Finite population-size effects in projection Monte Carlo methods

Nicolas Cerf and Olivier C. Martin
Division de Physique Théorique, Institut de Physique Nucléaire, Orsay Cedex 91406, France
(Received 27 December 1993; revised manuscript received 28 September 1994)

Projection (Green's function and diffusion) Monte Carlo techniques sample a wave function by a stochastic iterative procedure. These methods converge to a stationary distribution which is biased, i.e., differs from the exact ground state wave function. This bias occurs because of the use of population control procedures. We demonstrate that these biased Monte Carlo algorithms lead to a modified effective mass which is equal to the desired mass only in the limit of an infinite population of walkers. In general, the bias for the energy scales as $1/N$ for a population of walkers of size N and is proportional to the expectation value of the kinetic energy. Finally, we consider various strategies to reduce this bias.

PACS number(s): 02.70.−c

I. INTRODUCTION

In the past two decades, Monte Carlo (MC) methods have been widely employed for studying quantum mechanical problems (see [1] for a recent review). In particular, they have proven to be valuable in determining ground state properties of particle [2–4], nuclear [5–7], atomic [8], and molecular systems [9], as well as of quantum fluids and solids [10–13].

Projection MC methods, which are our concern here, attempt to project out the ground state of a quantum system. The various implementations correspond to different choices of the projection operator. It is convenient to divide them into two main classes: Green's function Monte Carlo (GFMC) methods, which essentially use the Green's function H^{-1} as a projector, and diffusion Monte Carlo (DMC) methods, which use the imaginary time evolution operator $\exp(-Ht)$ as a projector. These methods are best understood as stochastic implementations of the power method. In that method, the dominant eigenvalue and eigenvector of a matrix or projection operator A are computed by iteratively applying A on an arbitrary initial vector $|\phi\rangle$ [14]. As the number n of iterations becomes large, one has, in Dirac's notation,

$$A^n|\phi\rangle = \lambda_0^n\langle\psi_0|\phi\rangle|\psi_0\rangle + O({\lambda_1}^n), \tag{1}$$

where λ_0 is the leading eigenvalue, $|\psi_0\rangle$ is the corresponding eigenvector, and λ_1 is the largest subleading eigenvalue. Then one has [14,15]

$$\lambda_0 = \lim_{n\to\infty}\left(\frac{\langle\psi_T|A^{n+k}|\phi\rangle}{\langle\psi_T|A^n|\phi\rangle}\right)^{1/k} \tag{2}$$

for any trial vector $|\psi_T\rangle$ and integer k.

In projection MC methods, this power method can be implemented stochastically as long as the operator A has non-negative matrix elements (a property that depends on the basis). For simplicity, we follow the notation of Ref. [16] and assume that A is a $d \times d$ matrix. One first defines a set of weights w_j, with $w_j = \sum_{i=1}^{d} A_{ij}$ and $j = 1, \ldots, d$. Then the matrix M, defined by

$$A_{ij} = M_{ij} w_j, \tag{3}$$

has each column summing to one ($\sum_{i=1}^{d} M_{ij} = 1$). This property ensures the conservation of probability and makes M a "stochastic" matrix. The matrix A is thus expressed as a product of the stochastic matrix M and the diagonal matrix w, i.e., $A = Mw$. The matrix M defines a Markov process that is used to generate the sequence of states (called configurations hereafter): given the configuration j, configuration i is chosen with probability M_{ij}. The matrix w makes it necessary to weight the configurations so that during evolution, the weight associated with configuration j is multiplied by w_j. It is easy to show that, on average, the evolution is identical to the standard power method (i.e., without stochasticity), so that Eq. (1) is reproduced for mean values. Thus the configurations after a large number n of steps will form a sample of the dominant eigenvector. If one considers a population of N random walkers ($\alpha = 1, \ldots, N$) in configurations i_α, carrying the weight w_α accumulated along the evolution, then

$$F(i) = \sum_{\alpha=1}^{N} \delta_{i,i_\alpha} w_\alpha \tag{4}$$

is an estimator of the ith component of the (unnormalized) dominant eigenvector of A, in the limit of many iterations. The number of iterations should be chosen large enough so that subdominant eigenvectors are projected out. However, in practice, this method of carrying weights does not work well because the variance of the above estimator grows exponentially with n, the number of iterations, while it decreases only with the usual $1/N$ law associated with random sampling. There are several possibilities to remedy this problem. One is to repeat the n iterations a large number of times using R independent runs (but still using the same initial vector

$|\phi\rangle$). The variance of any estimator will then diminish as $N^{-1}R^{-1}$ since the walkers are all independent. This slow decrease will have to compensate a variance that grows exponentially in n; thus, in practice, it is necessary to resort to variance reduction techniques. One such method is to do importance sampling, i.e., to guide the random walk using a trial wave function. The result is a smaller variance, though still exponentially increasing with n. A further reduction of the variance is achieved by "replicating" the population of walkers [17,18]. The idea here is to discard low weight configurations and duplicate large weight configurations. Then, the weights of individual configurations are kept near a target value, typically 1; as a consequence, the number $N^{(n)}$ of configurations at iteration n (or their total weight) either increases or decreases exponentially with n because of the λ_0^n factor in Eq. (1). To avoid this, it would be best to evolve using the projection operator A/λ_0, but of course λ_0 is not known. So, in practice, one uses a population control method whereby the random walkers are evolved using A, then the weights are divided by $\lambda(n)$, and finally the walkers are replicated. The factor $\lambda(n)$ is adjusted during the simulation to keep the population size or its total weight approximately constant. This population control is sometimes called renormalization because the weights are rescaled.

This evolution using renormalization forms a Markov chain in the space of populations of configurations and has a stationary probability distribution: configuration i appears with a relative frequency $\psi^*(i)$. It is common practice to consider that $|\psi^*\rangle$ is proportional to the ground state wave function. However, this is not justified; in fact, $|\psi^*\rangle$ is *not* a multiple of $|\psi_0\rangle$; rather $|\psi^*\rangle$ is a biased estimator of $|\psi_0\rangle$. The amplitude of the bias goes to zero as $1/N$, where N is the characteristic size of the population used in the evolution. The renormalization introduces a bias as soon as the w matrix is not a multiple of the identity matrix. A qualitative way of arguing for the existence of such a bias is obtained by considering the case of a single random walker ($N = 1$) in the absence of replication. In this case, the renormalization procedure amounts to forgetting the weight after each evolution (the weight of the walker is reset to one). The evolution thus proceeds in fact according to the transition probabilities M_{ij} only; the corresponding stationary distribution (the eigenvector of M_{ij} associated with the eigenvalue 1) clearly differs from the desired eigenvector $|\psi_0\rangle$ of A, so there is a large bias. It is reasonable to assume that this bias remains for other values of N and that it goes to 0 as $N \to \infty$, since the procedure in that limit is equivalent to the nonstochastic case. It is intuitively clear that renormalization introduces some *approximation* in the Monte Carlo algorithm since information is lost at each rescaling: the absolute scale of the weights is thrown away; only relative magnitudes are kept.

Soon after the first use of projection Monte Carlo, it was realized (Kalos [19] and Ceperley and Kalos [20]) that the standard eigenvalue estimator was biased. This bias was rediscovered by Nightingale and Blöte [21] for transfer matrix calculations and by Gelbard [22] in the context of neutronics computations. Methods to reduce the bias for the energy were proposed by Ceperley and Kalos [20], Reynolds *et al.* [9], Nightingale and Blöte [21], and Umrigar *et al.* [23]. However, it was Hetherington [16] who first explicitly stated that $|\psi^*\rangle$ was biased by order $1/N$, albeit for a particular choice of replication and renormalization procedure. In spite of these (few) papers, in many works using projection Monte Carlo, it is not realized that $|\psi^*\rangle$ is biased or at best the bias is believed to be negligible. The aim of the present paper is to show the origin of this population control bias, calculate its magnitude in a model case as well as its scaling with various simulational parameters, and consider ways to remove it.

Our discussion of the bias applies to the GFMC method, to the DMC method, and also to all stochastic implementations of the power method. In particular, it applies to MC calculations for transfer matrices in statistical mechanics and to neutronics. However, we will present most of the derivations for the DMC method; the advantage of this method is that it follows from the standard Feynman path integral formalism, rendering the discussion simpler for our purposes. After briefly reviewing the DMC method in Sec. II, we calculate in Sec. III the errors introduced by a particular choice of the replication and renormalization. Section IV illustrates the bias on a pairing Hamiltonian, showing that it can be rather significant. Finally, we review different strategies that can be used or have been proposed to reduce this bias in Sec. V.

II. DIFFUSION MONTE CARLO METHODS

A. Time evolution and estimators

The DMC method was first developed in [18], providing a technique for calculating the ground state energy of a quantum system using random walks. There are three phases in a DMC method: (imaginary) time evolution, replication, and renormalization. These last two phases will be discussed in Sec. II B.

Consider a point $\mathbf{r}$ of mass m in configuration space, placed in the potential $V(\mathbf{r})$. The Schrödinger equation in imaginary time (with $\hbar \equiv 1$) reads

$$\frac{\partial \psi}{\partial t} = -H\psi = \frac{1}{2m}\nabla^2\psi(\mathbf{r},t) - V(\mathbf{r})\psi(\mathbf{r},t) \ . \qquad (5)$$

This is identical to a reaction-diffusion equation: the right-hand side can be interpreted as a diffusion term (with a diffusion constant $D = 1/2m$) and a reaction term describing a growth or decay process (with the source-sink term $-V\psi$). As is well known, the diffusion process can be treated via a random walk, whereas the reaction term necessitates the introduction of a multiplicative process.

The imaginary time evolution given by Eq. (5) acts as a projector which, at large time, selects out the lowest energy state. It is convenient to shift the potential by a trial energy E_T, giving

$$\frac{\partial\psi}{\partial t} = (E_T - H)\psi = \frac{1}{2m}\nabla^2\psi(\mathbf{r},t) - [V(\mathbf{r}) - E_T]\psi(\mathbf{r},t) \ . \tag{6}$$

The projection operator in the DMC method is thus the (imaginary) time evolution operator $U_t = e^{-(H-E_T)t}$. If E_T is larger than E_0, the wave function will grow in normalization whereas the opposite occurs for E_T smaller than E_0. As will be explained in Sec. II B, E_T is adjusted during time evolution to avoid an exponential increase or decrease in the normalization. In order to evaluate U_t in practice, it is necessary to resort to the short time approximation. One first divides the time interval t into M infinitesimal intervals Δt and uses the relation

$$U_t = \prod_{n=1}^{M} e^{-(H-E_T)\Delta t}. \tag{7}$$

The infinitesimal propagators are then approximated, for instance, by use of the Trotter formula as

$$e^{-(H-E_T)\Delta t} = e^{-T\Delta t}e^{-(V-E_T)\Delta t} + O(\Delta t^2) \tag{8}$$

where T and V stand for the kinetic and the potential energy, respectively. (It is of course possible to use breakups, which are more accurate.) Applying the evolution operator on an initial wave function $|\phi\rangle$ yields the ground state $|\psi_0\rangle$ at large t:

$$\psi_0(\mathbf{r}) \simeq e^{(E_0-E_T)t}\langle\psi_0|\phi\rangle^{-1}\int d\mathbf{r}^{(M)}\cdots d\mathbf{r}^{(0)}\ \langle\mathbf{r}|\mathbf{r}^{(M)}\rangle\langle\mathbf{r}^{(M)}|e^{-\Delta tT}|\mathbf{r}^{(M-1)}\rangle e^{-\Delta t[V(\mathbf{r}^{(M-1)})-E_T]}$$
$$\times\cdots\langle\mathbf{r}^{(1)}|e^{-\Delta tT}|\mathbf{r}^{(0)}\rangle e^{-\Delta t[V(\mathbf{r}^{(0)})-E_T]}\phi(\mathbf{r}^{(0)}). \tag{9}$$

We have assumed that the potential $V(\mathbf{r})$ is local in configuration space, so that applying $e^{-\Delta t(V-E_T)}$ simply amounts to multiplying by a weight w equal to that factor. The operator $e^{-\Delta tT}$ corresponds to a diffusion operator whose matrix elements are

$$P(\mathbf{r}',\mathbf{r}) \equiv \langle\mathbf{r}'|e^{-\Delta tT}|\mathbf{r}\rangle = \left(\frac{m}{2\pi\Delta t}\right)^{d/2}\exp\left(-\frac{m}{2\Delta t}(\mathbf{r}-\mathbf{r}')^2\right), \tag{10}$$

where d stands for the dimension of the problem.

The idea behind the DMC method is to obtain numerical estimators of the right-hand side of Eq. (9) by use of sampling techniques. To do so, the repeated iteration of the propagator $e^{-(H-E_T)\Delta t}$ is simulated stochastically using a population of (weighted) random walkers. At each iteration, a random walker at $\mathbf{r}$ is weighted by a factor $w(\mathbf{r}) = e^{-\Delta t[V(\mathbf{r})-E_T]}$ and then diffused according to a (d-dimensional) Gaussian step of variance $\Delta t/m$. This is analogous to the Feynman separation of the time evolution operator into two separate propagators (corresponding to the potential and the kinetic energy) [24]. The change in the total population weight can be used to estimate the energy E_0 of the ground state. Indeed, a naive normalization (or growth) estimator for E_0 can be written as

$$E_N = E_T - \frac{1}{\Delta t}\ln\left(\frac{W^{(n+1)}}{W^{(n)}}\right) \tag{11}$$

with $W^{(n)}$ being the total weight of the walkers at the nth iteration. As first noted in Ref. [19], this estimator is biased even after relaxation (after the excited states have decayed away) as a consequence of the statistical fluctuations of the $W^{(n)}$'s. The point is that any estimator should be expressed in terms of matrix elements. The error in Eq. (11) can thus be interpreted as coming from taking the ratio before doing the average. To correct this, one should write instead

$$E_N = E_T - \frac{1}{\Delta t}\ln\left(\frac{\langle 1|e^{-\Delta t(H-E_T)}|\psi_0\rangle}{\langle 1|\psi_0\rangle}\right), \tag{12}$$

where $\langle 1| \equiv \int d\mathbf{r}\langle\mathbf{r}|$. These matrix elements must be estimated separately using Eq. (9), the ratio being taken afterwards. (It is true that the statistical errors in the numerator and denominator will introduce a small bias in E_N because it is a nonlinear function of its arguments. However, this bias will disappear with increasing statistics.)

Another estimator of E_0, based on the coordinate positions of the random walkers, can be constructed by use of the expression

$$E_m = \frac{\langle\psi_T|H|\psi_0\rangle}{\langle\psi_T|\psi_0\rangle}, \tag{13}$$

where $|\psi_T\rangle$ is an arbitrary trial wave function, and E_m is called the mixed (or trial or variational) estimator of E_0. As was the case for the growth estimator, these matrix elements can be evaluated using Eq. (9); for instance, $\langle\psi_T|\psi_0\rangle$ can be estimated numerically by

$$\langle\psi_T|\psi_0\rangle \sim e^{(E_0-E_T)t}\langle\psi_0|\phi\rangle^{-1}\ \frac{1}{N}\sum_{i=1}^{N}\psi_T(\mathbf{r}_i^{(M)})\phi(\mathbf{r}_i^{(0)})$$
$$\times\prod_{n=1}^{M} w(\mathbf{r}_i^{(n-1)}) \ . \tag{14}$$

So far, we have only considered averages. For the above estimators to be useful, their variance cannot be too large. In fact, what counts is the signal to noise ratio, that is, the *relative* variance defined as the ratio of the variance to the average squared. It is relatively easy to show that the above estimator has a relative variance that grows exponentially with $n\Delta t$, which is the physical (imaginary) time over which the evolution operator is applied. Since it is necessary to take this time large enough to eliminate excited states, this approach is very

inefficient. In practice, one has to find ways to reduce the variance of estimators. A natural way to do that is to guide the random walk rather than to use the simple diffusion operator. For completeness, we now very briefly describe this procedure.

One starts with the density function $\rho(\mathbf{r},t) \equiv \psi(\mathbf{r},t)\psi_T(\mathbf{r})$, with $\psi_T(\mathbf{r})$ being an arbitrary (time-independent) trial wave function. The guided random walk is obtained by modifying the diffusion procedure so as to sample $\rho(\mathbf{r},t)$ rather than $\psi(\mathbf{r},t)$. For this, one uses the propagator $\psi_T(\mathbf{r})e^{-(H-E_T)\Delta t}\psi_T^{-1}(\mathbf{r})$. Physically, this propagator corresponds to a diffusion in an external potential [related to $\nabla\psi_T(\mathbf{r})$], implemented using the modified weight factor $w(\mathbf{r}) = e^{-\Delta t[E_L(\mathbf{r})-E_T]}$ with the "local" energy $E_L(\mathbf{r}) = \psi_T^{-1}(\mathbf{r})H\psi_T(\mathbf{r})$. At large times, the density function $\rho(\mathbf{r},t)$ converges to $e^{-(E_0-E_T)t}\rho_0(\mathbf{r})$ with $\rho_0(\mathbf{r}) = \psi_0(\mathbf{r})\psi_T(\mathbf{r})$. Therefore, one can write a mixed estimator of E_0 as

$$E_m = \frac{\langle\psi_T|H|\psi_T^{-1}\rho_0\rangle}{\langle\psi_T|\psi_T^{-1}\rho_0\rangle}\ , \tag{15}$$

this estimator being calculated by averaging separately the numerator and the denominator along the MC run for the reason mentioned above. If $|\psi_T\rangle$ is chosen close to the exact $|\psi_0\rangle$, the variance on the energy estimators will be significantly smaller than in the case of no guiding. In fact, in the limit $|\psi_T\rangle \to |\psi_0\rangle$, the evolution operator is proportional to a stochastic matrix and the variance of the estimators E_N and E_m vanishes. Of course, if we set $\psi_T(\mathbf{r}) \equiv 1$, the evolution comes back to the standard $e^{-(H-E_T)\Delta t}$, so that the random walk is unguided. Note that the unguided random walk asymptotically samples $|\psi_0\rangle$, whereas the guided one samples the physical probability distribution $|\psi_0|^2$ (to the extent to which $|\psi_T\rangle$ is a good approximation of $|\psi_0\rangle$). In the following sections, we will restrict the analysis to unguided random walks [$\psi_T(\mathbf{r}) \equiv 1$] for simplicity, but our results could be extended to the guided case.

B. Replication, renormalization, and the stationary distribution

As mentioned previously, if one follows N random walkers, accumulating the weights through successive iterations, the relative variance on the observables increases exponentially in n, the number of iterations. In fact, we have calculated that this variance grows as $\exp\left(n\Delta t\langle H - E_L(\mathbf{r})\rangle\right)$, where $\langle\ \rangle$ stands for the expectation value in the ground state. In the unguided case, this becomes $\exp(n\Delta t\langle T\rangle)$, where T is the kinetic energy operator. Since it is necessary to follow a large number of steps n to project out the ground state, this approach leads to large statistical noise, unless the trial wave function used for guiding the random walk is quite close to the exact wave function (see [25]). Thus, in practice, one uses an additional variance reduction scheme, called replication. In this process, walkers with small weights are discarded and walkers with large weights are duplicated. Replication is implemented so as to preserve all average quantities.

Among workers in the field, the simplest replication is very frequently used (see, e.g., [9,17–19,26]). It consists of considering the fractional part of the weights as a probability in order to deal with walkers of unit weight only. For each walker of weight w, one creates $\mathrm{int}(w+\xi)$ walkers of unit weight, where ξ is a random number uniformly distributed in $[0,1]$. Thus weights less than 1 are considered "small" and weights greater than 1 are considered "large." For reasons that will be apparent soon, it is convenient to generalize this standard replication procedure to any dividing value V between large and small: it is enough to first divide all the weights by V, apply the above replication algorithm to these rescaled weights, and correct afterwards by multiplying all the weights by V. (The standard replication simply corresponds to the choice $V = 1$; in addition, various other modifications to this procedure have been proposed [16,21].) The use of replication leaves average values unaffected (for instance, estimators have the same average value), but lowers the variance.

We assume the replication procedure given and now move on to what we call renormalization. An important consequence of replication is that the number of walkers is no longer constant in time. Furthermore, replication introduces noise that is amplified with time. It can be shown that, even if $E_T = E_0$ exactly, the fluctuations in the total weight of the walkers increase indefinitely with the number of iterations. For the standard replication, the weights are equal to 1, so it is the number of walkers that has a diverging variance. This is inconvenient for practical reasons: computer memory is limited, and one does not want the population size to vanish. Thus it is common procedure to implement a kind of renormalization [9,17–19,26] to keep the number of walkers in some convenient range. The notion of exactly what constitutes a renormalization is not part of the standard lore and sometimes renormalization is referred to as "population control" [23]. The simplest explanation of this procedure follows from the history of the GFMC and DMC methods. First, in order to avoid large fluctuations in the total weight or number of walkers, it is convenient to replicate using a time dependent V. A simple choice is $V^{(n)} = W^{(n)}/\mathcal{N}$, where $W^{(n)}$ is the total population weight at step n. This choice ensures that the population size stays near its target value $\mathcal{N}$. [In the notation used in the Introduction, this corresponds to evolving according to $A/\lambda(n)$ with $\lambda(n) = V^{(n)}$. It is also equivalent to adjusting the trial energy E_T during the DMC simulation.] Second, in the early days of the DMC method, the main energy estimator used was the naive normalization (or growth) estimator [Eq. (11)]. If one applies this estimator using the ratio of the total population weight at the beginning and at the end of each Δt evolution (during which there is no replication), the target weight value V does not enter, so it does not really matter whether or not one rescales the weights by V after replication. Thus it became common practice to simply skip the phase where the weights are multiplied by $V^{(n)}$. In view of this, we hereby define *renormalization* (or, equivalently, population control) as the process of maintaining the population size near its target *and* omitting to rescale the weights by

$V^{(n)}$ after replication. As a consequence, all the weights are equal to 1 after replication with renormalization.

With the introduction of renormalization, the Markov process that evolves the random walkers has a stationary distribution. This process describes the dynamics of a population of $N^{(n)}$ walkers of coordinates $\{\mathbf{r}_i^{(n)}\}$, with $i = 1, \ldots, N^{(n)}$. We now wish to make sense of the common claim that this distribution gives a sample of the exact ground state $|\psi_0\rangle$. Thus we need to "project" the (population) stationary distribution to extract a wave function of a single configuration $\mathbf{r}$. However, such a projection is ambiguous because $N^{(n)}$ varies with n. Perhaps the simplest definition of a stationary wave function $|\psi^*\rangle$ is to say that the $\mathbf{r}_i$ are generated with a relative frequency $\psi^*(\mathbf{r}_i)$, allowing for the simple definition

$$\psi^*(\mathbf{r}) = \mathrm{E}\left[\delta(\mathbf{r} - \mathbf{r}_i)\right] . \tag{16}$$

The average, denoted by E [ ], is over all the $\mathbf{r}_i$ that could be generated along the run. Alternatively, one may consider the other natural definition of $|\psi^*\rangle$: at each time step, one has a sample of $|\psi^*\rangle$, which means that one should take all the walkers at a given time and weight them with $1/N^{(n)}$. This interpretation leads to the following definition:

$$\psi^*(\mathbf{r}) = \mathrm{E}\left[\sum_{i=1}^{N^{(n)}} \delta(\mathbf{r} - \mathbf{r}_i^{(n)})/N^{(n)}\right] . \tag{17}$$

Note that these two interpretations are identical if $N^{(n)}$ is independent of n, which is the case in our model calculation of Sec. III. Our point in this paper is that neither definition of $|\psi^*\rangle$ gives the ground state distribution $|\psi_0\rangle$. The stationary distribution of the Markov process *cannot* be interpreted as being proportional to the ground state; there is a bias which we will show decreases as $1/N$. This bias on the wave function survives asymptotically, even after the trivial bias due to the initial condition (i.e., $|\phi\rangle \neq |\psi_0\rangle$) has disappeared.

Given the result $|\psi^*\rangle \neq |\psi_0\rangle$, it is clear that all the standard energy estimators will also be biased. Interestingly, almost from the very beginnings of the GFMC and DMC methods, it was realized that the naive normalization (or growth) estimator was biased [19,20]. It was considered that the bias came from the correlation between the numerator and the denominator. Thus it was suggested that one takes averages first before ratios [20]. This improves the situation, but does not eliminate the bias precisely because the stationary distribution is still not the ground state. Alternatively, one can regard the feedback of the number of walkers (or the total weight) into E_T (which is adjusted along the random walk) as the origin of the bias, called therefore the population control bias in Ref. [23]. Another procedure for reducing the bias was proposed by Reynolds *et al.* [9], whereby the renormalization (or population control) was applied as infrequently as possible. We will discuss this in Sec. V D. Qualitatively, one can understand which way the bias goes: the throwing away of the absolute weight in the renormalization (neglecting to rescale the weights by $V^{(n)}$ after replication) tends to oversample the classically forbidden regions; the energy thus found is too large compared to the exact ground state value. The conclusion is that all simple-minded estimators (cf. Sec. V D) will be biased because the stationary wave function $|\psi^*\rangle$ is.

In the next section, we propose a simple model of replication and renormalization that enables us to derive the form of the bias. The case of the standard replication and renormalization will be considered in Secs. IV and V.

III. SYSTEMATIC ERROR DUE TO RENORMALIZATION IN A SIMPLE MODEL

A. Evolution equation for the wave function

In the DMC method, there is usually a residual time-step error due to the use of an approximate infinitesimal propagator. Thus computer runs are traditionally repeated with different Δt's in order to extrapolate to the $\Delta t = 0$ limit. This error is not our concern in this paper; we will assume that the time step is small enough to make this error negligible, making the DMC method analogous to the GFMC method. Then, in both methods, the random process converges to a stationary distribution $|\psi^*\rangle$ that is supposed to be equal to the exact ground state. However, we will show that the finite size of the population of random walkers introduces a bias in the method. Because of the use of renormalization (population control), $|\psi^*\rangle$ and $|\psi_0\rangle$ differ by on the order of $1/N$, the typical number of walkers. This will be shown in the DMC method, but the conclusion is valid for the GFMC method as long as population control is also used.

For the sake of simplicity, we analyze here the case of a particle in a one-dimensional potential governed by the Hamiltonian $H = p^2/2m + V(x)$. This example illustrates the origin of the bias, but our results hold for the multidimensional case (as well as for problems in other representations or basis sets). We follow the time evolution with replication and renormalization of an ensemble of N points $\{x_i\}$ chosen from an arbitrary initial distribution $\psi(x)$. Our aim is to show that the evolved distribution will asymptotically lead to a wave function that is *not* the ground state solution $\psi_0(x)$ of the time-independent Schrödinger equation

$$-\frac{1}{2m}\psi_0'' + [V(x) - E_0]\psi_0(x) = 0, \tag{18}$$

where E_0 is the ground state energy. We take the wave function (*not* its square) to be a probability distribution, so that $\int \psi(x)\, dx = 1$. We evolve the N points using the infinitesimal imaginary time evolution operator. Its matrix elements (with $\hbar \equiv 1$ and $E_T \equiv 0$) are given by

$$U_{\Delta t}(x', x) = \langle x'|e^{-\Delta t\, H}|x\rangle = P(x', x)w(x) + O(\Delta t^2), \tag{19}$$

where $w(x) = e^{-\Delta t\, V(x)}$ is the weight associated to the point x and $P(x', x)$ is a Gaussian distribution in x' with

mean x and variance $\Delta t/m$. Note that the normalization condition for the probability distribution $P(x',x)$ imposes that $w(x) = \int U_{\Delta t}(x',x)\,dx'$, implying that both $w(x)$ and $P(x',x)$ always have an implicit dependence on Δt, for any choice of breakup of H. After evolution by Δt, we replicate in the following way. The weights after evolution are normalized in order to yield a probability distribution for the position after evolution. Then we average over all the possible realizations (according to the initial wave function), thus giving a continuous probability distribution which yields the wave function after evolution. This leads to an evolution equation for the wave function as a function of imaginary time.

We will apply estimators acting on both the initial ensemble of points $\{x_i\}$ and the corresponding ensemble of points $\{x_i'\}$ after the evolution for a time Δt. Consider the random variables

$$W = \frac{1}{N}\sum_{i=1}^{N} w(x_i), \tag{20}$$

$$F(\tilde{x}) = \frac{1}{N}\sum_{i=1}^{N} w(x_i)\delta(\tilde{x} - x_i'). \tag{21}$$

W is an estimator for the average weight of the population of points after time evolution, from which the standard normalization (or growth) estimate for the energy E_0 can be deduced:

$$E_N = -\frac{1}{\Delta t}\ln(E[W]) \tag{22}$$

with $E[\,]$ denoting the expectation value. Also, the random variable $F_N(\tilde{x}) = F(\tilde{x})/W$ is an estimator of the *normalized* wave function $\tilde{\psi}(\tilde{x})$ after time evolution, that is,

$$\tilde{\psi}(\tilde{x}) = E[F_N(\tilde{x})] \tag{23}$$

with obviously $\int \tilde{\psi}(\tilde{x})\,d\tilde{x} = 1$.

We are thus interested in determining the expectation value of both estimators W and F_N. As we are only concerned with the $\Delta t \to 0$ limit, the calculation is carried out by expanding $E[W]$ and $E[F_N]$ in powers of Δt, keeping only the leading terms. This is done in Appendix A, along with a determination of the covariance $\mathrm{cov}[F,W]$. In our simple model, the points x_i are independent, leading to a feasable calculation of this covariance. The drawback of our choice of replication is that it cannot be readily implemented in practice. For more general choices, the correlations induced by the replication and renormalization process modify the stationary state from the calculated one, but the scaling in N remains the same. Note that the degree of correlation between the walkers is dependent on the details of the replication algorithm, so that we have not been able to calculate exactly the stationary state for replication procedures other than the one used here. On the other hand, it is assumed in this model that population control is continuously achieved (i.e., at each time step in the limit $\Delta t \to 0$), in contrast with more refined Monte Carlo procedures [9]. However, as we will explain in Sec. V D, this does not change our conclusions.

Given the above estimators and population control method, the distribution of $\tilde{x}$ after time evolution by Δt is well defined. As shown in Appendix A, one has

$$\begin{aligned}\tilde{\psi}(\tilde{x}) = E[F_N(\tilde{x})] &= \psi(\tilde{x}) + \frac{\Delta t}{2m}\psi''(\tilde{x}) \\ &-\Delta t\left(1-\frac{1}{N}\right)[V(\tilde{x}) - E_0]\psi(\tilde{x}) \\ &+\Delta t\left(1-\frac{1}{N}\right)\psi(\tilde{x})\int dx\psi(x) \\ &\times[V(x) - E_0] + O(\Delta t^2).\end{aligned} \tag{24}$$

Letting Δt tend to zero, one obtains the continuous time evolution equation for the wave function:

$$\begin{aligned}\frac{\partial\psi(\tilde{x},t)}{\partial t} &= \lim_{\Delta t\to 0}\frac{\tilde{\psi}(\tilde{x}) - \psi(\tilde{x})}{\Delta t} \\ &= \frac{1}{2m}\psi''(\tilde{x}) - \left(1-\frac{1}{N}\right)[V(\tilde{x}) - E_0]\psi(\tilde{x}) \\ &\quad + \left(1-\frac{1}{N}\right)\psi(\tilde{x})\int dx[V(x) - E_0]\psi(x)\ .\end{aligned} \tag{25}$$

Except for the scale factor $(1 - 1/N)$ in the energies, this is exactly the evolution equation for the normalized wave function $\phi(x)/\int\phi(x')\,dx'$, with $\phi(x)$ obeying the standard imaginary time Schrödinger equation. Thus the third term on the right-hand side of (25) simply keeps the norm of $\psi(x)$ constant. The correction in $1/N$ originates from the correlation between F and W in the expression of this (normalized) wave function.

B. Stationary distribution

The stationary distribution of our stochastic procedure is a solution of the time-independent equation

$$\begin{aligned}-\frac{1}{2m^*}\psi''(x) + [V(x) - E_0]\psi(x) \\ - \psi(x)\int dx'\ [V(x') - E_0]\psi(x') = 0\end{aligned} \tag{26}$$

with the *effective* mass m^* defined as

$$m^* = m\left(1-\frac{1}{N}\right)\ . \tag{27}$$

Note that by integrating over x the time-independent Schrödinger equation (18) for the ground state wave function, one gets the identity

$$E_0 = \int dx\ V(x)\psi_0(x)\ , \tag{28}$$

where E_0 is the *true* (i.e., with $N \to \infty$) ground state energy and $\psi_0(x)$ is the exact ground state wave function. This expression corresponds to our mixed estimate for the ground state energy E_0, in view of the normalization of $\psi_0(x)$. As a consequence of (28), $\psi_0(x)$ is also solution

of the equation

$$-\frac{1}{2m}\psi_0'' + [V(x) - E_0]\psi_0(x) - \psi_0(x)\int dx' \, [V(x') - E_0]\psi_0(x') = 0 \ , \quad (29)$$

equivalent to our Eq. (26) except for the correction in $1/N$. Now we define $\psi^*(x)$ as the solution of the standard Schrödinger equation for a particle of mass m^*,

$$-\frac{1}{2m^*}\psi^{*\prime\prime} + [V(x) - E^*]\psi^*(x) = 0, \quad (30)$$

where E^* stands for the *perturbed* ground state energy. Integration as before over x yields

$$E^* = \int dx \ V(x)\psi^*(x) \quad (31)$$

and, then, taking $\psi(x) = \psi^*(x)$, expression (26) becomes an identity. Therefore, $\psi^*(x)$ is also a solution of our time-independent equation [Eq. (26)]; it thus corresponds to the stationary solution of our evolution equation (25), associated with an energy E^*.

In summary, we have shown that the sampling procedure leads to a stationary state that is equal to the solution of a modified problem with an effective mass m^*. Since the effective mass is always lower than the real mass m, the resulting wave function penetrates deeper in the classically forbidden regions and alternatively decreases in the classically allowed regions. As a consequence, the mixed energy estimate E^* [i.e., Eq. (31)] is clearly larger than the exact value E_0. This tendency is in qualitative agreement with what was noticed by several authors (see, e.g., [16,23]). It was shown, e.g., in Ref. [23] that if a fluctuation increases the fraction of walkers in a region where $V(x) < E_0$, the population size should increase, but the population control will moderate this trend. The consequence is that the equilibrium distribution $\psi^*(x)$ is too small for low $V(x)$ and too high for high $V(x)$. However, this somewhat qualitative result has never been related to an effective mass of the particle. Here we have devised a method to understand and quantitatively estimate the bias in the wave function and observables. Note that, for $N = 1$, the effective mass is zero or, alternatively, there is no potential, so that the particle undergoes free diffusion as expected.

One may also calculate the normalization (or growth) energy estimate [see Eq. (22)] by use of the expression of $E[W]$ derived in Appendix A [see Eq. (A2)], to find

$$E_N = E_0 - \frac{1}{\Delta t}\ln\left[1 - \Delta t \int dx \ \psi^*(x) \times [V(x) - E_0] + O(\Delta t^2)\right] = E_0 + \int dx \ \psi^*(x)[V(x) - E_0] + O(\Delta t) \ , \quad (32)$$

similar to Eq. (31), given the normalization convention for $\psi^*(x)$. Thus the normalization energy estimate will also be biased (with the same bias as the mixed energy estimate). Now, in order to estimate this energy bias (due to a finite N), we make use of a simple perturbation calculation (expansion in $1/N$). The derivation of the biased wave function $\psi^*(x)$ is reported in Appendix B. With Eqs. (27) and (30), the Hamiltonian of the modified problem can be written as $H = H_0 + \Delta H$, where $H_0 = T + V$ is the Hamiltonian of the original problem, and $\Delta H = T/N$ is the perturbation. Thus the energy bias can be written as

$$\Delta E = E^* - E_0 = \langle\psi_0|\Delta H|\psi_0\rangle = \langle T\rangle_0/N \ , \quad (33)$$

where $\langle T\rangle_0$ stands for the expectation value of the kinetic energy in the *unperturbed* ground state. In the case of guided random walks (Sec. II A), the operator T is to be replaced by $H - E_L(\mathbf{r})$. This expression proves quite naturally that the energy bias is always positive and of order $1/N$, as already mentioned. It has been already noticed in Refs. [16,23] that the MC estimator tends to overestimate the ground state energy and that the bias scales like $O(N^{-1})$, but its exact form was not known. It is worth noticing that, although the factor $1/N$ suggests that the bias will rapidly become small with increasing population sizes (e.g., $N \sim 100$), this is not necessarily the case because the factor $\langle T\rangle_0$ may be large, as we observe for our problem in discrete space (see Sec. IV). Equation (33) is also checked in Appendix B for the simple case of an harmonic oscillator. It is shown that the perturbed ground state wave function can be written as

$$\psi^*(x) = \psi_0(x) + \frac{1}{8N}[\psi_2(x) - \psi_0(x)], \quad (34)$$

where $\psi_2(x)$ stands for the second excited state wave function (normalized with our convention). The resulting estimate for the ground state energy is then given by

$$E^* = E_0 + \frac{1}{N}\langle T\rangle_0 = \omega\left(\frac{1}{2} + \frac{1}{4N}\right). \quad (35)$$

Our result is general and applies also for many-dimensional problems as well as problems in discrete space (see Sec. IV). The only condition is that the time evolution operator $\exp(-H\Delta t)$ can be divided into a diffusion part $\exp(-T\Delta t)$ and a weight $\exp(-V\Delta t)$. The diffusion operator must be such that $\langle 1|\exp(-T\Delta t)|x\rangle = 1$ for all x, with $\langle 1| \equiv \int dx\langle x|$, which amounts to saying that the associated matrix must be stochastic. One can also extend our result to the case of stochastic iteration of matrices (cf. [16]). In this case, following the notation of Sec. I, we can see that the random walk tends to a stationary state that is the dominant eigenvector of the modified matrix

$$A^*_{ij} = M_{ij}{w_j}^{1-1/N}, \quad (36)$$

differing from the exact one by a $1/N$ correction. When $N = 1$, one has trivially $A^*_{ij} = M_{ij}$, as already pointed out in Sec. I.

The major hypotheses used to obtain Eqs. (26), (27), and (33) are (i) a very small time step Δt, (ii) population control at each time step, and (iii) statistical independence between the random walkers. This last hy-

pothesis is the most stringent one; it was valid in our idealized replication and renormalization process, but we cannot hope to get better than the order of magnitude of the bias in other replication and renormalization procedures. Nevertheless, in many cases, it will be possible to calculate the scaling of the bias with different parameters of the problem (see, e.g., Sec. IV). The other hypotheses are much less stringent. For (i), it is sufficient to use reasonably small values for Δt. Moreover, it is possible to improve the convergence in Δt by using a symmetric form for the breakup (see [27]): $e^{-H\Delta t} = e^{-V\Delta t/2}e^{-T\Delta t}e^{-V\Delta t/2} + O(\Delta t^3)$. With this breakup, extrapolation to small Δt's can be unnecessary and assumption (i) is satisfied. Note also that there is no time-step error at all in the GFMC method since the propagator is exactly simulated. Finally, for (ii), performing a renormalization at each time step in our model is not essential. Indeed, we could have done otherwise, the stationary distribution $|\psi^*\rangle$ not being very sensitive to the renormalization period. Suppose, for instance, that renormalization is done every k time steps; when Δt is small, the finite limit for $|\psi^*\rangle$ as Δt tends to zero implies that the result is relatively insensitive to k. Thus the above model analysis should be applicable to cases such as that in Ref. [9], where renormalization is rather infrequent. In fact, we will see in Sec. V D that another replication and renormalization procedure has $|\psi^*\rangle$ *perfectly* independent of the renormalization period for any Δt.

IV. EXAMPLE IN A DISCRETE SPACE: PAIRING HAMILTONIAN

The purpose of this section is to show that the previously described bias can be important and indeed scales as $1/N$ for the standard replication and renormalization procedure. We consider the so-called pairing Hamiltonian that describes the residual interaction between nucleons in nuclei. This many-body system is described by the Hamiltonian

$$H = \sum_{k=1}^{\Omega} \epsilon_k (a_k^\dagger a_k + a_{\bar{k}}^\dagger a_{\bar{k}}) - G \sum_{k,k'=1}^{\Omega} a_{k'}^\dagger a_{\bar{k}'}^\dagger a_{\bar{k}} a_k, \quad (37)$$

where k and $\bar{k}$ are time-reversed conjugate states (with energies ϵ_k), Ω is the total number of conjugate state pairs, and G (≥ 0) stands for the strength of the pairing force. One is interested in calculating the exact ground state energy E_0 for a system of n pairs of particles by a diffusion Monte Carlo procedure. We define the breakup $H = V + T$, with

$$V = \sum_{k=1}^{\Omega} \epsilon_k (a_k^\dagger a_k + a_{\bar{k}}^\dagger a_{\bar{k}}) - E_{sen},$$

$$T = E_{sen} - G \sum_{k,k'=1}^{\Omega} a_{k'}^\dagger a_{\bar{k}'}^\dagger a_{\bar{k}} a_k \ , \quad (38)$$

where $E_{sen} = n(\Omega - n + 1)G$ is the seniority energy, that is, the ground state energy E_0 in the case where all the single-particle levels are degenerate (i.e., $\epsilon_k = 0$ for all k). The operator $\exp(-T\Delta t)$ is then a diffusion operator, as can be seen by checking that $\sum_{C'} \langle C'|T^l|C\rangle = 0$ for all integers $l > 0$ and for any configuration C. The diffusion follows a Poisson law (see [28,29]) and $\exp(-V\Delta t)$ can be used as a weight since it is diagonal in C space. In analogy to Eq. (28), the ground state energy is given by

$$E_0 = \sum_C V(C)\psi_0(C), \quad (39)$$

where $\psi_0(C)$ is the component of C in the exact ground state, $V(C) = \langle C|V|C\rangle$, and we normalize $|\psi_0\rangle$ so $\langle 1|\psi_0\rangle = 1$. Then a MC mixed estimator for E_0 is given by

$$E_m = \frac{1}{N} \sum_{i=1}^{N} V(C_i) \ , \quad (40)$$

where the $\{C_i\}$ are the configurations generated by the (unguided) MC run. As mentioned previously, this estimator is biased because the $\{C_i\}$ ensemble is not distributed as $|\psi_0\rangle$. The same conclusion holds for the growth estimators for E_0.

In Fig. 1 we give the dependence of the computed ground state energy for a typical model system with $n = 12$ pairs of particles in $\Omega = 24$ equispaced doubly degenerate levels, with a pairing strength $G = 0.5$ (ex-

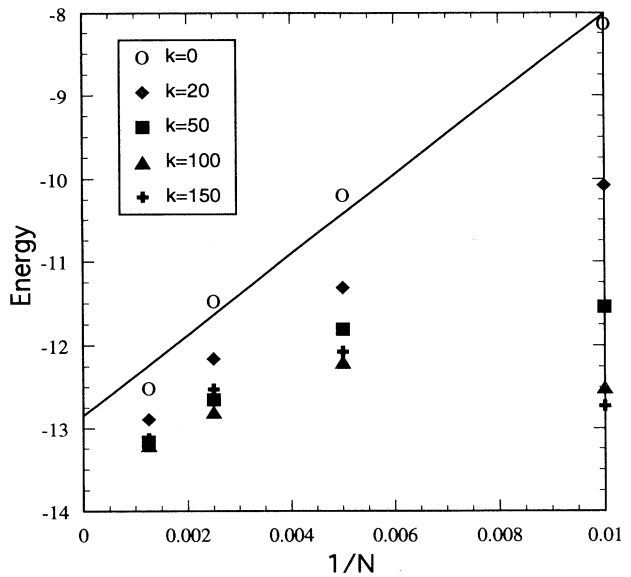

FIG. 1. Growth estimator for the ground state energy as a function of $1/N$, with N being the size of the population of random walkers. The data fit to a line corresponds to the simple estimator (i.e., $k = 0$). The corrected estimators (as discussed in Sec. V D) for $k = 20, 50, 100, 150$ are also shown. The model consists of $n = 12$ pairs of particles placed in $\Omega = 24$ equispaced doubly degenerate levels and interacting through a pairing force of strength $G = 0.5$ (see [28,29]). The exact value of the energy is 13.096 (expressed in units of the level spacing).

pressed in units of the level spacing). The data points of interest to us here (i.e., with $k = 0$, as explained in Sec. V D) have been fitted to a linear law in $1/N$; one sees that the fit is very good. To get 1% accuracy without extrapolation, it is necessary to take $N \geq 3000$. The data shown are for the mixed estimator E_m, but the other estimators (such as the growth estimator with different interpretations for $|\psi^*\rangle$ and ways to take the ratios) give almost identical results. We have also considered observables other than the energy. In general, their estimators have still larger biases, so that it is necessary to go to larger values of N before being able to extrapolate in $1/N$.

It is of interest to understand how this bias depends on the model parameters. When $G = 0$ (i.e., no pairing), we have $T = 0$, in which case the bias disappears. For degenerate states (i.e., $\epsilon_k = 0$ for all k), we have $V = E_0 = -E_{sen}$, so that $\langle T \rangle = 0$ and again the bias is zero. Thus the bias vanishes for both limiting cases and tends to a maximum somewhere in between as we have confirmed numerically [29]. Furthermore, when the model space dimension Ω increases, with G varying such that the ground state energy is approximately independent of Ω, it is easy to show that $\langle T \rangle$ is approximately proportional to E_{sen}. Then the bias for the energy scales as Ω^2/N, which explains why it becomes important for large model spaces (see [29]).

V. SOLUTIONS TO SUPPRESS THE BIAS

Section III showed how to estimate the magnitude of the bias both for the wave function and for observables such as the energy. If the bias is comparable to or larger than the statistical error, it is necessary to remove it or at least reduce it. The simplest procedure consists in extrapolating to the $N \to \infty$ limit. Other possibilities include changing the evolution or replication or renormalization procedure so that the bias on the wave function is smaller. We discuss these three choices successively and illustrate them on simple model Hamiltonians. Finally, we consider modifications of the estimator itself that remove most of the bias whether or not $|\psi^*\rangle$ is close to $|\psi_0\rangle$.

A. Extrapolation in N

We saw in Sec. IV that an energy estimator extrapolates well with a $1/N$ scaling for large N. For a different observable or a different problem, the onset of the $1/N$ scaling will occur at smaller or larger values of N. Here we quantify the convergence of $|\psi^*\rangle$ to $|\psi_0\rangle$ as $N \to \infty$. To do this, we measure how various moments of the distribution $|\psi^*\rangle$ converge to their $N = \infty$ limit, using both $|\psi^*\rangle$ definitions. We consider the observables [see Eq. (16)]

$$\mu_k \equiv \int_{-\infty}^{\infty} x^k \psi^*(x) dx \simeq \frac{\sum_i (x_i)^k}{\sum_i 1} \tag{41}$$

for a one-dimensional anharmonic oscillator to check the convergence. The potential was taken to be $V(x) = x^4 - x^2 + x/2$, with $\hbar = m = 1$. If we choose the other definition [see Eq. (17)] for $|\psi^*\rangle$, the observables become

$$\mu_k \simeq \left\langle \frac{\sum_{i=1}^{N^{(n)}} (x_i^{(n)})^k}{N^{(n)}} \right\rangle . \tag{42}$$

This last estimator corresponds to the Kalos prescription [17], i.e., the ratio is taken at each time step and then averaged over iterations. In the Ceperley-Kalos prescription [19,20], corresponding to the first estimator [Eq. (41)], the numerator and denominator are averaged separately before taking ratios, with the expectation that this will reduce the bias.

Figure 2 shows the convergence of the first four moments for the Ceperley-Kalos prescription using the standard replication [18]. Interestingly, although the Ceperley-Kalos prescription is better justified theoretically, its bias is nearly identical for that of the Kalos prescriptions (not shown). From the figure, we see that it is possible to use the $1/N$ extrapolation from very small values of N. However, for more realistic problems such as the one described in Sec. IV, much larger values of N are necessary before the linear dependence sets in.

Doing an extrapolation in N is inconvenient for two reasons. First, it is necessary to check that the values of N considered are in the $1/N$ scaling regime. Thus at least three values of N have to be used. Second, the extrapolation to $N = \infty$ magnifies the statistical errors. For instance, if the data points at N and $2N$ are used in a linear extrapolation and if each data point has a statistical error of σ, the extrapolated value has a statistical error of $\sqrt{5}\sigma$. If it were possible to eliminate the bias, the computer time could be used entirely for the data point

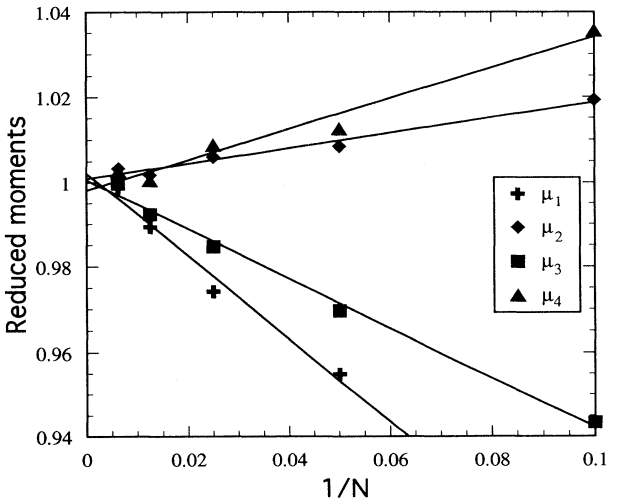

FIG. 2. Four first reduced moments ($\mu_k/\mu_k^{\text{exact}}$, with $k = 1, \ldots, 4$) as a function of $1/N$, with N being the size of the population of random walkers. The system is a one-dimensional anharmonic potential $V(x) = x^4 - x^2 + x/2$. (The time step used is $\Delta t = 0.05$.)

at N, leading to a statistical error of $\sigma/\sqrt{2}$. The ratio of the statistical errors in the two cases is then equal to $\sqrt{10}$, which corresponds to a factor of 10 of computer time. This provides strong motivation for searching for bias-free methods.

B. Modifying the evolution operator

It is well known that if the evolution is done using a guided random walk (also sometimes called importance sampling or the generalized Feynman-Kac method), the variance of estimators is reduced. In fact, if the guiding function used (cf. Sec. II) is equal to $|\psi_0\rangle$, the weights have no variance, no replication is needed, and the energy estimators are exact. A natural question is whether there exists a guiding function or equivalently a modified evolution operator that leads to $|\psi^*\rangle = |\psi_0\rangle$ in the presence of replication. Since it is difficult to analyze the effects induced by almost all replication procedures, we consider here the less ambitious goal of finding an evolution procedure for which $|\psi^*\rangle$ converges towards $|\psi_0\rangle$ faster than $1/N$.

Our approach is empirical, numerical, but motivated by the perturbative calculation of Sec. III. There we saw that, for a particular choice of replication and renormalization, $|\psi^*\rangle$ was the solution of a Schrödinger equation with a perturbed mass. This suggests that by using a different value of the mass in the evolution operator, it may be possible to reduce or eliminate the $1/N$ bias. Unfortunately, as previously remarked, the replication and renormalization used in the theoretical analysis is not readily implemented in practice. For other replication or renormalization, the analysis is not feasible because the random walkers become correlated, so that the effective ensemble size N_{eff} is less than the number of walkers. Nevertheless, the analytic calculations are suggestive, so we consider rescaling the time step appearing in the diffusion operator using

$$\Delta t \to \Delta t(1 - 1/N_{\text{eff}}) \ , \tag{43}$$

while leaving the weight factors unchanged. This is equivalent to increasing the effective mass m^* by a factor $(1 - 1/N_{\text{eff}})^{-1}$. Here, N_{eff} is something like the number of independent configurations, but in practice it is a free parameter which we set so that the first moment has no bias. Then we determine numerically the convergence of the other moments of $|\psi^*\rangle$ (cf. the study in Sec. V A) to their $N \to \infty$ limit. Though we did find that the correction improved the results, they were erratic in the sense that there was no unique value of N_{eff} that gave the correct value for all the moments. Also, the scaling of the biases remained proportional to $1/N$. We checked this within the two interpretations of observables (taking average of ratios or ratios of averages) and using the standard replication. No doubt the failure to find a $1/N^2$ convergence arises because the theoretical calculations assumed a specific replication and renormalization procedure, whereas this numerical study used the standard procedure. A different kind of replication and renormalization was studied by Hetherington [16]; he estimated the form of the bias neglecting correlations within walkers. His analysis suggests that the $1/N$ term of the bias might be removed by the modification of the weights (cf. the matrix notation in Sec. I):

$$w_j \to w_j(1 + w_j/N_{\text{eff}}) \ , \tag{44}$$

which is equivalent (to the order $1/N$) to the modification resulting from our analysis [see Eq. (36)]. Thus, roughly speaking, the weights must be enhanced in order to compensate for the factor $w_j^{1-1/N}$ in Eq. (36). We have numerically checked that one obtains the same negative result using his prescription as with the first one. Thus our conclusion is negative: we have not been able to find a way to remove the leading term of the bias. Nevertheless, we think this question is worth pursuing.

C. Modifying the replication or renormalization procedure

Above we considered possible modifications of the time evolution operator to remove the $1/N$ bias in $|\psi^*\rangle$. It is also possible to affect the bias by modifying the replication or renormalization procedure. Since there is no analytic work here to guide us, we will simply compare the magnitude of the bias for three often used replication and renormalization methods: the "standard" method [9,17–19], a method due to Hetherington [16], and an "improved" method due to Nightingale and Blöte [21]. For completeness, we first describe these three methods and then summarize in Fig. 3 their influences on the bias.

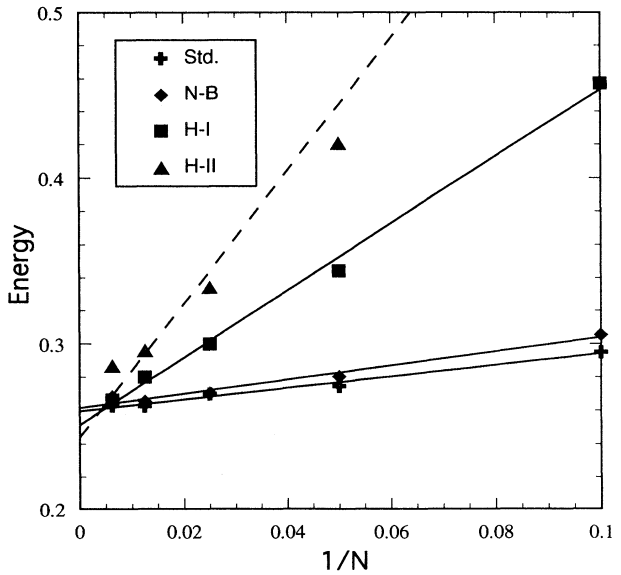

FIG. 3. Estimator of the energy as a function of $1/N$, for the three different types of replication procedures (with the same time step $\Delta t = 0.05$). Std., standard method; N-B, Nightingale-Blote method; H-I, Hetherington method. The Δt dependence of the Hetherington replication is also illustrated by the dashed line at $\Delta t = 0.025$ (H-II). The system is a one-dimensional anharmonic potential $V(x) = x^4 - x^2 + x/2$.

The standard replication procedure, exposed in Sec. II B, can also be presented in a continuous time formalism [18]. Starting with Eq. (5), we see that the potential term gives a probability per unit time to either annihilate or to multiply. Thus for each time step Δt, $[V(x) - E_0]\Delta t$ gives the decrease or the increase in the population at that point in x and for that time interval. If it is negative, a configuration at that x is to be duplicated with that probability; if it is positive, the configuration has a "death" probability. In practice, this birth and death process is implemented [17,18] by picking a random number ξ between 0 and 1 and duplicating (or "killing") whenever $\xi < |[V(x) - E_0]\Delta t|$.

Another replication was proposed by Hetherington [16]. First, the N weights of the configurations after evolution by Δt are rescaled to give a probability distribution; then N new configurations are chosen at random from this probability distribution and each is given weight 1. Although it is rarely used, this method has the advantage of keeping N fixed, enabling Hetherington to estimate the scaling of the bias (neglecting correlation effects).

The third replication discussed here is that of Nightingale and Blöte [21]; it can be motivated as follows. Imagine using the standard replication method and calculating the naive growth estimator at time step n from

$$e^{-\Delta t E_0} \simeq \left\langle \frac{\sum_{i=1}^{N^{(n)}} w_i}{N^{(n)}} \right\rangle_{\text{before}} = \left\langle \frac{N^{(n+1)}}{N^{(n)}} \right\rangle_{\text{after}}, \quad (45)$$

where $N^{(n)}$ is the number of configurations at time step n and the average is over time steps. The first estimator is evaluated after evolution but before replication. In the second estimator, obtained after replication, $\sum_{i=1}^{N^{(n)}} w_i$ is replaced by $N^{(n+1)}$, which is simply a random variable whose average is $\sum_{i=1}^{N^{(n)}} w_i$. We see therefore that measuring observables *after* replication introduces unnecessary noise. Nightingale and Blöte realized that it was possible to perform a different replication that introduced no noise at all into the total population weight, so that the growth estimator was exactly the same before and after replication in their method. The idea is to keep weighted configurations at all times, and to do the birth and death processes only when the weights have varied significantly from 1. Their replication is defined as follows. First, for a birth process, one duplicates each walker with a large weight (e.g., larger than 2): it is replaced by two identical new walkers, each with a weight equal to half of the old weight. This part is deterministic, no noise is introduced. Second, for a death process, given two walkers with low weights w_i and w_j (e.g., smaller than 1/2), one selects one of them with a probability given by its weight relative to $w_i + w_j$. The other one is killed, while the first is maintained and given the new weight $w_i + w_j$. The remaining walkers are unchanged by the replication (i.e., they keep their weights). This procedure has the appealing feature that it preserves exactly the total weight of the population during replication and preserves all other averages, just as the standard replication does.

How do these replication and renormalization procedures influence the bias $|\psi^*\rangle - |\psi_0\rangle$? Not surprisingly, the three methods lead to biases that scale with $1/N$, but with different coefficients. In addition, the Hetherington method gives rise to an exceptionally large bias because its scaling parameter is $1/(N\Delta t)$. This extra factor of Δt shows that as $\Delta t \to 0$, Hetherington's replication and renormalization is no longer effective. Indeed, in that limit, it is no longer affected by the potential and all the configurations become identical to within Δt. The effective number of random walkers N_{eff} is then equal to 1 and there is no guiding at all. To keep the bias constant as $\Delta t \to 0$ with this method, N must scale as $1/\Delta t$. In contrast, the two other methods both have finite $\Delta t \to 0$ limits and have very similar biases. In Fig. 3 we show the dependence on $1/N$ of the energy estimators for the anharmonic oscillator used in Sec. V A. Note that, as expected, the different replications all lead to a linear dependence at large N. Also, there is no improvement in this case when going from the standard to the Nightingale-Blöte replication. Probably the main advantage of the Nightingale-Blöte replication is that it always maintains a finite population; on the contrary, the standard replication can lead (with a low probability) to an empty population, so its large time behavior leads to difficulties.

D. Modifying the estimator

Once it is accepted that $|\psi^*\rangle$ is inevitably different from $|\psi_0\rangle$, it is natural to modify the estimators so as to remove the bias. The idea is to use matrix elements between $|\psi^*\rangle$ and any given trial wave function $|\psi_T\rangle$, while inserting projection operators so that only the ground state component contributes. For instance, the ratio of matrix elements

$$\frac{\langle\psi_T|e^{-(k+1)\Delta t H}|\psi^*\rangle}{\langle\psi_T|e^{-k\Delta t H}|\psi^*\rangle} = e^{-\Delta t E_0}\left[1 + O(e^{-k\Delta t(E_1 - E_0)})\right] \quad (46)$$

is a growth estimator that is unbiased in the $k \to \infty$ limit. The numerator and the denominator can be estimated simultaneously within one Monte Carlo run. It is most convenient to choose the definition of Eq. (16) for $|\psi^*\rangle$ and to consider configurations only after replication. For instance, if we take the standard replication where all weights become unity, the denominator can be estimated using

$$\langle\psi_T|e^{-k\Delta t H}|\psi^*\rangle \simeq \sum_{i=1}^{N^{(n+k)}} \psi_T\left(x_i^{(n+k)}\right) \Pi_k^{(n)} . \quad (47)$$

In this expression, n is the time-step number at which $|\psi^*\rangle$ is sampled and $\simeq$ means up to a multiplicative constant that is related to the normalization of $|\psi^*\rangle$. (It is thus the same for both the numerator and the denominator.) Finally, $\Pi_k^{(n)} = \Pi_{l=n}^{l=n+k-1}\pi(l)$ and $\pi(l)$ is the factor

by which the weights are divided before the replication or renormalization number l, aiming at maintaining the population size around some average value. (This factor was called $V^{(l)}$ in our discussion on replication; see Sec. II B.) Nightingale and Blöte [21] introduced the same estimators, which they motivated by "undoing" the renormalization process. It seems to us that the natural way to define estimators is to start with the notion of a *biased* stationary distribution $|\psi^*\rangle$ and then to express any desired quantity in terms of matrix elements.

Given the above estimator as a function of k, the exact answer (up to statistical errors) is obtained by extracting the $k \to \infty$ limit, with an exponential convergence. However, it is important to note that, in this limit, the variance of the estimators for both the numerator and the denominator grow exponentially, so one must in practice extract the limiting behavior from small values of k. One has thus to choose a compromise between accuracy (bias reduction) and statistical error. We choose to illustrate this compromise in a realistic model—the pairing Hamiltonian described in Sec. IV. We saw previously that the bias was large in that system and had a $1/N$ scaling. In Fig. 1 we show the results using the modified estimators [Eq. (46)] along with the simple estimator of Sec. IV, which corresponds to $k = 0$ using the same statistics. We see that when the error in the $k = 0$ estimator is large (N is too small), the extrapolation in k greatly improves things. However, when the bias of the $k = 0$ estimator is already small, increasing k rapidly leads to noise levels where there is no improvement.

More generally, it is possible to introduce unbiased (up to exponentially small corrections) estimators for local operators. Let $\mathcal{O}$ be a general (local) observable; if we assume that $|\psi^*\rangle = |\psi_0\rangle$, an estimator for the ground state expectation value is obtained from

$$\frac{\langle\psi_0|\mathcal{O}|\psi_0\rangle}{\langle\psi_0|\psi_0\rangle} \simeq \frac{\langle\psi_T|e^{-\tau H}\mathcal{O}|\psi_0\rangle}{\langle\psi_T|e^{-\tau H}|\psi_0\rangle} , \tag{48}$$

where τ should be taken large enough to eliminate contributions from excited states. Since we know that in fact $|\psi^*\rangle \neq |\psi_0\rangle$, a computable observable is

$$\frac{\langle\psi_T|e^{-\tau H}\mathcal{O}e^{-k\Delta t H}|\psi^*\rangle}{\langle\psi_T|e^{-(\tau+k\Delta t)H}|\psi^*\rangle} , \tag{49}$$

for which there is an obvious numerical estimator. This estimator should then be extrapolated to the $\tau, k \to \infty$ limit. (With finite statistics, the numerator and the denominator fluctuate, so taking the ratio introduces a small bias; this bias can be made arbitrarily small by extending the length of the run.)

Prior to Nightingale and Blöte, Reynolds *et al.* introduced a procedure to reduce the bias by doing population control as infrequently as possible [9]. More precisely, their runs are divided into blocks and renormalization is done only at the *end* of the blocks (e.g., every k time steps, if k is the block size or renormalization period). Inside each block, the walkers are evolved using a time-independent E_T and replicated. The population then has a size that fluctuates more than when renormalization is performed, but the correct weightings are kept during the time of a block. Thus, for instance, the growth estimator

$$E_N = E_T - \frac{1}{\Delta t}\ln\left(\frac{\langle W^{(n+1)}\rangle}{\langle W^{(n)}\rangle}\right), \tag{50}$$

with averages over runs or blocks at fixed n, has a bias that decreases exponentially with n, the time evolved since the last renormalization. On the other hand, this bias reduction is accompanied by a significant increase in statistical noise: the variance of the statistical error *increases* exponentially with n. To improve statistics, it is tempting to extend the averaging to all n values within each block; however, the resulting bias will then only decrease as the inverse of the renormalization period k (or block size). If a small systematic error is required, it is preferable to skip the first time steps of a block in the average to keep only the tail of the negative exponential. Essentially, this method is closely related to the Nightingale-Blöte method for improving estimators, the moving average being replaced by a block average. There is therefore the same compromise between bias reduction and minimization of the statistical error.

Another interesting point is the dependence of the wave function $|\psi^*\rangle$ on the renormalization period k. Here $|\psi^*\rangle$ is the stationary distribution obtained just after renormalization (i.e., every k time steps). In our idealized model of replication and renormalization (Sec. III), we have seen that $|\psi^*\rangle$ tends to a finite limit when Δt tends to zero. This suggests that the wave function $|\psi^*\rangle$ is not very sensitive to the renormalization frequency k. It turns out that, in the generalized replication defined in Sec. II B using $V^{(n)} = W^{(n)}/\mathcal{N}$ and with renormalization implemented only every k time steps (i.e., not rescaling the weights by $V^{(n)}$ after replication every k time steps), the wave function $|\psi^*\rangle$ is *totally* independent of k. Indeed, since renormalization is achieved by throwing away the normalization factor after replication, it is easy to see that this procedure leaves the the actual random walk (evolution and replication) unchanged, so that $|\psi^*\rangle$ is not affected by k. This shows that our effective mass calculation is also relevant when population control is achieved periodically or in blocks (not constantly along the evolution), even if some bias reduction is obtained through the observables in that case.

VI. CONCLUSION

We have shown how projection MC methods for calculating ground state properties are affected by a bias related to the finite size of the population of random walkers. The source of this bias is the renormalization that is introduced to avoid large fluctuations in the population size or weight. This difficulty occurs whenever the propagator differs from a diffusion operator (i.e., $\sum_i A_{ij}$ varies with j in matrix notation). The renormalization enables the population of walkers to converge to a stationary distribution which we call $|\psi^*\rangle$, that is not equal to the exact ground state $|\psi_0\rangle$. Our analytic calculation shows that, for a particular replication and renormalization procedure, $|\psi^*\rangle$ corresponds to the ground state of a modified Schrödinger equation in which the mass m is

replaced by an effective mass $m^* = m(1-1/N)$ for a population of walkers of size N. As a consequence, the bias on any observable scales as $1/N$. In particular, the bias for the energy estimator is proportional to the expectation value of the kinetic energy in the exact ground state $|\psi_0\rangle$. Nevertheless, for the replication or renormalization procedures used in practice, we have not been able to obtain the exact form of the correction, so this is still an open problem. Finally, we have considered different strategies to reduce the bias, namely, extrapolation in N, modifications of the evolution or replication procedure, or the introduction of new observables. This latter approach proves to be the most efficient as it can deal with any $|\psi^*\rangle \neq |\psi_0\rangle$.

ACKNOWLEDGMENTS

We thank M. Caffarel, W. Krauth, D. Levesque, and N. Trivedi for sharing their knowledge with us and are very grateful to P. Nightingale for pointing us to the right references. N.C. is supported by a grant from the Human Capital and Mobility Program of the Commission of the European Communities. The Division de Physique Théorique is Unité de Recherche des Universités Paris XI et Paris VI associée au CNRS.

APPENDIX A: DERIVATION OF THE MOMENTS OF W, F, AND F_N

Here we calculate the various moments of W, F, and F_N. [The calculation requires keeping first-order terms in Δt in the expansion of $w(x)$ and $P(x',x)$.] In order to calculate the expectation value of F_N, we expand in powers of Δt the expression for F_N, leading to

$$E[F_N] = \frac{E[F]}{E[W]}\left[1 - \frac{\mathrm{cov}[F,W]}{E[F]\,E[W]} + \frac{\mathrm{var}[W]}{E[W]^2}\right] + O(\Delta t^2)\ , \tag{A1}$$

where cov[ ] and var[ ] stand for the covariance and the variance of the arguments, respectively. We are interested first in calculating the expectation value of the estimators W and F. Using the evolution operator $\mathrm{e}^{-H\Delta t}$, we obtain

$$E[W] = E[w(x)] = e^{-\Delta t\ E_0}\left[1 - \Delta t \times \int dx\ \psi(x)[V(x) - E_0] + O(\Delta t^2)\right], \tag{A2}$$

where we have made use of the normalization of $P(x',x)$ and we have kept only the contribution of first-order terms in Δt in the expansion of $w(x)$. The expectation value is calculated by use of

$$E[X] = \int\int dx\ dx'\ P(x',x)\psi(x)X(x,x')\ . \tag{A3}$$

For the expectation value of $F(\tilde{x})$, we have

$$\begin{aligned} E[F(\tilde{x})] &= E[w(x)\delta(\tilde{x}-x')] \\ &= \psi(\tilde{x})w(\tilde{x}) + \frac{\Delta t}{2m}[\psi(x)w(x)]''_{x=\tilde{x}} \\ &= e^{-\Delta t\ E_0}\left[\psi(\tilde{x}) + \frac{\Delta t}{2m}\psi''(\tilde{x}) - \Delta t\ [V(\tilde{x}) - E_0]\psi(\tilde{x}) + O(\Delta t^2)\right], \end{aligned} \tag{A4}$$

where we have kept the second-order term in the Gaussian random walk expansion, yielding a term in Δt. Note that all the derivatives of $w(x)$ are neglected because they give one supplementary order in Δt. The ratio of the expectation values of both estimators is then expressed as

$$\frac{E[F(\tilde{x})]}{E[W]} = \psi(\tilde{x}) + \frac{\Delta t}{2m}\psi''(\tilde{x}) - \Delta t\ [V(\tilde{x}) - E_0]\psi(\tilde{x}) + \Delta t\ \psi(\tilde{x}) \int dx\ \psi(x)[V(x) - E_0] + O(\Delta t^2) \tag{A5}$$

after neglecting second- (and higher-) order terms in Δt. Since in our model the points x_i are independent, we get, for the covariance of F and W,

$$\begin{aligned} \mathrm{cov}[F(\tilde{x}),W] &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{cov}[w(x_i)\delta(\tilde{x}-x_i'), w(x_j)] \\ &= \frac{1}{N}[E[w^2(x)\delta(\tilde{x}-x')] - E[F(\tilde{x})]\ E[W]]. \end{aligned} \tag{A6}$$

It is worth noting that this independence is a crucial point of the calculation, as explained in Sec. III. We also have

$$\begin{aligned} E[w^2(x)\delta(\tilde{x}-x')] &= \psi(\tilde{x})w^2(\tilde{x}) + \frac{\Delta t}{2m}[\psi(x)w^2(x)]''_{x=\tilde{x}} \\ &= e^{-2\Delta t\ E_0}\left[\psi(\tilde{x}) + \frac{\Delta t}{2m}\psi''(\tilde{x}) - 2\Delta t\ [V(\tilde{x}) - E_0]\psi(\tilde{x}) + O(\Delta t^2)\right]. \end{aligned} \tag{A7}$$

In addition, we can express the covariance of F and W as

$$\mathrm{cov}[F(\tilde{x}),W] = -\frac{1}{N}e^{-2\Delta t\ E_0}\left[\Delta t\ [V(\tilde{x}) - E_0]\psi(\tilde{x}) - \Delta t\ \psi(\tilde{x}) \int dx\ [V(x) - E_0]\psi(x) + O(\Delta t^2)\right]. \tag{A8}$$

Finally, it is easy to see that the contribution of the third term on the right-hand side of Eq. (A1) is of the order Δt^2 and thus the variance of W can be neglected. Then, using (A1), (A2), (A4), (A5), and (A8), we obtain, for the expectation value of F_N,

$$E[F_N(\tilde{x})] = \psi(\tilde{x}) + \frac{\Delta t}{2m}\psi''(\tilde{x}) - \Delta t\left(1-\frac{1}{N}\right)[V(\tilde{x})-E_0]\psi(\tilde{x}) + \Delta t\left(1-\frac{1}{N}\right)\psi(\tilde{x})\int dx\ \psi(x)[V(x)-E_0] + O(\Delta t^2)\ . \quad \text{(A9)}$$

APPENDIX B: BIAS IN THE WAVE FUNCTION

We are now searching for an expression of the biased wave function $\psi^*(x)$. Using Eq. (31), the energy bias is written as

$$\Delta E = E^* - E_0 = \int dx\ V(x)[\psi^*(x) - \psi_0(x)]. \quad \text{(B1)}$$

Let us write

$$\psi^*(x) = \psi_0(x) + \frac{1}{N}f(x) \quad \text{(B2)}$$

since it is clear that the sampling introduces a perturbation of the order $1/N$ (via the effective mass). Note that the normalization of $\psi^*(x)$ imposes the condition

$$\int f(x)\ dx = 0\ . \quad \text{(B3)}$$

As a consequence of (B1), it is evident that the energy bias will also be of order $1/N$, but we are interested here in an absolute value, not just the scaling with N. Inserting (B2) into Eq. (26) and keeping only the first-order term in the expansion in $1/N$ leads to the equation for $f(x)$:

$$-\frac{1}{2m}f'' + [V(x)-E_0]f(x) - [V(x)-E_0]\psi_0(x) - \psi_0(x)\int dx\ V(x)\ f(x) = 0, \quad \text{(B4)}$$

where we have made use of Eqs. (18) and (28). It is apparent from Eq. (B1) that the fourth term can be written $\psi_0(x)\ N\ \Delta E$ and thus this equation can also be used to deduce ΔE. Let us expand $f(x)$ in terms of the eigenfunctions of the *unperturbed* Schrödinger equation:

$$f(x) = \sum_{n=0}^{\infty} a_n\ \psi_n(x) \quad \text{with} \quad -\frac{1}{2m}\psi_n'' + [V(x)-E_n]\psi_n(x) = 0\ . \quad \text{(B5)}$$

Note that the eigenfunctions $\psi_n(x)$ have an arbitrary normalization here, that is, $\int \psi_n(x)\ dx = I_n$, so that Eq. (B3) imposes that $\sum_{n=0}^{\infty} a_n I_n = 0$. Inserting into Eq. (B4) yields

$$\sum_{n=0}^{\infty} a_n(E_n - E_0)\psi_n(x) = [V(x) - E_0 + N\ \Delta E]\,\psi_0(x)\ , \quad \text{(B6)}$$

which must be valid for all x. Thus, in order to determine the coefficients a_n, the right-hand side of Eq. (B6) has to be expanded in a series of the $\psi_n(x)$. First, we can calculate the energy bias ΔE by writing the equation corresponding to $n = 0$:

$$\int |\psi_0(x)|^2\ [V(x) - E_0 + N\ \Delta E]\ dx = 0\ . \quad \text{(B7)}$$

Thus we have

$$\Delta E = \frac{1}{N}\,[E_0 - \langle V\rangle_0] = \frac{\langle T\rangle_0}{N} \quad \text{(B8)}$$

where

$$\langle V\rangle_0 = \frac{\int |\psi_0(x)|^2\ V(x)\ dx}{\int |\psi_0(x)|^2\ dx} \quad \text{(B9)}$$

stands for the expectation value of the potential energy in the unperturbed ground state, while $\langle T\rangle_0$ is the analogous expression for the kinetic energy. Note that this expression is of course equivalent to Eq. (33). The other terms ($n \neq 0$) of the expansion along with the obtained value for ΔE can be used in order to evaluate the a_n's. Then a_0 is obtained using (B3), leading to the determination of $f(x)$.

Let us check our results in the simple case of an harmonic oscillator, with a potential

$$V(x) = \frac{1}{2}m\omega^2 x^2\ . \quad \text{(B10)}$$

The normalized (with our convention) ground state wave function is given by

$$\psi_0(x) = \sqrt{\frac{\alpha^2}{2\pi}}\ e^{-\frac{\alpha^2 x^2}{2}} \qquad \text{with } \alpha^2 \equiv \frac{m\omega}{\hbar}\ . \quad \text{(B11)}$$

Our estimate for the ground state energy is then

$$E_0 = \int V(x)\psi_0(x)\ dx = \frac{\hbar\omega}{2}. \quad \text{(B12)}$$

It can be checked that the bias of the wave function can be simply written in this case as

$$f(x) = a_0\psi_0(x) + a_2\psi_2(x), \quad \text{(B13)}$$

where $\psi_2(x)$ is the normalized wave function of the second excited state, that is,

$$\psi_2(x) = \sqrt{\frac{\alpha^2}{2\pi}}\ (2\alpha^2x^2 - 1)e^{-\frac{\alpha^2x^2}{2}}\ . \quad \text{(B14)}$$

Then, Eqs. (B3) and (B4) are verified on the condition

that $a_2 = -a_0 = 1/8$, so that $f(x)$ is exactly given by

$$f(x) = \frac{1}{4}\sqrt{\frac{\alpha^2}{2\pi}}\,(\alpha^2 x^2 - 1)e^{-\frac{\alpha^2 x^2}{2}} \quad . \tag{B15}$$

Finally, the estimate for the energy bias can be calculated as

$$\Delta E = \frac{1}{N}\int dx\; V(x)[a_0\psi_0(x) + a_2\psi_2(x)]$$
$$= \frac{1}{N}\left[\frac{E_2 - E_0}{8}\right] = \frac{1}{N}\left[\frac{\hbar\omega}{4}\right] \quad . \tag{B16}$$

Using the virial theorem for the harmonic oscillator, $\langle V\rangle = \langle T\rangle = E_0/2$, and replacing it into Eq. (B8), one indeed finds this same expression for the bias.

[1] W. von der Linden, Phys. Rep. **220**, 53 (1992).
[2] M. Creutz, L. Jacobs, and C. Rebbi, Phys. Rev. Lett. **42**, 1390 (1979).
[3] F. Fucito, E. Marinari, G. Parisi, and C. Rebbi, Nucl. Phys. B **180**, 369 (1981).
[4] D. J. Scalapino and R. L. Sugar, Phys. Rev. Lett. **46**, 519 (1981).
[5] J. B. Zabolitzky and M. H. Kalos, Nucl. Phys. A **356**, 114 (1981).
[6] G. Sugiyama and S. E. Koonin, Ann. Phys. (N.Y.) **168**, 1 (1986).
[7] C. Alexandrou, J. Myczkowski, and J. W. Negele, Phys. Rev. C **39**, 1076 (1989).
[8] K. E. Schmidt and J.W. Moskowitz, J. Stat. Phys. **43**, 1027 (1986); J. W. Moskowitz and K. E. Schmidt, J. Chem. Phys. **85**, 2868 (1985).
[9] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, J. Chem. Phys. **77**, 5593 (1982).
[10] M. H. Kalos, D. Levesque, and L. Verlet, Phys. Rev. A **9**, 2178 (1974).
[11] P. A. Whitlock, D. M. Ceperley, G. V. Chester, and M. H. Kalos, Phys. Rev. B **19**, 5598 (1979).
[12] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
[13] M. H. Kalos, M. A. Lee, P. A. Whitlock, and G. V. Chester, Phys. Rev. B **24**, 115 (1981).
[14] J. H. Wilkinson, *The Algebraic Eigenvalue Problem* (Clarendon, Oxford, 1965).
[15] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods* (Methuen, London, 1964).
[16] J. H. Hetherington, Phys. Rev. A **30**, 2713 (1984).
[17] M. H. Kalos, Phys. Rev. **128**, 1791 (1962).
[18] J. B. Anderson, J. Chem. Phys. **63**, 1499 (1975).
[19] M. H. Kalos, Nucl. Phys. A **126**, 609 (1969).
[20] D. M. Ceperley and M. H. Kalos, in *Monte Carlo Methods in Statistical Physics,* edited by K. Binder (Springer-Verlag, New York, 1979).
[21] M. P. Nightingale and H. W. J. Blöte, Phys. Rev. B **33**, 659 (1986); Phys. Rev. Lett. **60**, 1562 (1988).
[22] E. M. Gelbard, Prog. Nucl. Energy **24**, 1 (1990).
[23] C. J. Umrigar, M. P. Nightingale, and K. J. Runge, J. Chem. Phys. **99**, 2865 (1993).
[24] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[25] M. Caffarel and P. Claverie, J. Chem. Phys. **88**, 1088 (1988).
[26] S. A. Chin, J. W. Negele, and S. E. Koonin, Ann. Phys. (N.Y.) **157**, 140 (1984).
[27] J. E. Hirsch, D. J. Scalapino, R. L. Sugar, and R. Blankenbecler, Phys. Rev. Lett. **47**, 1628 (1981).
[28] N. Cerf and O. Martin, Phys. Rev. C **47**, 2610 (1993).
[29] N. Cerf, Nucl. Phys. A **564**, 383 (1993).