Diffusion Monte Carlo methods with a fixed number of walkers

Roland Assaraf,[1] Michel Caffarel,[1] and Anatole Khelif[2]
[1]*CNRS, Laboratoire de Chimie Théorique, Université Pierre et Marie Curie, 4 Place Jussieu, 75252 Paris, France*
[2]*CNRS, Laboratoire de Logique Mathématique, Université Denis Diderot, 2 Place Jussieu, 75251 Paris, France*
(Received 14 June 1999)

In this paper we discuss various aspects of diffusion Monte Carlo methods using a fixed number of walkers. First, a rigorous proof of the divergence of pure diffusion Monte Carlo (PDMC) methods (DMC without branching in which the weights are carried along trajectories) is given. Second, a bias-free Monte Carlo method combining DMC and PDMC approaches, and based on a minimal stochastic reconfiguration of the population, is discussed. Finally, some illustrative calculations for a system of coupled quantum rotators are presented.

PACS number(s): 02.70.Lq, 75.40.Mg

I. INTRODUCTION

Quantum Monte Carlo (QMC) methods are powerful approaches to compute the ground-state properties of quantum systems. They have been applied with success to a great variety of problems including quantum liquids and solids, nuclear matter, spin systems, the electron gas, the electronic structure of small atoms and molecules, etc. (see, e.g., Refs. [1–4]). The basic idea of QMC is to extract from a known trial vector $|\psi_T\rangle$ its exact ground-state component $|\psi_0\rangle$. This is realized by using an operator $G(\mathcal{H})$ acting as a filter,

$$\lim_{L\to\infty} G(\mathcal{H})^L|\psi_T\rangle \sim |\psi_0\rangle, \tag{1}$$

where $\mathcal{H}$ is the Hamiltonian operator of the system. For problems defined in a continuous configuration space two forms for $G(\mathcal{H})$ are usually introduced; they define the two following types of approaches.

(i) Diffusion Monte Carlo (DMC) methods

$$G(\mathcal{H}) = e^{-\tau(\mathcal{H}-E_T)}. \tag{2}$$

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

$$G(\mathcal{H}) = \frac{1}{1+\tau(\mathcal{H}-E_T)}, \tag{3}$$

where E_T is some reference energy and τ plays the role of a time step. For lattice problems or any problem described by a Hamiltonian matrix in a finite linear space, a most natural choice is

$$G(\mathcal{H}) \equiv 1-\tau(\mathcal{H}-E_T) \tag{4}$$

and the method is usually referred to as lattice Green's function Monte Carlo. Note that the denomination ''projector Monte Carlo'' is also found in the literature to refer to any of the previous variants of the method. For simplicity we shall use here the general denomination ''diffusion Monte Carlo'' for QMC methods based on Eq. (1) and present our results for a *finite* linear space with the choice (4) for the operator $G(\mathcal{H})$. All results presented in this paper can be straightforwardly generalized to continuous models.

In a Monte Carlo scheme, successive applications of $G(\mathcal{H})$ are done using probabilistic rules. In short, it is based on the fact that the quantity

$$P^*_{i\to j}(\tau) \equiv \psi_T(j)\langle j|[1-\tau(\mathcal{H}-E_T)]|i\rangle \frac{1}{\psi_T(i)} \tag{5}$$

can be viewed as a ''generalized'' transition probability and can be used to sample stochastically the action of $G(\mathcal{H})$ on an arbitrary vector. This statement can be made more explicit by rewriting $P^*_{i\to j}(\tau)$ under the form

$$P^*_{i\to j}(\tau) \equiv P_{i\to j}(\tau) w_{ij}, \tag{6}$$

where

$$P_{i\to j}(\tau) \equiv \psi_T(j)\langle j|[1-\tau(\mathcal{H}-E_L)]|i\rangle \frac{1}{\psi_T(i)} \tag{7}$$

is now a *genuine* transition probability: $P_{i\to j}(\tau) \geqslant 0$ and $\Sigma_j P_{i\to j} = 1$ (the latter condition is not fulfilled by $P^*_{i\to j}$, except when $|\psi_T\rangle$ is the exact ground state) and where the quantity w_{ij} is defined as follows:

$$w_{ij} \equiv \frac{\langle i|[1-\tau(\mathcal{H}-E_T)]|j\rangle}{\langle i|[1-\tau(\mathcal{H}-E_L)]|j\rangle}. \tag{8}$$

In both expressions E_L is the so-called local energy which plays an important role in any QMC scheme

$$E_L(i) = \frac{\langle i|\mathcal{H}|\psi_T\rangle}{\langle i|\psi_T\rangle}. \tag{9}$$

In order to apply stochastically $G(\mathcal{H})$, two type of approaches have been considered. A first type of approaches consists in using the transition probability $P_{i\to j}$ to generate successive states and then introducing at each step the quantity w_{ij} as a weight in the averages (''to carry'' the weights). In this type of approaches the number of configurations (or ''walkers'') is constant by the very definition of the stochastic process. These methods are usually referred to as pure (no branching) diffusion Monte Carlo (PDMC) methods. In the second type of approach a birth-death (or branching) process associated with the local weight is introduced. In practice, it

consists in adding to the stochastic move defined by the transition probability, a new step in which the current configuration is destroyed or copied a number of times proportional to the local weight w_{ij}. In these methods—generically referred to as the diffusion Monte Carlo method—the number of configurations is no longer constant. Remark that in theory there is no need to go beyond the pure diffusion Monte Carlo method. In practice, this is not true since the numerical experience has shown that for extended and/or complex systems, the efficiency (computer time needed to achieve a given accuracy) is drastically reduced when configurations are let to go to regions of configuration space where the weights are small. In other words, it is important to sample less frequently regions where the total weight is small and to accumulate statistics where it is large. This is the basic reason which motivates the introduction of the branching process and justifies the widespread use of DMC compared to PDMC methods. Now, since in DMC the number of walkers can fluctuate, some sort of population control is required. Indeed, nothing prevents the total population from exploding or collapsing entirely. Various solutions to this problem have been proposed. The most employed approaches consist either in performing from time to time a random deletion/duplication step or in varying slowly the reference energy to keep the average number of walkers approximately constant. In both cases, a finite bias is introduced by the population control step. In order to minimize this undesirable source of error it is important to control the size of population as rarely as possible and in the most gentle way [1].

Very recently, following an idea introduced by Hetherington [5], Sorella and co-workers [6–8] have reconsidered the use of stochastic reconfiguration in diffusion Monte Carlo. Their motivation is to combine the best of both worlds: efficiency of DMC and absence of bias as in PDMC. Their approach is derived within a PDMC framework (the walkers ''carry'' some weight) but the population is ''reconfigured'' using specific rules. The reconfiguration is done in a such way that the number of walkers is kept constant at each step.

In this work we present a number of results regarding diffusion Monte Carlo methods and stochastic reconfiguration strategies. First, we present a rigorous proof that any PDMC method is expected to diverge as the simulation time and the number of iterations L [as defined by Eq. (1)] are let to go to infinity. This result is not surprising and has already been realized by a number of authors. However, to our knowledge no rigorous arguments have been given so far to clarify this point. In general, it is stated in a more or less detailed fashion that the variance of the product of weights w_{ij} explodes as the number of iterations is made large. Quite interestingly, the derivation of the proof of the divergence of PDMC presented here shows that this result is in fact far from being trivial. In particular, the proof of the divergence requires some care from a mathematical point of view. Second, we present a variant of the stochastic reconfiguration method which we consider to be a minimal bias-free QMC method combining efficiently PDMC and DMC ideas. This approach is built such as to minimize as much as possible the fluctuations associated with the reconfiguration step and also to recover the PDMC and DMC methods as two well-defined limits. Finally, we illustrate and compare the respective qualities and drawbacks of the different approaches on some numerical examples.

The organization of the paper is as follows. In Sec. II we give the basic ingredients of the diffusion Monte Carlo methods. Section III is devoted to the derivation of the proof of the divergence of PDMC approaches. Section IV discusses the construction of a DMC method including a minimal reconfiguration process. In Sec. V some practical calculations for a system of coupled quantum rotators are shown. Calculations are intended to illustrate the important aspects of the various DMC approaches discussed in this work. Finally, a summary of our results is presented in Sec. VI.

II. DIFFUSION MONTE CARLO METHODS

In this section we give a very brief account of the main aspects of diffusion Monte Carlo methods. This part is essentially designed to introduce formulas and notations used in the following sections. It will also enable the nonexpert to understand the major steps of DMC approaches. For more detailed presentations of the implementation of DMC to lattice (finite) systems the reader is referred to Refs. [9–12].

A. Pure diffusion Monte Carlo

As already mentioned in the Introduction the basic idea of QMC approaches is to extract from a known trial vector $|\psi_T\rangle$ its exact ground-state component $|\psi_0\rangle$. Note that such approaches are in a very close relation with power-type methods in which the ground-state eigenvector is obtained by applying a large number of times the matrix on an arbitrary initial vector. Here, the major difference is that the basic step (matrix times a vector) is no longer done exactly (the size of the linear space is too large) but in a probabilistic way using a Markov chain. Once the ground-state eigenvector has been determined, a number of properties can be obtained. As an important example, the energy is given by

$$E_0 = \lim_{L\to\infty} \frac{\langle \psi_T | \mathcal{H}[1-\tau(\mathcal{H}-E_T)]^L | \psi_T \rangle}{\langle \psi_T | [1-\tau(\mathcal{H}-E_T)]^L | \psi_T \rangle}. \tag{10}$$

Using the basic formula relating the matrix elements of the Hamiltonian and the ''generalized'' transition probability already presented in the Introduction, Eqs. (5)–(8), we easily get

$$E_0 = \lim_{L\to\infty},$$

$$\left\langle\left\langle E_L(i_L) \prod_{k=0}^{L-1} w_{i_k i_{k+1}} \right\rangle\right\rangle \Big/ \left\langle\left\langle \prod_{k=0}^{L-1} w_{i_k i_{k+1}} \right\rangle\right\rangle. \tag{11}$$

In this formula the symbol $\langle\langle \cdots \rangle\rangle$ denotes the stochastic average over all realizations of the Markov chain described by $P_{i\to j}$, Eq. (7). It is easily checked that the stationary density of the process verifying

$$\sum_i \Pi_i P_{i\to j}(\tau) = \Pi_j \tag{12}$$

is given by

$$\Pi_i = \psi_T(i)^2. \tag{13}$$

The probability of a given realization of the chain corresponding to L steps and a total time of $t = L\tau$ is

$$P[i_0 \to i_1 \to \cdots i_{L-1} \to i_L] = \Pi_{i_0} P_{i_0 \to i_1}(\tau) \cdots P_{i_{L-1} \to i_L}(\tau). \tag{14}$$

Remark that in the limit $P \to \infty$ and $\tau \to 0$ with $t = L\tau$ fixed, this probability defines a functional measure on the set of all ''trajectories'' of time-length t. View from that point, formula (11) is nothing but a generalized version of the well-known Feynman-Kac formula [13–15]. Using the ergodic (recurrent) property of the Markov chain, the sum-over-all trajectories restricted to a finite time interval can be rewritten as a sum along *one single* arbitrary infinite realization of the chain

$$E_0 = \lim_{L \to \infty} \frac{\frac{1}{L} \sum_{j=1}^{L} E_L(j) \prod_{k=0}^{j-1} w_{kk+1}}{1/L \sum_{j=1}^{L} \prod_{k=0}^{j-1} w_{kk+1}}, \tag{15}$$

where different states are denoted for simplicity as $[0,1,2,\ldots,L]$. In practice, numerical calculations are based on this formula which is particular simple to implement on a computer. Now, for later use, let us remark that the basic equation (11) can be rewritten as a simple sum-over states under the form

$$E_0 = \sum_i E_L(i) P_i / \sum_i P_i, \tag{16}$$

where the probability P_i associated with a given state i is given by

$$P_i \equiv \lim_{L \to \infty} \sum_{i_0, i_1, \ldots, i_{L-1}} P[i_0 \to i_1 \to \cdots i_{L-1} \to i_L] \prod_{k=0}^{L-1} w_{i_k i_{k+1}}, \tag{17}$$

where, for simplicity of notation, state i is identified to state i_L. By using Eqs. (7), (8), (13), and (14) it can be verified that P_i is given by

$$P_i = \psi_T(i) \psi_0(i) \tag{18}$$

up to an immaterial normalization constant. Note that when the weights are all taken to be equal to one, P_i reduces to the stationary density Π_i of the Markov chain as it should be.

B. Diffusion Monte Carlo

In the pure DMC method just described the number of configurations is kept fixed and the weights are carried out along random sequences of states. In DMC approaches the weight is introduced directly into the stochastic process via a birth-death or branching process. In practice, it consists of adding to the standard stochastic move of the PDMC method a new step in which the current configuration is destroyed or copied a number of times proportional to the local weight. Denoting m_{ij} the number of copies (multiplicity) of the state j, we take

$$m_{ij} \equiv \mathrm{int}(w_{ij} + \eta), \tag{19}$$

where $\mathrm{int}(x)$ denotes the integer part of x, and η a uniform random number on $(0,1)$. In theory, such a process is properly defined only for an infinite number of walkers. Of course, in practice only a large but finite number of walkers (a population) is considered. Adding a branching process can be viewed as sampling directly with the generalized transition probability $P^*_{i \to j}(\tau)$ defined above, Eq. (5). The fact that its normalization is not constant is responsible for the fluctuations of population. However, a stationary density for this modified process can still be defined. By writing the stationary condition

$$\sum_i P_i P^*_{i \to j}(\tau) = P_j \tag{20}$$

we see from Eq. (5) that this relation is fulfilled if E_T is chosen to be the exact energy E_0 and for the following stationary density:

$$P_i = \psi_T(i) \psi_0(i). \tag{21}$$

By using a stabilized population of configurations the exact energy may be obtained as

$$E_0 = \langle \langle E_L \rangle \rangle_w. \tag{22}$$

Note the use of an additional subscript w in the average to recall the presence of the branching process. Formally, expressions (16) and (22) for the estimate of the exact energy in PDMC and DMC methods, respectively, are identical. The same for the expressions of the probability P_i of a given state i in both approaches, Eqs. (18),(21). However, there is an essential difference which distinguishes both methods. This is the way that this probability is realized stochastically. In PDMC, the stationary density of the Markov chain is $\Pi = \Psi_T^2$ and P_i represents some effective probability obtained from averaging the weights along trajectories of infinite time-length, formula (17). In DMC, the probability P_i is realized by the stochastic process itself. There is no need to introduce additional weights in averages [see formula (22)]. As a consequence, the DMC approach is a much more stable method from a numerical point of view. The price to pay for that is the introduction of a bias resulting from the population control (done either by random deletion/duplication or smooth variation of the reference energy, see discussion in the Introduction). In contrast, with PDMC there is no need for population control. However, as we shall see in the next section, the method is intrisically unstable.

III. DIVERGENCE OF THE PDMC

In this section it is shown that the estimate of the effective probability P_i associated with a given state i as defined in PDMC, Eq. (17), does not converge to a finite deterministic value. Let us define $\lambda_n(i)(n \geqslant 1)$ the product of weights between the $(n-1)$th and nth occurences of state i in the Markov sequence:

$$\lambda_n(i) \equiv \prod_{l=N_{n-1}}^{N_n - 1} w_{i_l i_{l+1}}, \tag{23}$$

where N_k denotes the time index of the Markov chain. Let us denote $X_n(i)$ the total weight associated with all states occuring between time N_{n-1} and time N_n

$$X_n(i) \equiv \sum_{k=N_{n-1}}^{N_n} \prod_{l=N_{n-1}}^{k-1} w_{i_l i_{l+1}}. \tag{24}$$

As an important consequence of the Markov property all pairs of random variables $[X_n(i),\lambda_n(i)]$ are independent and equidistributed (same law). Only random variables X and λ corresponding to the same index n and same state i are dependent. Using the ergodic property of the chain and previous definitions, expression (17) for P_i can be rewritten as

$$P_i(n) \equiv \frac{\lambda_1(i)+\lambda_1(i)\lambda_2(i)+\cdots+\lambda_1(i)\lambda_2(i)\cdots\lambda_n(i)}{X_1(i)+\lambda_1(i)X_2(i)+\cdots+\lambda_1(i)\lambda_2(i)\cdots\lambda_{n-1}(i)X_n(i)} \tag{25}$$

when the number n of occurences of state i becomes large. Now, our problem is to determine whether or not this quantity has a well-defined limit as n goes to infinity. For reasons we shall understand later, two different cases must be distinguished. Denoting $E[\ln(\lambda)]$ the finite expectation value of the random variable λ we consider separately the two cases $|E[\ln(\lambda)]|>0$ and $E[\ln(\lambda)]=0$. Note that the time or state indices are not specified since all random variables λ are independent and of the same law.

A. $|E[\ln\lambda]|>0$

Let us first consider the case $E[\ln\lambda]>0$. After some elementary manipulations the inverse of $P_i(n)$ as expressed by Eq.(25) can be rewritten in the equivalent form (same law)

$$\frac{1}{P_i} = \frac{X_1(i)+\dfrac{X_2(i)}{\lambda_2(i)}+\dfrac{X_3(i)}{\lambda_2(i)\lambda_3(i)}+\cdots+\dfrac{X_n(i)}{\lambda_2(i)\cdots\lambda_n(i)}}{\lambda_1(i)+1+\dfrac{1}{\lambda_2(i)}+\dfrac{1}{\lambda_2(i)\lambda_3(i)}+\cdots\dfrac{1}{\lambda_2(i)\cdots\lambda_n(i)}}. \tag{26}$$

Note that, while deriving this expression, subscripts of random variables have been interchanged. Such a manipulation is allowed since random variables are independent and equidistributed. To proceed further we define the following quantities:

$$Y_n \equiv \frac{X_2(i)}{\lambda_2(i)}+\frac{X_3(i)}{\lambda_2(i)\lambda_3(i)}+\cdots+\frac{X_n(i)}{\lambda_2(i)\cdots\lambda_n(i)}$$

and

$$Z_n \equiv 1+\frac{1}{\lambda_2(i)}+\frac{1}{\lambda_2(i)\lambda_3(i)}+\cdots+\frac{1}{\lambda_2(i)\cdots\lambda_n(i)}. \tag{27}$$

We then have

$$\frac{1}{P_i} = \frac{X_1+Y_n}{\lambda_1+Z_n} \tag{28}$$

and

$$\frac{Y_n}{Z_n} = \frac{X_2+Y_{n-1}}{\lambda_2+Z_{n-1}}. \tag{29}$$

Let us now suppose that $1/P_i$ converges to a constant. Then it follows that in the limit of large n, the random variables Y_n/Z_n and $(X_1+Y_n)/(\lambda_1+Z_n)$ converge to the same constant. Now, since (Y_n,Z_n) are independent of (X_1,λ_1) [and the same for (Y_{n-1},Z_{n-1}) and (X_2,λ_2)] it follows that the random variables (X_1,λ_1) must reduce to some constants, and the same for all (X_i,λ_i). This result shows that, except in the trivial case where the weights are equal to one (no branching), P_i cannot converge to a well-defined limit as n goes to infinity. Note that similar arguments can be given in the case $E(\ln\lambda)<0$, after the transformation $\lambda_i \to 1/\lambda_i$.

Now, the important remark is that all these arguments are valid only if the random variable Z_n converges to a finite distribution. For our purposes, the convergence of Y_n has not to be considered here since the two conditions (Z_n converges and P_i finite and different from zero) implies the convergence of Y_n. In the case $E[\ln\lambda]>0$ the convergence of Z_n is a consequence of the law of large numbers. Indeed, according to this theorem, for a given realization of the Markov chain there exist two constants $C>0$ and $\beta>1$ such that

$$\lambda_1(i)\cdots\lambda_n(i) \geqslant C\beta^n \tag{30}$$

for n large enough. As a consequence Z_n converges to a positive and *finite* distibution almost surely. In the case $E[\ln\lambda]=0$ this is no longer true: Z_n tends to infinity for large n and no direct constraint on the law of random variables X_i or λ_i can be drawn. As a consequence, this case must be treated separately. Before doing that, let us emphasize that this case is in fact general. Indeed, the expectation value of $\ln\lambda$ does not depend on the particular state (as a result of the Markov property) and by multiplying all weights by a suitable constant we can always impose $E(\ln\lambda)=0$.

B. $E(\ln\lambda)=0$

To treat this case we try to depart as less as possible from the previous case. For that we introduce some new quantities $\gamma_n(i)$ which will play a role similar to that played by quantities $\lambda_n(i)$, except that by their very definition $\gamma_n(i)\geqslant C$ where C is some constant strictly greater than 1. As a direct consequence $E(\ln\gamma)>0$ and arguments similar to those employed previously will be invoked.

Let us define $\gamma_n(i)$ as the product of weights $w_{i_k i_{k+1}}$ between two occurences of state i such that the ratio of the total product of weights at the two occurences is greater than the

constant $C>1$. $\gamma_n(i)$ can be written as

$$\gamma_n(i)=\prod_{l=N_{\phi(n-1)}}^{N_{\phi(n)-1}} w_{i_l i_{l+1}}, \tag{31}$$

where $\phi(n)$ denotes the $\phi(n)$th occurence of state i verifying the condition associated with the threshold C. More precisely, $\phi(n)$ is defined as

$$\phi(n)=\inf_{(k>\phi(n-1)+1)}\left\{\prod_{l=N_{\phi(n-1)+1}}^{N_k} w_{i_l i_{l+1}}>C\right\}. \tag{32}$$

Note that the function $\phi(n)$ is well defined [successive values of $\phi(n)$ are finite] because we have

$$\sup_{1\leqslant k\leqslant n}\sum_{l=1}^{k}\ln\lambda_l(i)\rightarrow+\infty \quad \text{as } n\rightarrow+\infty \quad \text{for a given realization.} \tag{33}$$

This property is a consequence of the theorem (40) which will be presented later. Roughly speaking, what is done here is to extract from the full set of occurences of state i a subset of occurences (labeled by the function ϕ) corresponding to a series of ''stopping times'' along the random sequence. Once more, as a result of the Markov property the random variables $\gamma_n(i)$ are independent and equidistributed. In addition, from their very definition $\gamma_n(i)\geqslant C>1$. Using previous definitions we can rewrite P_i as

$$P_i(n)\equiv\frac{U_1(i)+\gamma_1(i)U_2(i)+\cdots+\gamma_1(i)\gamma_2(i)\cdots\gamma_{n-1}(i)U_n(i)}{V_1(i)+\gamma_1(i)V_2(i)+\cdots+\gamma_1(i)\gamma_2(i)\cdots\gamma_{n-1}(i)V_n(i)}, \tag{34}$$

where $U_n(i)$ represents the sum of the products of $\lambda(i)$ between the occurences $\phi(n-1)$ and $\phi(n)$

$$U_n(i)\equiv\sum_{k=N_{\phi(n-1)}}^{N_{\phi(n)}}\prod_{l=N_{\phi(n-1)}}^{k}\lambda_l(i) \tag{35}$$

and V_n the sum of all weights between the two occurences

$$V_n(i)\equiv\sum_{k=N_{\phi(n-1)}}^{N_{\phi(n)}}\left[\prod_{l=N_{\phi(n-1)}}^{k}\lambda_l(i)\right]\frac{X_k(i)}{\lambda_k(i)}. \tag{36}$$

The triplets (U_n,V_n,γ_n) are independent and equidistributed. After some elementary manipulations the inverse of P_i as expressed by Eq. (34) can be rewritten in the equivalent form (same law)

$$\frac{1}{P_i}=\frac{V_1+T_n}{U_1+W_n} \tag{37}$$

and

$$\frac{T_n}{W_n}=\frac{V_2+T_{n-1}}{U_2+W_{n-1}} \tag{38}$$

with the following definitions:

$$T_n\equiv\frac{V_2(i)}{\gamma_2(i)}+\frac{V_3(i)}{\gamma_2(i)\gamma_3(i)}+\cdots+\frac{V_n(i)}{\gamma_2(i)\cdots\gamma_n(i)}$$

and

$$W_n\equiv\frac{U_2(i)}{\gamma_2(i)}+\frac{U_3(i)}{\gamma_2(i)\gamma_3(i)}+\cdots+\frac{U_n(i)}{\gamma_2(i)\cdots\gamma_n(i)}. \tag{39}$$

In order to complete the proof we need to show that the series associated with W_n converges to some finite distribution (as already discussed in the preceding section, the derivation of the convergence of T_n is not necessary). If Z_n converges we can conclude (same arguments as before) that P_i converges to a finite (deterministic) value only if the ratio V_n/U_n is a constant, which is not the case. To do that let us first introduce the following theorem.

Theorem. Let X_l be a family of independent, equidistributed and centered (zero mean) random variables. If all moments of the random variables are finite we have

$$\frac{\sup_{1\leqslant k\leqslant n}\sum_{l=1}^{k}X_l}{n^{\alpha}}\rightarrow+\infty \quad \text{as } n\rightarrow+\infty \quad \forall\alpha<\frac{1}{2}. \tag{40}$$

This theorem is a consequence of the central-limit theorem. Using rough arguments we can say that the sum of the independent variables in the numerator converges to some Gaussian distribution with a variance proportional to n and that the greatest value is expected to behave as the square root of the variance $\sqrt{n}$. As a consequence, the ratio of the numerator and denominator must diverge as soon as $\alpha>1/2$. Although these arguments are correct, a rigorous derivation is actually not so simple. It requires some mathematical care which is beyond the scope of this work. The derivation will be presented elsewhere [20]. Now, the important point is that the random variables $\ln\lambda_l(i)$ verify the hypotheses of the theorem. They are independent, equidistributed (with zero mean), and as a consequence of the finite variation of the weights and ''stopping times'' [as defined by the function $\phi(n)$] all their moments are finite. Using the fact that $\phi(n)$ is a series extracted from the full series of occurences of state i we obtain the following property:

$$\frac{\sup_{1\leqslant k\leqslant\phi(n)}\left[\sum_{l=1}^{k}\ln\lambda_l(i)\right]}{\phi(n)^{\alpha}}\rightarrow+\infty \quad \text{as } n\rightarrow+\infty \quad \forall\alpha<\frac{1}{2} \tag{41}$$

or, equivalently,

$$\frac{\sum_{l=1}^{n} \ln \gamma_l(i)}{\phi(n)^{\alpha}} \to +\infty \quad \text{as} \quad n \to +\infty \quad \forall \quad \alpha < \frac{1}{2}. \tag{42}$$

Note that, in the particular case $\alpha = 0$, we recover the property (33), a result which guarantees that the function $\phi(n)$ is well defined. From its definition (35) the series $U_n(i)$ is a sum of at most (but not equal) $\phi(n)$ terms all smaller than the constant C except the last one which is $\gamma_n(i)$. Therefore, we have

$$0 \leqslant U_n(i) \leqslant C\phi(n) + \gamma_n(i). \tag{43}$$

From this relation we can write

$$0 \leqslant \frac{U_n(i)}{\gamma_2(i)\cdots\gamma_n(i)} \leqslant \frac{C\phi(n)}{\gamma_2(i)\cdots\gamma_n(i)} + \frac{1}{\gamma_2(i)\cdots\gamma_n(i)}. \tag{44}$$

The series of general term $1/\gamma_2(i)\cdots\gamma_n(i)$ is convergent since all $\gamma_n(i)$ are greater or equal to $C > 1$. Regarding the other term $A_n \equiv C\phi(n)/\gamma_2(i)\cdots\gamma_n(i)$ we can write as a result of Eq. (42) that there exists a constant $M(\alpha)$ such that

$$\phi(n) \leqslant M\left[\sum_{l=1}^{n} \ln \gamma_l(i)\right]^{1/\alpha} \quad \text{for} \quad 0 < \alpha < \frac{1}{2}. \tag{45}$$

Therefore, we have

$$\frac{C\phi(n)}{\gamma_2(i)\cdots\gamma_n(i)} \leqslant CM\gamma_1(i)\frac{\left[\ln \prod_{l=1}^{n} \gamma_l(i)\right]^{1/\alpha}}{\prod_{l=1}^{n} \gamma_l(i)}. \tag{46}$$

From the fact that the function $[\ln x]^{1/\alpha}/x$ decreases for x large enough and that $\gamma_l > C$ it follows that A_n is bounded from above by $CM\gamma_1(i)[n \ln C]^{1/\alpha}/C^n$, the general term of a convergent series. Finally, we can conclude that the series W_n converges to some finite positive distribution. This result completes our proof of the nonconvergence of the PDMC estimate of the effective probability P_i.

IV. DMC WITH MINIMAL STOCHASTIC RECONFIGURATION

As seen in the previous section PDMC is intrinsically unstable. As already remarked, the basic reason for that is the increase of variance of the products of weights as a function of the number of iterations (or projecting time). However, as illustrated by a number of applications performed using this type of approaches (e.g., Refs. [15,11,16–19]) the method has proven to be very useful. This is the case when the trial wave function is accurate enough to allow the convergence of the various averages before large fluctuations associated with large projecting times arise. When convergence is achieved no finite bias due to a population control process is introduced. In order to make PDMC approaches effective the fluctuations of the weights must be decreased in some way. Before considering this point, let us determine the dependence of the error as a function of the computational effort in a PDMC scheme. The fluctuations of the weight from iteration n to iteration $n+1$ will be described by the variance β^2 defined as

$$\beta^2 = \frac{\left\langle \left(\frac{w^{(n+1)}}{w^{(n)}} \right)^2 \right\rangle}{\left\langle \frac{w^{(n+1)}}{w^{(n)}} \right\rangle^2}. \tag{47}$$

By definition β is greater or equal to one. The equality is obtained in the optimal case corresponding to constant weights (no branching). Let N be the total number of Monte Carlo steps of the simulation (the computational effort is proportional to N). The systematic error due to a finite projecting time T (number of iterations $L = T/\tau$) is of order $\exp(-T\Delta)$ where Δ is the gap in energy of the model ($\Delta = E_1 - E_0$, where E_0 is the ground-state energy and E_1 the energy of the lowest state having a nonzero overlap with the trial wave function). The statistical error due to the finite statistics on some quantity evaluated at some fixed projecting time T is given by $\beta^T/\sqrt{N/T}$. By equating both errors an estimate of the relation between the computational effort (via N) and a given accuracy ϵ can be obtained. In the large-N limit the relation is easily found to be

$$\epsilon \sim \frac{1}{N^{\gamma/2}}$$

with

$$\gamma = \frac{\Delta}{\ln \beta + \Delta}. \tag{48}$$

When $\beta = 1$ (no fluctuations of the weights) $\gamma = 1$ and the efficiency of the simulation is optimal: the standard $1/\sqrt{N}$ law of diffusion processes is recovered. As β is increased the efficiency of the simulations can decrease quite rapidly. Accordingly, to enhance the efficiency of PDMC the fluctuations of the weights must be decreased. An elegant solution to this problem has been introduced more than ten years ago by Hetherington [5]. The idea consists in carrying many walkers simultaneously and introducing a global weight associated with the entire population instead of a local weight for each walker. The global weight W is chosen to be the average of the local weights w_{i_l}

$$W_{i_1 \cdots i_M} \equiv \frac{1}{M} \sum_{l=1}^{M} w_{i_l}, \tag{49}$$

where M is the number of walkers considered (to avoid confusion between various indices only one subscript has been used for individual weights). By increasing the number M of walkers the fluctuations of the global weight is reduced and β as defined in Eq. (47) is decreased. It is easy to check that the quantity $\ln \beta$ decreases as the inverse of the number of walkers. As a consequence of Eq. (48) the gain in computational efficiency can be very important. Now, the method consists in defining a PDMC scheme in the enlarged linear

space defined by the tensorial product (M times) of the initial linear space. In this new space the full transition probability is defined as the tensorial product of individual transition probabilities. Note that no correlation between the stochastic moves of different walkers is introduced at this level. Second, and this is the important point, each individual weight carried by a walker is rewritten as a function of the global weight

$$w_{i_l} = \tilde{w}_{i_l}(i_1 \cdots i_M) W_{i_1 \cdots i_M}$$

with

$$\tilde{w}_{i_l}(i_1 \cdots i_M) = \frac{w_{i_l}}{W_{i_1 \cdots i_M}}. \tag{50}$$

This rewriting allows us to introduce the global weight as a weight common to all walkers and thus to define a standard PDMC scheme in the tensorial product of spaces. To take into account the new weight $\tilde{w}_{i_l}$ a so-called reconfiguration process is introduced. At each step the total population of M walkers is ''reconfigured'' by selecting with probability proportional to $\tilde{w}$ the same number M of walkers. Note that, at this point, some correlation between different walkers is introduced. Now, let us discuss the two important limits of the algorithm, namely, the case of an infinite number of walkers, $M \to \infty$ and the case of constant weights, $w_i \to 1$. When $M \to \infty$ the global weight converges to its stationary exact value. As a consequence, the different weights $\tilde{w}$ associated with each walker [as given by Eq. (50)] become independent from each other and the reconfiguration process reduces to the usual branching process (19) without population control and systematic bias since the population is infinite. In the limit $w_i \to 1$ the method does not reduce to the standard PDMC approach. Indeed, the reconfiguration step ''reconfigures'' the entire population whatever the values of the weights. In order to improve the efficiency of such methods this undesirable source of fluctuations must be reduced and the limit of the exact PDMC should be implemented in the method. For that we divide the population of walkers into two different sets. A first set of walkers corresponds to all walkers verifying $\tilde{w} \geqslant 1$. These walkers can be potentially duplicated and will be called ''positive'' walkers. The other walkers verify $0 \leqslant \tilde{w} < 1$, they can be potentially destroyed and will be called ''negative walkers.'' The number of reconfigurations is defined as

$$N_{\mathrm{Reconf}} = \sum_{i+} |\tilde{w}_i - 1| = \sum_{i-} |\tilde{w}_i - 1|, \tag{51}$$

where Σ_{i+} (Σ_{i-}) indicates that the summation is done over the set of positive (negative) walkers. The equality in Eq. (51) is a simple consequence of the definition of positive and negative walkers. In practice, an integer number of reconfigurations is obtained by considering $\mathrm{int}(N_{\mathrm{reconf}} + \eta)$, where η is a uniform random number on the interval (0,1). Once the number of reconfigurations has been drawn, N_{reconf} walkers are added to or removed from the current population by drawing separately N_{reconf} walkers among the lists of positive and negative walkers. It is easily verified that by doing this no source of systematic error has been introduced and that it is equivalent to the original reconfiguration process of Hetherington. However, in contrast with the latter the average number of reconfigurations is kept minimal and, consequently, the efficiency of the simulation is significantly enhanced. In addition, the average number of reconfigurations vanishes as the weights become constant. In other words, the reconfiguration method reduces in this limit to the standard PDMC method. In their recent work Calandra-Buonaura and Sorella [7] (CBS) have proposed to use a reconfiguration process which is essentially identical to that of Hetherington, except that the reconfiguration step is not necessarily done at each iteration. Besides reducing the finite bias on the stationary density they have shown that their approach allows to calculate efficiently ground-state correlation functions within a forward walking approach. Here, our reconfiguration process is built in order to minimize as much as possible the fluctuations of the weights at each step. As a consequence, the finite bias on the stationary density is also reduced as much as possible. In particular, and in contrast with the CBS scheme, our algorithm is found to be optimal when the reconfiguration process is applied at each iteration.

V. AN ILLUSTRATIVE EXAMPLE

In this section we present some calculations illustrating the various aspects of DMC approaches discussed in the preceding sections. The system considered is a chain of N_s coupled quantum rotators (one per site). In the angular representation the Hamiltonian is written

$$H \equiv -\sum_{i=1}^{N_s} \frac{\partial^2}{\partial \theta_i^2} - \frac{x}{2} \sum_{i=1}^{N_s} \cos(\theta_{i+1} - \theta_i), \tag{52}$$

where $(\theta_1 \cdots \theta_{N_s})$ are angular variables $\theta_i \in \mathbb{R}/2\pi\mathbb{Z}$ and periodic boundary conditions are used ($\theta_{N_s+1} = \theta_1$). In this formula x is a parameter defining the relative weight of the potential and kinetic terms. It can be shown that the model described by this Hamiltonian has the same critical properties as the two-dimensional XY spin model [21]. The finite-temperature Kosterlitz-Thouless (KT) classical phase transition of the spin model is equivalent to a zero-temperature quantum phase transition in the rotator model occuring at some critical value for the parameter x. Monte Carlo simulations have been done in the angular momentum representation. In this representation H is expressed in the discrete basis, $|l_1 \cdots l_{N_s}\rangle$ ($l_i \in \mathbb{Z}$), consisting of the eigenvectors of the angular momentum operators at different sites. We have

$$H = \sum_{i=1}^{N_s} \hat{J}_i^2 - \frac{x}{2} \sum_{i=1}^{N_s} (\phi_{i+1}^+ \phi_i + \mathrm{H.c.}), \tag{53}$$

where the operators $(\phi_i^+, \phi_i, \hat{J}_i)$ are defined as follows [Lie-algebra of $O(2)$]:

$$\begin{aligned} \phi_i^+ |l_1 \cdots l_i \cdots\rangle &= |l_1 \cdots l_i + 1 \cdots\rangle, \\ \phi_i |l_1 \cdots l_i \cdots\rangle &= |l_1 \cdots l_i - 1 \cdots\rangle, \\ \hat{J}_i |l_1 \cdots l_i \cdots\rangle &= l_i |l_1 \cdots l_i \cdots\rangle. \end{aligned} \tag{54}$$

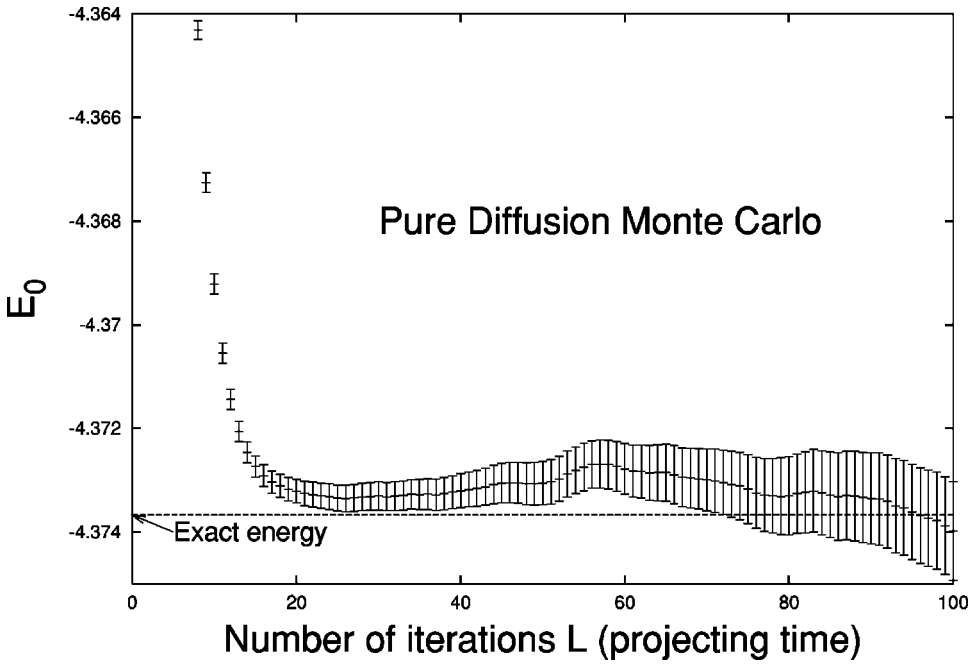

FIG. 1. PDMC calculation of the energy as a function of the projecting time. E_0 is in units of $\hat{J}^2$, Eqs. (53),(54).

Here we are interested in calculating the ground-state energy of the model. Note that the ground state belongs to the fundamental representation of $O(2)$ corresponding to a total momentum equal to zero $\Sigma_{i=1}^{N_s} l_i = 0$. In what follows the parameter x is taken to be 1.8, a value expected to be very close to the exact critical value (see, Hamer *et al.* [21]). In actual calculations we have taken $N_s = 6$. By using exact diagonalization methods (Lanczòs algorithm) and after extrapolation to an infinite basis set ($l_i \to \infty$) we get $E_0 = -4.37367626$ (all digits converged) for parameters ($x = 1.8$, $N_s = 6$).

In Fig. 1 a PDMC calculation of the exact energy is presented. The trial wave function used is given by

$$\psi_T = e^{-k \sum_{i=1}^{N_s} l_i^2}, \tag{55}$$

where k is some positive parameter.

The unstable character of PDMC at large times is clearly illustrated. At zero-projecting time the variational energy associated with the trial wave function is recovered with small fluctuations, $E_v = -4.10284(25)$. The fact that this value is quite different from the exact one illustrates the poor quality of the trial wave function. Now, when the projecting time is increased the estimate of the energy converges to the exact value (number of iterations of about 25). For larger times the estimate of the energy begins to wander and no stabilization is observed.

In Fig. 2 we present some DMC calculations performed by using the standard branching process associated with w_{ij} [Eq. (19)] and a population control step to keep the number of walkers under control. The population control has been done by adjusting the reference energy to the fluctuations of population by using a formula of the type

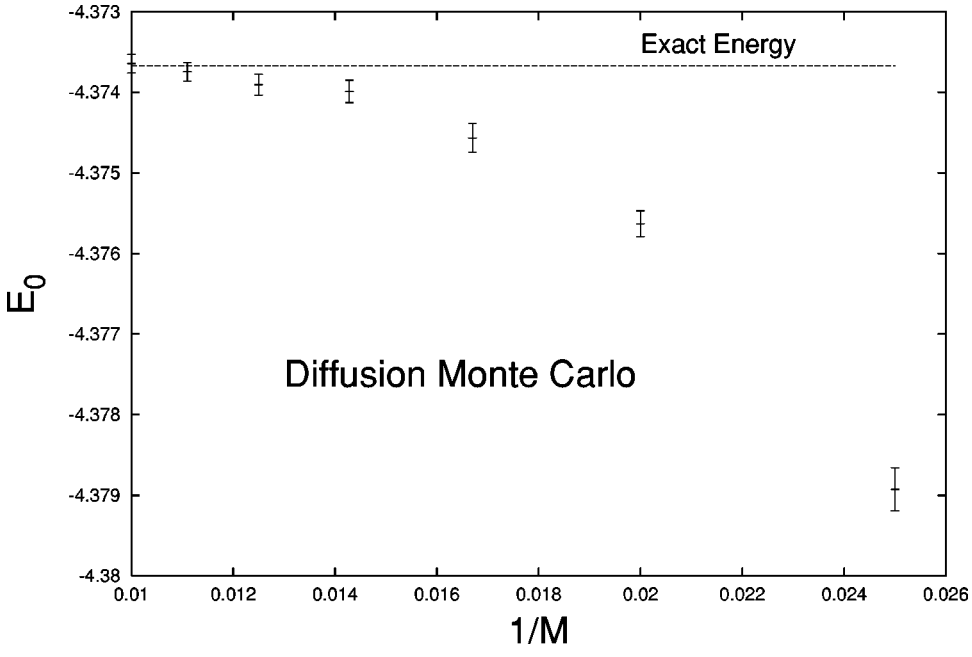

FIG. 2. DMC calculation of the exact energy as a function of the size of the population. E_0 is in units of $\hat{J}^2$, Eqs. (53),(54).

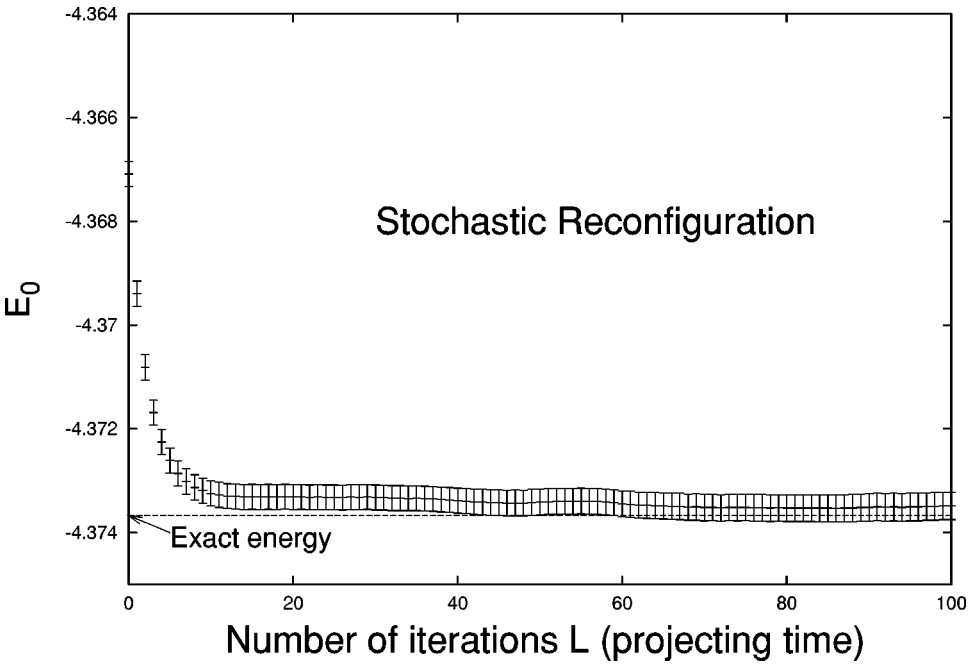

FIG. 3. PDMC with stochastic reconfiguration method. E_0 is in units of $\hat{J}^2$, Eqs. (53),(54).

$$E_T(t+\tau)=E_T(t)+K/\tau \ln[M(t)/M(t+\tau)], \qquad (56)$$

where $M(t)$ is the total number of walkers at time t and K is some positive constant. When $M(t+\tau)>M(t)$ the reference energy is reduced and more walkers are killed at the next step. In the opposite case E_T is raised and more walkers are duplicated.

Calculations have been done with populations of different sizes ranging from $M=40$ to $M=100$. At $M=40$ the bias is small (systematic error of about 1/1000) but much greater than the statistical error. The error is seen to decrease as the size of the population is increased. For $M=100$ it is smaller than the statistical error. It should be emphasized that the magnitude of the systematic error is very dependent on the quality of the trial wave function. Here a quite simple trial wave function has been used. With more sophisticated and fully optimized forms the error would be much smaller.

In Fig. 3 we present a PDMC calculation with the original reconfiguration process of Hetherington. The number of walkers used is $M=50$. When compared to the PDMC calculation of Fig. 1 (same range for the projecting time) the stabilization in time resulting from the use of the global weight and the reconfiguration process is clearly seen. A chaotic behavior similar to that observed in Fig. 1 at large times is also expected but for much larger projecting times.

In the next figure, Fig. 4, we present our improved version for the stochastic reconfiguration process. The convergence as a function of time of the energy is very satisfactory and the fluctuations are reduced. Note that the value of the energy at the origin (no projection) $E_0=-4.37115(16)$ is

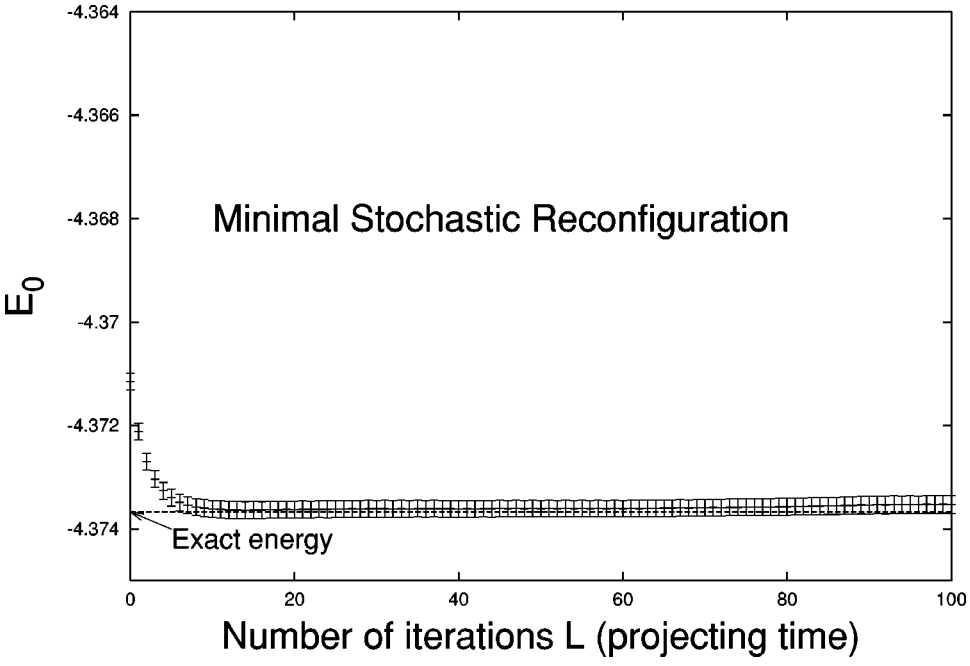

FIG. 4. PDMC with minimal stochastic reconfiguration. E_0 is in units of $\hat{J}^2$, Eqs. (53),(54).

much closer to the exact result ($E_0 = -4.37367\cdots$) than in the standard case (Fig. 3) $E_0 = -4.36708(24)$. This result is a direct consequence of the fact that the average number of reconfigurations with our minimal scheme is much smaller than in the previous case. In other words, the effective number of walkers has been increased and, then, the finite bias on the stationary density has been reduced. Note that in the limit of an infinite number of walkers the finite error on the energy would entirely disappear.

VI. SUMMARY

In this paper we have discussed various aspects of diffusion Monte Carlo methods at fixed number of walkers. First, we have concentrated our attention on the so-called pure diffusion Monte Carlo (PDMC) methods in which no branching process is introduced (the weights are carried) and for which the number of configurations is kept fixed at any level of the algorithm. As already remarked by a number of authors, PDMC methods are powerful, but they suffer from a severe problem at large projecting times (rapid increase of the variance). In this paper this statement has been made much more precise by showing that the statistical estimate of the effective probability associated with a given state as calculated in a PDMC scheme does not converge to a finite deterministic value. This is in sharp contrast with what happens in DMC where a different —but biased— estimate is used for the same quantity. Quite interestingly, the derivation of the proof turns out to be far from being trivial. In particular, it was necessary to deal in detail with a difficult case [$E(\ln\lambda)=0$ with our notations] from which a convergent variant of PDMC could have emerged. Second, based on an original estimate of the PDMC error [formula (48)] we have discussed the most natural generalization of PDMC which can make the method effective for problems associated with large fluctuations of the weights. By introducing stochastic reconfiguration processes as proposed by Hetherington and very recently reconsidered by Sorella and co-workers we have proposed an alternative approach to realize what can be called a minimal stochastic reconfiguration DMC approach. The method has been designed to reduce as much as possible the statistical fluctuations associated with the reconfiguration process and also to recover both PDMC and DMC limits. The numerical calculations presented have illustrated the validity of such an approach.

[1] D.M. Ceperley and M.H. Kalos, in *Monte Carlo Method in Statistical Physics*, edited by K. Binder (Springer-Verlag, Heidelberg, 1992).

[2] D.M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995).

[3] W. von der Linden, Phys. Rep. **220**, 53 (1992).

[4] B.L. Hammond, W.A. Lester, Jr., and P.J. Reynolds, in *Monte Carlo Methods in Ab Initio Quantum Chemistry*, Vol. 1 of *World Scientific Lecture and Course Notes in Chemistry* (World Scientific, Singapore, 1994).

[5] J.H. Hetherington, Phys. Rev. A **30**, 2713 (1984).

[6] S. Sorella, Phys. Rev. Lett. **80**, 4558 (1998).

[7] M. Calandra Buonaura and S. Sorella, Phys. Rev. B **57**, 11 446 (1998).

[8] S. Sorella and L. Capriotti, Phys. Rev. B **61**, 2599 (2000).

[9] J. Carlson, Phys. Rev. B **40**, 846 (1989).

[10] N. Trivedi and D.M. Ceperley, Phys. Rev. B **41**, 4552 (1990).

[11] K.J. Runge, Phys. Rev. B **45**, 7229 (1992).

[12] R. Assaraf, P. Azaria, M. Caffarel, and P. Lecheminant, Phys. Rev. B **60**, 2299 (1999).

[13] J. Glimm and A. Jaffé, *Quantum Physics. A Functional Integral Point of View* (Springer, New York, 1981).

[14] M. Caffarel and P. Claverie, J. Chem. Phys. **88**, 1088 (1988).

[15] M. Caffarel and P. Claverie J. Chem. Phys. **88**, 1100 (1988).

[16] D.M. Ceperley and B. Bernu, J. Chem. Phys. **89**, 6316 (1988).

[17] B. Bernu, D.M. Ceperley, and W.A. Lester, Jr., J. Chem. Phys. **93**, 552 (1990).

[18] M. Caffarel, M. Rérat, and C. Pouchan, Phys. Rev. A **47**, 3704 (1993).

[19] F. Schautz and H.-J. Flad, J. Chem. Phys. **110**, 11 700 (1999).

[20] A. Khelif, R. Assaraf, and M. Caffarel (unpublished).

[21] C.J. Hamer, J.B. Kogut, and L. Susskind, Phys. Rev. D **19**, 3091 (1979).