)}$ and $h^{(1)}$ in doing the statistical analysis, since it is in the correlation between these two functions that gives a low variance energy estimate.

Downloaded 28 Apr 2003 to 130.126.9.235. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/jcpo/jcpcr.jsp

Knowledge of all three correlation functions implies that the feasible images have values of the ground-state energy bounded from both above and below. The upper bound is given by the transient estimate energy and a lower bound by the Temple bound[12]

$$E_{\mathrm{TE}}(t)-\frac{[h^{(2)}(t)-E_{\mathrm{TE}}(t)^2]}{E_{\mathrm{gap}}}\leqslant E_0\leqslant E_{\mathrm{TE}}(t), \tag{27}$$

where E_{gap} is the gap to the next state of the same symmetry. Note that these upper and lower bounds squeeze in the exact energy exponentially fast in t because the energy E_{TE} of the projected trial function $e^{-tH/2}\Psi_T$ converges exponentially fast to the ground state. Now the spread in ground-state energies for feasible images will be smaller than the bounds in the above inequalities because we are asking that the spectral overlap and energy fit the Monte Carlo data at many times simultaneously. For example, the results of the Lanczós method, described below, give tighter upper bounds than the transient estimate energy using the same functions $h^{(0)}(t)$ and $h^{(1)}(t)$.

The fact that the fitting gives tighter energy bounds is only one of the reasons to use this approach. It is perhaps more important that the information in the correlation functions is combined in a statistically robust way. The error bars on the correlation functions are all exponentially increasing in time. Thus it is important to strongly weight the small time data. Bayesian statistics provides a systematic framework for balancing the statistically accurate, but biased data at short times with the noisy, but more converged data at large times.

IV. APPLICATION TO THE HARMONIC OSCILLATOR

In this section, we apply our approach to the Hamiltonian

$$H=-\frac{1}{2}\frac{d^2}{dx^2}+\frac{1}{2}x^2+\gamma x, \tag{28}$$

where γ is some constant defining the magnitude of the linear perturbation. By using the Green's function Monte Carlo method presented in Sec. II B, we calculate the time-correlation functions with a trial wave function Ψ_T chosen to be the ground-state wave function of the unperturbed ($\gamma=0$) harmonic oscillator

$$\Psi_T\sim e^{-x^2/2} \tag{29}$$

and $\Psi_G=\Psi_T$. Then one can show

$$h^{(i)}(n\tau)=e^{-\gamma^2/2}\sum_{k=0}^{\infty}\frac{1}{k}\left(\frac{\gamma^2}{2}\right)^k(k+E_0)^i\left(\frac{1}{1+k\tau}\right)^n \tag{30}$$

when the reference energy is equal to the ground-state energy $E_T=(1/2)-(\gamma^2/2)$. We have performed a GFMC calculation using the time-dependent Green's function of the unperturbed harmonic oscillator as trial Green's function.

The upper curve of Fig. 1 shows the transient energy $E_{\mathrm{TE}}(t)$. The open circles are the QMC results, to be compared with the solid line representing the analytical results

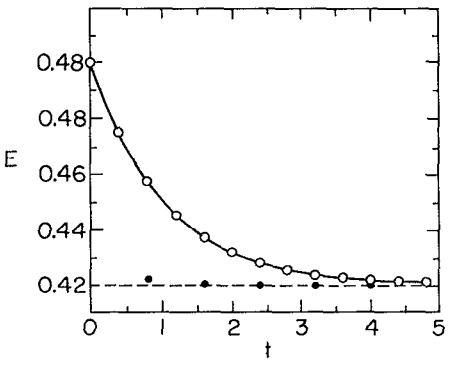

FIG. 1. Energy as a function of the projecting time for the harmonic oscillator. The solid line is the exact transient estimate energy from Eq. (30); the open circles are $E_{\mathrm{TE}}(t)$ from QMC. Results obtained from a Lanczós-type analysis are shown by the filled circles and t labels the amount of data used in the analysis. The dashed line shows the exact ground-state energy.

obtained from Eq. (30). The agreement between exact and computed values is excellent.

The lower curve gives the result of a Lanczós-type analysis using the *same* data. The Lanczós algorithm used here has been presented elsewhere[1] and may be summarized as follows: Consider the following projected trial wave function at some given time $t_p=p\tau$:

$$\tilde{\Psi}_T(t_p)=\left[\frac{1}{1+\tau(H-E_T)}\right]^p\Psi_T. \tag{31}$$

The overlap and Hamiltonian matrix elements between any two such states, say, at times t_p and t_q, may be expressed in terms of the time-correlation functions $h^{(0)}$ and $h^{(1)}$. We have

$$\begin{aligned}\langle\tilde{\Psi}_T(t_p)\,|\,\tilde{\Psi}_T(t_q)\rangle&=h^{(0)}(t_p+t_q),\\ \langle\tilde{\Psi}_T(t_p)\,|\,H\,|\,\tilde{\Psi}_T(t_q)\rangle&=h^{(1)}(t_p+t_q).\end{aligned} \tag{32}$$

Having computed all the time-correlation functions $h^{(i)}$ up to a given maximum time t, we consider the generalized eigenvalue problem defined in the basis set consisting of all of the projected trial wave functions defined at different times with $t_i\leqslant t/2$. By doing this, tighter upper bounds are obtained from the QMC data than from the transient estimate energy since more variational freedom is introduced. One is constructing the best linear combination of the projected trial wave function defined at different times $\Sigma_k c_k\tilde{\Psi}_T(t_k)$. One sees from Fig. 1 that the exact result is obtained almost immediately from the "short-time" information. However, this procedure, based on a nonlinear relation between QMC correlation functions and the energy, is unstable with respect to statistical errors.

Because of the discrete nature of the energy levels of the oscillator, we represent the spectral overlap by a sum of delta functions

$$c^*(E)=\sum_{i=0}^{P-1}\delta(E-E_i)c_i^*. \tag{33}$$

Downloaded 28 Apr 2003 to 130.126.9.235. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/jcpo/jcpcr.jsp

TABLE I. Maximum entropy analysis for the harmonic oscillator.

	$c^*(E)$	Most probable $c(E)$	Average $c(E)$	Exact $c(E)$
E_0	0.500	0.419 8	0.419 9(2)	0.420 00
c_0	0.900	0.923 21	0.923 2(2)	0.923 12
c_1	0.025	0.073 78	0.073 8(1)	0.073 85
c_2	0.025	0.003 13	0.003 10(5)	0.002 95
c_3	0.025	0	0	7.8×10^{-5}
c_4	0.025	0	0	1.5×10^{-6}

The coefficient representing the ground state is by far the largest since the overlap between the trial wave function and the ground state has been optimized. The energies E_i for $i>1$ could be incorporated into the fitting procedure, but for simplicity in this model, we only optimized the location of the main peak corresponding to the ground-state energy E_0. The other energies were fixed at their exact value. In applications to many-electron systems with a quasicontinuum of excited states, the assumption of a uniform grid in energy is more appropriate.

We assumed a "flat" default model, i.e., a model in which a uniform weight is used for the excited peaks. It is essential to introduce in the default model the fact that there is a very dominant peak close to the variational energy, but its precise magnitude does not matter. Once the magnitude of this peak has been chosen (we have taken 0.9), the common magnitude of the other peaks is chosen uniformly

$$\sum_{i=0}^{P-1} c_i^* = h^{(0)}(0) = 1. \tag{34}$$

In our numerical applications, we have chosen $P=5$ and therefore $c_i^* = 0.025$ (for $i>0$). We have then used the GFMC method to calculate the values of $h^{(0)}(t)$, $h^{(1)}(t)$, and $h^{(2)}(t)$ at the 13 times shown in Fig. 1.

A very important point to notice is that the transient estimate of the ground-state energy is a relatively smooth function of time despite the stochastic nature of its evaluation. This is due to the fact that we use a common set of random walks to compute the correlation functions at different times. In other words, all the QMC data on which we base our analysis are highly correlated. It is essential to include this correlation between the data in calculating the χ^2. In fact, when we tried to perform the MaxEnt analysis neglecting this correlation, the resulting ground-state energy was systematically biased because of a tendency to fit the noise. Therefore, our first step consists of performing a singular value decomposition of the covariance matrix C_{ij} to determine the degree of correlation between data and to discard those eigenvalues with singular values less than the computer's precision 10^{-16}. For the case shown in Fig. 1, 12 eigenvectors are kept out of 39 original data points.

The maximum entropy solution is given in Table I. It is remarkable that the analysis succeeds in reproducing accurately the magnitude of the first three peaks. The error bars have been obtained by averaging over the probability distribution of feasible images. This simple example illustrates the feasibility of this approach on an exact model.

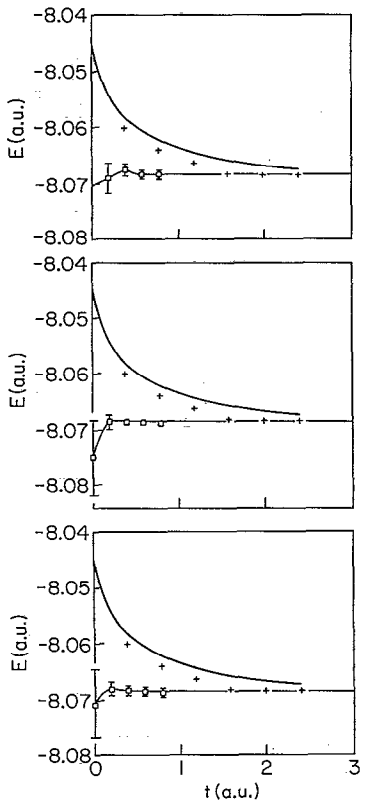

FIG. 2. Fixed node energy as a function of the projecting time t for LiH. The upper curve of the three pictures is $E_{TE}(t)$ and +'s are the Lanczós results. The lower curve of the first picture represents the average MaxEnt results obtained from the time-correlation function $h^{(0)}(t')$ only using the data $t'\leqslant t$. The second picture gives the result obtained with $h^{(0)}$ and $h^{(1)}$ and the third with $h^{(0)}$, $h^{(1)}$, and $h^{(2)}$. The solid horizontal line indicates the fixed-node energy.

V. A REALISTIC APPLICATION: THE LiH MOLECULE

Let us now apply the maximum entropy ideas to the determination of the ground-state energy of the LiH molecule. We begin with the fixed-node, pure DMC method of Sec. II A. In the calculations below, the Li and H ions were fixed with a bond length of 3.015 bohr, the trial wave function was Ψ_{II} of Ref. 13, and the guiding function from Ref. 8. The exact electronic energy, corrected from zero-point and relativistic effects, is estimated at $-8.070\,23$ hartrees. This number is computed by adding together the nonrelativistic energies of Li and H atoms,[14] subtracting the experimental binding energy of the molecule,[15] and the zero point energy of LiH. This experimental number is lower than a modern configuration interaction (CI) calculation[16] (which obtained an energy of -8.06904 hartrees) by 1.2 mhartrees.

A. Fixed node approach

The Monte Carlo data input to the Bayesian analysis consist of a set of 13 values for each correlation function $h^{(i)}(t)$ $(i=0,1,2)$ starting at $t=0$ and uniformly distributed with a spacing of $\delta t=0.2$ a.u. The upper curve of Fig. 2(a) [upper curves of Figs. 2(a)–2(c) are all identical] shows the transient energy $E_{TE}(t)$. Because the fixed-node

Downloaded 28 Apr 2003 to 130.126.9.235. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/jcpo/jcpcr.jsp

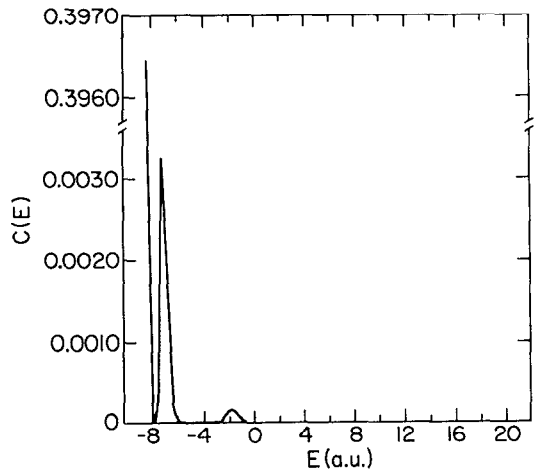

FIG. 3. Fixed node spectral overlap for LiH using all the Monte Carlo data.

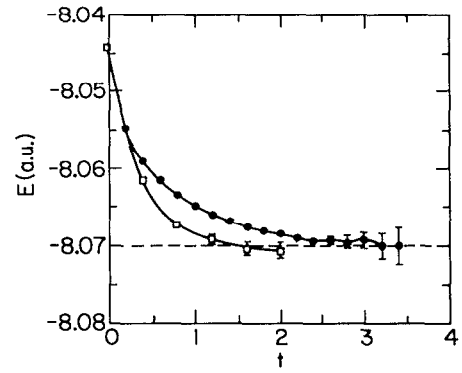

FIG. 4. Transient estimate energy (filled circles) and Lanczós energies (open squares) for LiH. The dashed line indicates the exact energy.

method is stable, one can extend this curve until the transient energy has converged. This occurs for times on the order of 3–4 a.u., where it reaches the value of $-8.0680(6)$ (represented by the horizontal solid line). This is higher than the exact energy because the assumed nodes of the trial function are not correct. We also show the Lanczós results [Figs. 2(a)–2(c), "+"s] obtained by using a two-dimensional basis set consisting of $\tilde{\Psi}_T(0)$ and $\tilde{\Psi}_T(n\tau)$ with $n=1,\ldots,6$. The convergence of the Lanczós energy is faster than that of the transient energy.

We have used a flat model for the spectral overlap, but in order to represent the continuum of states present, it was necessary to use a large number of fixed energies. P in Eq. (19) was on the order of a few hundreds. The spacing between these peaks needed to describe the details of the spectral overlap was found to be 0.1 a.u. In order to get a good fit to the data, energies up to 20 a.u. were included. Only the ground-state energy was varied in the analysis; the other energies (but not the spectral overlaps) were fixed. We note that one can prove by considering the existence of the integrals $\int \Psi_T H_m \Psi_T = \int dE\, E^m c(E)$ that the spectral overlap for a trial function of a Coulombic interaction will decay as $O(E^{-k})$ at large energies where the exponent k depends on whether the trial function has the correct two-particle cusp condition. We have chosen not to use this additional analytic prior information in our analysis. Indeed, our results are insensitive to the default model.

Figure 3 presents the maximum entropy spectral overlap obtained when all the data available [$h^{(0)}(t_i)$, $h^{(1)}(t_i)$, and $h^{(2)}(t_i)$ for $i=1$–13] are used. We see clearly besides the very dominant peak associated with the ground state, a structure for the excited states. More precisely, a second peak is seen ~ -7 a.u. as well as a smaller peak ~ -2 a.u. Note that the location of the first peak is in agreement with the estimate of the gap in energy which can be obtained from the exponential decay of the transient estimate curve of Fig. 2(a). This gap does not represent the first excited state, which is much smaller, but represents the lowest excited state having a large overlap with the trial wave function. The small peak at an energy of -2 a.u., which is always present in our analysis, represents a nontrivial feature of $c(E)$.

The lower curves of Figs. 2 (squares with error bars) show the dependence of the estimated ground-state energy on the quantity of information available. For example, at $t=0.2$ [Fig. 2(a)], we use only $h^{(0)}(0.2)$, perform the fit, and calculate errors. Obviously, the quantity of information is very limited and the error bar obtained on the energy is therefore large. Nonetheless, the error bars overlap with the exact result. Then, we incorporate the next value $h^{(0)}(2\tau)$, etc. The convergence is very rapid. Figures 2(b) and 2(c) show how the information contained in $h^{(1)}$ and $h^{(2)}$ affects the estimated energy and errors. The convergence is much enhanced, particularly at the very short times for which any additional information improves our knowledge of the ground-state energy substantially. However, we also see when convergence is reached, at approximately 1 a.u., introducing more data does not significantly reduce the error bars since little additional information is contained in the correlation functions. The only way of decreasing this error bar would be to decrease the statistical errors on the data by making a much longer Monte Carlo run.

In this example, the trial function was chosen equal to the guiding function which implies that the information in $h^{(1)}$ and $h^{(2)}$ is essentially present in $h^{(0)}$. Hence there is not a rapid decrease in errors bars as $h^{(1)}$ and $h^{(2)}$ are added to the analysis.

B. Transient method

We have computed the correlation functions using a strictly positive guiding function, thus removing the fixed-node restriction. The results in this section were done with the pure diffusion Monte Carlo scheme presented in Sec. II A. The upper curve of Fig. 4 reproduces the convergence of the transient estimate energy $E_{\mathrm{TE}}(t)$. The sign problem is evident for $t \approx 2.0$ a.u. The lower curve shows the results obtained by using the Lancós algorithm for these data. In contrast to the fixed-node case for which the MaxEnt analysis was successful, we encountered some se-

Downloaded 28 Apr 2003 to 130.126.9.235. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/jcpo/jcpcr.jsp

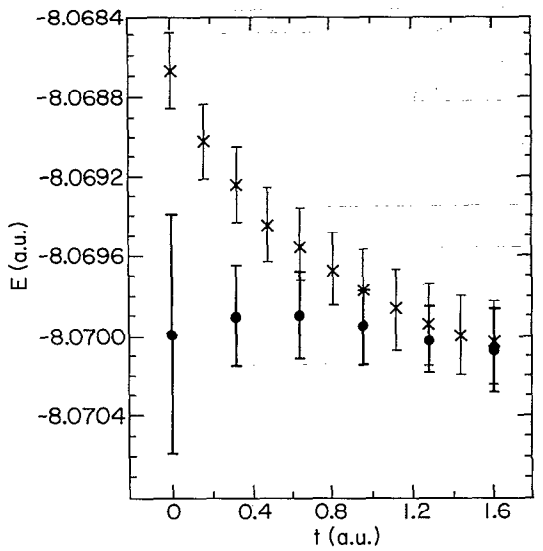

FIG. 5. Released-node energy as a function of the projecting time for LiH. The upper curve (X) is E_{TE}. The average MaxEnt results are given by filled circles. Note the expanded scale.

rious problems in doing the analysis for the transient data. Indeed, we found that the analysis is very sensitive to systematic time-step errors in the data. Very small time steps are required to obtain a time step-independent image. Unfortunately, in that case, the corresponding total simulation time (roughly, the number of MC steps multiplied by the time step) is too short to give reliable QMC data.

C. Nodal-release approach

To avoid the time-step errors and to speed up the convergence in imaginary time, we used the GFMC method with released node where one starts from the fixed-node output $\Psi_T\Psi_{\mathrm{FN}}$ as initial population instead of the transient method which starts from Ψ_T^2. This means that we compute slightly different matrix elements of the evolution operator and that

$$c(E)=\sum_i \delta(E-E_i)\langle\Psi_T|\Phi_i\rangle\langle\Phi_i|\Phi_{\mathrm{FN}}\rangle \qquad (35)$$

is not necessarily positive. However, since the trial wave function Ψ_T is chosen quite close to the fixed-node solution, it is likely that significant values of $c(E)$ are positive and we continued to use the entropic prior function. We are currently investigating a more rigorous prior function which will be needed to extend this analysis to excited state energies.

The transient energy is shown as the upper curve (X) of Fig. 5. The Bayesian results are given by filled circles. Since there is no simple way of computing the correlation function $h^{(2)}(t)$, the fit only includes $h^{(0)}$ and $h^{(1)}$. We see that the convergence of the MaxEnt solution is very rapid and leads to a stable solution without going to large projecting times. We get a very accurate value of -8.0700 ± 0.0002 for the ground-state energy. In contrast to previous calculations, our errors include both statistical errors and systematic errors. The only uncontrolled systematic error arises from the assumption of the prior probability. Our result is 0.23 ± 0.2 mhartree above the experimental result.

VI. CONCLUSIONS

In this paper, we have presented an application of Bayesian statistics to the determination of the ground-state energy of quantum systems. We analyze time-correlation functions obtained from zero-temperature quantum Monte Carlo calculations (projector methods) to obtain the spectral overlap function of a given trial function. This spectral overlap contains a dominant peak at the ground state and small components at higher energies. It has been found that the default model appearing in the prior probability does not play a role as important as in other applications, so a flat model for the density of excited states is sufficient. In order to calculate the ground-state energy, we considered the location of the main peak as a parameter to be optimized. By using a chi-squared likelihood function and an entropic prior function, we have computed the average and the dispersion of the estimator of the energy (average maximum entropy) with the Metropolis method. In that way, a reliable estimate of the errors is obtained. Our numerical applications have demonstrated the efficiency of this approach for simple problems, e.g., the LiH molecule treated with the fixed-node method. However, the situation is a little more difficult when real Fermion data (data exhibiting the sign problem) are analyzed. In particular, we found that the maximum entropy analysis for LiH was very sensitive to systematic errors such as the finite time-step error. By using a GFMC scheme free of systematic error and a released-node approach starting from an initial fixed-node population, we demonstrated feasibility of the method on the four-electron LiH molecule. It is not yet clear how this method will scale with the number of Fermions, but this way of analyzing the correlation functions is guaranteed to be better than the transient estimate method, simply because all information generated in the QMC is used.

The Bayesian approach also appears to provide a way of calculating excited state properties with quantum Monte Carlo. Our previous work[17] has used a generalization of the transient estimate method and was bedeviled with instabilities coming from statistical fluctuations. The approach considered here is the most general way of treating systematic and statistical errors and uses all information in the correlation functions. At present, we are testing various choices for the nonpositive prior functions needed for excited states. Applications to these more difficult problems will determine the usefulness and generality of this type of analysis.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation through grant number NSF DMR88-08126 by the Department of Physics at the University of Illinois, by the National Center for Supercomputing Applications, by

Downloaded 28 Apr 2003 to 130.126.9.235. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/jcpo/jcpcr.jsp

NATO, and by the CNRS. The computational aspects of this work used the CRAY-YMP at the National Center for Supercomputing Applications.

[1] M. Caffarel, F. X. Gadea, and D. M. Ceperley, Europhys. Lett. **16**, 249 (1991).
[2] S. F. Gull and J. Skilling, IEEE Proc. **131**, 646 (1984).
[3] R. N. Silver, D. S. Sivia, and J. E. Gubernatis, Phys. Rev. B **41**, 2380 (1990).
[4] R. N. Silver, J. E. Gubernatis, D. S. Sivia, and M. Jarrell, in *Condensed Matter Theories 6*, edited by S. Fantoni and A. Fabrocini (Plenum, New York, 1991), pp. 189–202.
[5] J. E. Gubernatis, M. Jarrell, R. N. Silver, and D. S. Sivia, Phys. Rev. B **44**, 6011 (1991).
[6] M. Caffarel and P. Claverie, J. Stat. Phys. **43**, 797 (1986); J. Chem. Phys. **88**, 1088 (1988); **88**, 1100 (1988).
[7] M. H. Kalos, in *Monte Carlo Methods in Quantum Problems, NATO ASI Series C* (Reidel, Dordrecht, 1982).
[8] D. M. Ceperley, J. Comp. Phys. **51**, 404 (1983), D. M. Ceperley and B. J. Alder, J. Chem. Phys. **64**, 5833 (1984).
[9] K. E. Schmidt and M. H. Kalos in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer, Berlin, 1984).
[10] J. Skilling and S. F. Gull, MaxEnt Conference SAMS-SIAM, 1988.
[11] S. R. White (unpublished).
[12] G. Temple, Proc. R. Soc. London, Ser. A **119**, 22 (1928).
[13] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, J. Chem. Phys. 77, 5593 (1982).
[14] C. F. Bunge, Phys. Rev. A **16**, 2496 (1977).
[15] K. Huber and G. Herzberg, *Constants of Diatomic Molecules* (Van Nostrand, New York, 1979).
[16] N. C. Handy, R. J. Harrison, P. J. Knowles, and H. F. Schaefer III, J. Phys. Chem. **88**, 4852 (1984). We have corrected the estimate of the experimental LiH energy by the exact energy of an isolated H atom, −0.499 73 a.u.
[17] D. M. Ceperley and B. Bernu, J. Chem. Phys. **89**, 6316 (1988); B. Bernu, D. M. Ceperley, and W. A. Lester, *ibid.* **93**, 552 (1990).

Downloaded 28 Apr 2003 to 130.126.9.235. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/jcpo/jcpcr.jsp