Zero-variance zero-bias principle for observables in quantum Monte Carlo: Application to forces

Roland Assaraf[a)] and Michel Caffarel[b)]
CNRS-Laboratoire de Chimie Théorique Tour 22-23, Case 137, Université Pierre et Marie Curie,
4 place Jussieu, 75252 Paris Cedex 05, France

A simple and stable method for computing accurate expectation values of observables with variational Monte Carlo (VMC) or diffusion Monte Carlo (DMC) algorithms is presented. The basic idea consists in replacing the usual "bare" estimator associated with the observable by an improved or "renormalized" estimator. Using this estimator more accurate averages are obtained: Not only the statistical fluctuations are reduced but also the systematic error (bias) associated with the approximate VMC or (fixed-node) DMC probability densities. It is shown that improved estimators obey a zero-variance zero-bias property similar to the usual zero-variance zero-bias property of the energy with the local energy as improved estimator. Using this property improved estimators can be optimized and the resulting accuracy on expectation values may reach the remarkable accuracy obtained for total energies. As an important example, we present the application of our formalism to the computation of forces in molecular systems. Calculations of the entire force curve of the H_2, LiH, and Li_2 molecules are presented. Spectroscopic constants R_e (equilibrium distance) and ω_e (harmonic frequency) are also computed. The equilibrium distances are obtained with a relative error smaller than 1%, while the harmonic frequencies are computed with an error of about 10%. © *2003 American Institute of Physics.* [DOI: 10.1063/1.1621615]

I. INTRODUCTION

Over the recent years quantum Monte Carlo (QMC) methods have become more and more successful in computing ground-state total energies of molecular systems. For systems with large number of electrons the accuracy obtained by QMC is very good. As illustrated by a number of recent calculations,[1–14] the quality of the results is comparable and, in most cases, superior to that obtained with more traditional techniques (density functional theory, multiconfiguration self-consistent field, or coupled cluster methods). Unfortunately, for properties other than energy the situation is much less favorable and accurate results are difficult to obtain. To understand this point let us first define what we mean here by accuracy. In standard quantum Monte Carlo schemes there exist essentially two types of error:

(i)   The usual statistical error resulting from the necessarily finite simulation time. This error present in any Monte Carlo scheme behaves as $\sim 1/\sqrt{N}$ where N is the number of Monte Carlo steps.

(ii) The systematic error (or "bias") associated with some particular choice of the trial wave function. In a variational Monte Carlo (VMC) scheme it is the systematic error resulting from the approximate trial probability density. In a fixed-node diffusion Monte Carlo (DMC) it is either the fixed-node error of energy calculations or the systematic error associated with the mixed DMC probability density for a more general observ-

able. Other types of systematic errors may also exist, e.g., the short-time error,[15] however, such errors can be easily controlled and, therefore, will not be considered here.

Now, to enlighten the major differences between energy and observable computations let us evaluate the expressions of these two errors. We shall do that within the framework of the variational Monte Carlo method where, as we shall see later, all the main aspects of this work are already present.

In a variational Monte Carlo simulation the variational energy

$$E_v \equiv \frac{\langle \psi_T | H | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle}, \tag{1}$$

where ψ_T is the approximate trial wave function used, is re-expressed as the statistical average of the local energy defined as

$$E_L = \frac{H \psi_T}{\psi_T} \tag{2}$$

over the probability density associated with ψ_T^2, namely

$$E_v = \langle E_L \rangle_{\psi_T^2}. \tag{3}$$

An accurate calculation of the energy requires the two following conditions.

(i) The systematic (or variational) error defined as

$$\Delta_E \equiv E_v - E_0 \geqslant 0, \tag{4}$$

where E_0 is the exact energy, must be as small as possible.

[a)]Electronic mail: ra@lct.jussieu.fr
[b)]Electronic mail: mc@lct.jussieu.fr

10536

J. Chem. Phys., Vol. 119, No. 20, 22 November 2003

Observables in QMC   10537

(ii) The variance of the local energy (which is directly related to the magnitude of the statistical error)

$$\sigma^2(E_L) = \langle (E_L - E_v)^2 \rangle_{\psi_T^2}, \tag{5}$$

must also be as small as possible.

To estimate both quantities we express them in terms of the trial wave function error, $\delta\psi = \psi_T - \psi_0$, where ψ_0 is the exact wave function. Regarding the systematic error it is easy to check that

$$\Delta_E = \frac{\langle \psi_T - \psi_0 | H - E_0 | \psi_T - \psi_0 \rangle}{\langle \psi_T | \psi_T \rangle}. \tag{6}$$

In other words, Δ_E is of order two in the wave function error

$$\Delta_E \sim O[\delta\psi^2]. \tag{7}$$

Now, regarding the variance, it is convenient to write the following equality:

$$E_L - E_v = \frac{(H - E_0)(\psi_T - \psi_0)}{\psi_T} - \Delta_E, \tag{8}$$

from which it is directly seen that $\sigma^2(E_L)$ is also of order two,

$$\sigma^2(E_L) \sim O[\delta\psi^2]. \tag{9}$$

Equations (7) and (9) are at the origin of the high-quality calculations of the energy. They show that accurate energy calculations are directly related to good trial wave functions: The more accurate the trial wave function is, the smaller the statistical and systematic errors are. In the limit of an exact trial wave function, both errors vanish and the energy estimator reduces to the exact energy. This most fundamental property is referred to in the literature as the "zero-variance property." Note that a much more preferable and accurate denomination should be "zero-variance-zero-bias property" to emphasize the existence of the *two* types of error. Of course, in the case of the energy this distinction is not necessary since, as just seen, the two errors are not independent and vanish *simultaneously* with the exact wave function. However, as we shall see in the following, this peculiar aspect will be no longer true for other properties.

Let us now turn our attention to the computation of a general observable. Defining the expectation value of some arbitrary observable O as

$$O_v \equiv \frac{\langle \psi_T | O | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle}, \tag{10}$$

its Monte Carlo expression is given by

$$O_v = \langle O \rangle_{\psi_T^2}. \tag{11}$$

It is easy to verify that the systematic error behaves as

$$\Delta_O \equiv \langle O \rangle_{\psi_T^2} - \langle O \rangle_{\psi_0^2} \sim O[\delta\psi], \tag{12}$$

while the variance is given by

$$\sigma^2(O) \sim O[1]. \tag{13}$$

Compared to the energy case we have two striking differences. First, the systematic error in the averages is much larger. This is a direct consequence of Eq. (12): the estimator of a general observable has only a *linear* zero-bias property instead of a quadratic one like in the energy case. Even worse, because trial wave functions are optimized to lower the systematic error in the energy (and/or its fluctuations) and not the error in the observable, the prefactor associated with the linear error contribution, Eq. (12), is usually much larger than in the energy case, Eq. (7). In practice, this important systematic error makes in general the quality of the expectation value, Eq. (11), very poor. The second important difference is that there is no zero-variance property at all for observables when Eq. (11) is used. Indeed, even when the exact wave function is used as trial wave function we are still left with some finite (and eventually large) statistical fluctuations, Eq. (13). Thus, statistical fluctuations are in general very large for properties. A simple and popular strategy to reduce the important systematic error on properties is to mix VMC and fixed-node DMC calculations to build up a so-called "hybrid" or "second-order" estimator, $\langle O \rangle_{\text{hybrid}} \equiv 2\langle O \rangle_{\text{DMC}} - \langle O \rangle_{\text{VMC}}$, whose error is reduced.[16] An elementary calculation shows that the error is now of order $O[(\psi_T - \psi_0)^2]$, plus a linear contribution $O(\psi_0^{FN} - \psi_0)$ due to the approximate nodes of the trial wave function. However, once again such a solution is not, in practice, as satisfactory as it appears at first glance because of the large prefactor associated with the second-order contribution and, also, because of the non-negligible linear error due to the nodes. A second possible strategy to cope with the systematic error is to perform an "exact" QMC calculation based on one of the variants of the so-called "forward walking" scheme.[17–19] Unfortunately, although such schemes can lead to satisfactory results for small systems,[20–22] they are known to be intrinsically unstable and, therefore, very time consuming for large systems. In practice, the possibility of getting or not getting a satisfactory answer depends very much on the accuracy required and on the type of observable considered. Therefore, forward walking is not considered as a general practical solution to the problem.

In this work, we propose to follow a quite different route. Our purpose is to show that it is possible to use much more efficient estimators for properties than the usual bare expression, Eq. (11). More precisely, it is shown how to construct in a simple and systematic way new estimators having the same remarkable quadratic zero-variance zero-bias property as the energy case. Very recently, we have made a first step in that direction by showing how to generalize the zero-variance part of this property.[23,24] In short, the basic idea consists of constructing a "renormalized" or improved observable having the same average as the original one but a lower variance. To build the renormalized observable, an auxiliary wave function is introduced. This function plays a role analogous to the one played by the trial wave function in the case of the energy: The closer the auxiliary function is to the exact solution of some zero-variance equation (the Schrödinger equation in the case of the energy), the smaller the statistical fluctuations of the renormalized observable are. Our approach has been illustrated on some simple academic examples[23] and also for the much more difficult case of the computation of forces for some diatomic molecules.[24] Numerical results on these examples are very satisfactory. When

suitably chosen auxiliary functions are used, statistical errors are indeed greatly reduced.

Here, we present the full generalization of the preceding idea: it is shown how to construct improved observables minimizing *both* systematic and statistical errors with a quadratic behavior similar to that obtained for the energy. As a consequence, any observable is expected to be calculated, at least in principle, with the remarkable accuracy achieved by QMC for total energies. The basic idea behind our approach is quite simple: it consists of making use of the relation between energy and observable calculations as expressed by the Hellmann–Feynman (HF) theorem. As is well known this theorem expresses any quantum average as a total energy derivative with respect to the magnitude of the external potential defined by the observable. It is shown how the zero-variance zero-bias principle valid for each value of the energy (as a function of the external potential) can be extended to the derivative and, therefore, to the observable. Note that in the context of QMC simulations, the idea of using the HF theorem to compute observables, using either a finite difference scheme or the analytic derivative, is not new and has been applied by several groups.[25–34] In general, the results are good for very small systems but rapidly disappointing for larger systems. Indeed, only when a clear physical insight into the origin of the fluctuations of the infinitesimal difference of energy (the derivative) is available is it possible to propose an efficient solution to the problem. A very nice example of such a possibility has been presented some time ago by Umrigar at the VMC level[28] and, more recently, by Filippi and Umrigar for the more general DMC case.[33] By using a finite representation of the energy derivative and by introducing a special coordinate transformation allowing the electrons close to a given nucleus to move almost rigidly with that nucleus, Umrigar[28] has shown how to correlate efficiently the calculation of the electronic energies associated with two slightly different nuclear configurations of a diatomic molecule. As a result it is possible to get accurate estimates of the energy derivatives (forces) for some diatomic molecules.[28,33] Here, we show how this correlated sampling method can be re-expressed in our framework. In addition, by generalizing this idea it is shown how general coordinate transformations can be used to define a new class of improved estimators.

While finishing this work, we became aware of a paper just published by Casalegno, Mella, and Rappe.[35] The idea underlying their work has some close relations to what is presented here. In short, they propose, as we do here, to compute forces using a Hellmann–Feynman-type formalism. Their expression to calculate forces is obtained by making the derivative of the VMC (or DMC) energy average with respect to nuclear positions. To reduce the systematic error these authors propose to employ trial wave functions which have been very carefully optimized via energy minimization (let us recall that the HF theorem is valid when *fully* optimized wave functions are used). To decrease the very large statistical fluctuations associated with the infinite variance, the improved estimator introduced in our previous work on forces[24] is used. As we shall see in the following, the approach proposed by Casalegno and collaborators can be viewed as a special case of the general method presented here, except that their estimator does not obey a zero-variance zero-bias property. This latter aspect has some practical consequences since, as we shall see in the following, having such a property at our disposal allows more flexibility to reduce both systematic and statistical errors.

The organization of the paper is as follows: In Sec. II we present the Hellmann–Feynman theorem and the construction of improved estimators for variational Monte Carlo calculations. It is also shown how the idea of Umrigar[28] consisting of introducing a special coordinate transformation can be used to build up some more general and more efficient improved estimators. In Sec. III we discuss the generalization of the formulas to the case of diffusion Monte Carlo calculations. In Sec. IV we present the application of the formalism to the computation of the entire force curve for the H_2, LiH, and Li_2 molecules. Calculations of the spectroscopic constants, R_e and ω_e, are also reported. Finally, in the last section we summarize our results and present some concluding remarks.

II. IMPROVED ESTIMATORS FOR OBSERVABLES

In order to make the connection between energy and observable computations we shall make use of the HF theorem, which expresses the expectation value of an observable as an energy derivative,

$$\frac{\langle \psi_0 | O | \psi_0 \rangle}{\langle \psi_0 | \psi_0 \rangle} = \left. \frac{dE_0(\lambda)}{d\lambda} \right|_{\lambda=0}, \tag{14}$$

where $E_0(\lambda)$ is the exact ground-state energy of the "perturbed" Hamiltonian defined as

$$H(\lambda) \equiv H + \lambda O. \tag{15}$$

By choosing various approximate expressions for the exact energy in Eq. (14), it is possible to derive various approximate estimates for the average. In the next sections we present two choices which turn out to be particularly efficient in practical applications.

A. Improved estimator built from the variational approximation of the energy

A most natural choice consists of replacing the exact energy of the HF theorem by a high-quality variational approximation. To do that, we introduce some λ-dependent approximate trial wave function, $\psi_T(\lambda)$, to describe the ground-state of $H(\lambda)$ [note that, for the sake of clarity and simplicity, we shall denote in what follows $\psi_T(0)$, $H(0)$, and $E_0(0)$ as ψ_T, H, and E_0, respectively].

The exact average of the observable can be decomposed as

$$\frac{\langle \psi_0 | O | \psi_0 \rangle}{\langle \psi_0 | \psi_0 \rangle} = \left. \frac{dE_v(\lambda)}{d\lambda} \right|_{\lambda=0} + \epsilon(\delta\psi, \delta\psi'), \tag{16}$$

where $E_v(\lambda)$ is the variational energy associated with $\psi_T(\lambda)$,

$$E_v(\lambda) \equiv \langle E_L(\lambda) \rangle_{\psi_T^2(\lambda)} = \left\langle \frac{H(\lambda)\psi_T(\lambda)}{\psi_T(\lambda)} \right\rangle_{\psi_T^2(\lambda)} \tag{17}$$

and ϵ is some correction depending on $\delta\psi = \psi_0 - \psi_T$ and its derivative, and vanishing when the exact wave function is used as trial wave function.

Now, the important point is that the derivative of the variational energy, $[dE_v(\lambda)/d\lambda]|_{\lambda=0}$, is expected to be a better estimate of the exact average than the ordinary average of the bare estimator, Eq. (11), when properly chosen λ-dependent trial wave functions are used. This is true since the standard estimator, Eq. (11), can be re-expressed as a particular case of the derivative of the variational energy for a λ-*independent* trial wave function, a choice which is clearly not optimal. Before more quantitatively justifying this statement, let us rewrite the derivative as an ordinary average over the density ψ_T^2. This can be easily done, it gives

$$\left.\frac{dE_v(\lambda)}{d\lambda}\right|_{\lambda=0} = \langle \tilde{O} \rangle_{\psi_T^2}, \tag{18}$$

where $\tilde{O}$ is a new modified local operator written as

$$\tilde{O} \equiv O + \frac{(H-E_L)\psi_T'}{\psi_T} + 2(E_L-E_v)\frac{\psi_T'}{\psi_T}. \tag{19}$$

In this latter formula, and in the formulas to follow, we shall use the following simplified notation:

$$f' \equiv \left.\frac{df(\lambda)}{d\lambda}\right|_{\lambda=0}, \tag{20}$$

where $f(\lambda)$ is some arbitrary function of λ.

Now, we have to justify the first important result, that the new estimator $\tilde{O}$ is a better estimator for the exact average than the bare observable O. For that purpose, we compute the systematic error in the corresponding average and the variance of the new operator. Regarding the systematic error we can write

$$\Delta_{\tilde{O}} \equiv \langle \tilde{O} \rangle_{\psi_T^2} - \langle O \rangle_{\psi_0^2} = \left.\frac{d[E_v(\lambda) - E_0(\lambda)]}{d\lambda}\right|_{\lambda=0}. \tag{21}$$

Let us denote $\psi_0(\lambda)$ as the exact ground state of $H(\lambda)$ [with $\psi_0(0) = \psi_0$]. Using the equality

$$E_v(\lambda) - E_0(\lambda)$$
$$= \frac{\langle \psi_T(\lambda) - \psi_0(\lambda) | H(\lambda) - E_0(\lambda) | \psi_T(\lambda) - \psi_0(\lambda) \rangle}{\langle \psi_T(\lambda) | \psi_T(\lambda) \rangle} \tag{22}$$

and choosing the following convention of normalization

$$\langle \psi_T(\lambda) | \psi_T(\lambda) \rangle = 1, \tag{23}$$

the derivative can be easily computed. We get

$$\Delta_{\tilde{O}} = \langle \psi_T - \psi_0 | O - \langle O \rangle_{\psi_0^2} | \psi_T - \psi_0 \rangle + 2\langle \psi_T - \psi_0 | H$$
$$- E_0 | \psi_T' - \psi_0' \rangle. \tag{24}$$

As can be seen, the systematic error is now of order two in the errors $\psi_T - \psi_0$ and $\psi_T' - \psi_0'$,

$$\Delta_{\tilde{O}} \sim O[(\psi_T - \psi_0)(\psi_T' - \psi_0')]. \tag{25}$$

Now, let us compute the variance defined as

$$\sigma^2(\tilde{O}) = \langle (\tilde{O} - \langle \tilde{O} \rangle_{\psi_T^2})^2 \rangle_{\psi_T^2}. \tag{26}$$

Using Eqs. (18) and (19) and the fact that $E_L' = O + (H - E_L)\psi_T'/\psi_T$ we can express the difference $\tilde{O} - \langle \tilde{O} \rangle_{\psi_T^2}$ as follows:

$$\tilde{O} - \langle \tilde{O} \rangle_{\psi_T^2} = E_L' - E_v' + 2(E_L - E_v)\frac{\psi_T'}{\psi_T}. \tag{27}$$

For the sake of clarity, let us distinguish two different contributions in the difference. The first contribution is given by

$$E_L' - E_v' = \left.\frac{d[E_L(\lambda) - E_v(\lambda)]}{d\lambda}\right|_{\lambda=0}. \tag{28}$$

Using expression (8) for $[E_L(\lambda) - E_v(\lambda)]$ and performing the derivative one obtains

$$E_L' - E_v' = \frac{(O - \langle O \rangle_{\psi_0^2})(\psi_T - \psi_0)}{\psi_T} + \frac{(H - E_0)(\psi_T' - \psi_0')}{\psi_T}$$
$$- \frac{(H - E_0)(\psi_T - \psi_0)}{\psi_T}\frac{\psi_T'}{\psi_T} + \langle O \rangle_{\psi_0^2} - \langle \tilde{O} \rangle_{\psi_T^2}. \tag{29}$$

This latter expression is clearly of order one in $\psi_T - \psi_0$ and its derivative, $\psi_T' - \psi_0'$. The second contribution on the right-hand side of Eq. (27) is proportional to $E_L - E_v$. We have already seen that it is of order one in $\psi_T - \psi_0$, Eqs. (7) and (8). Finally, $\tilde{O} - \langle \tilde{O} \rangle_{\psi_T^2}$ is found to be of order one in $\psi_T - \psi_0$ and $\psi_T' - \psi_0'$. The variance, Eq. (26), is therefore of order two,

$$\sigma^2(\tilde{O}) \sim O[(\psi_T - \psi_0)(\psi_T' - \psi_0')]. \tag{30}$$

To summarize, using the HF theorem we are able to construct an improved observable $\tilde{O}$, Eq. (19), having a quadratic zero-variance zero-bias property, Eqs. (25) and (30), similar to what is known for the energy case, Eqs. (7) and (9). The improved estimator $\tilde{O}$ depends only on one single quantity, namely $\psi_T(\lambda)$. Accordingly, to get accurate results we need to choose in the neighborhood of $\lambda = 0$ a trial function accurate enough to get not only a small difference in wave functions but also the derivative of the wave functions. In practice, this latter point is particularly difficult to fulfill. Indeed, at fixed values of λ, it is known that the minimization of the fluctuations of the local energy can allow an important reduction of the error in the trial wave function. However, there is no reason why it should also lead to a satisfactory representation of the derivative of the trial wave function.

In order to escape from this difficulty we propose here to work directly at $\lambda = 0$ and to optimize *independently* the trial wave function ψ_T and its derivative ψ_T'. Such procedure is justified since it corresponds to choosing the following expression as λ-dependent trial wave function

$$\psi_T(\lambda) = \psi_T + \lambda \tilde{\psi}, \tag{31}$$

10540   J. Chem. Phys., Vol. 119, No. 20, 22 November 2003

R. Assaraf and M. Caffarel

where $\tilde{\psi}$ is some new *independent* function playing the role of a trial function for the derivative of the ground state at $\lambda = 0$. In this case, the renormalized observable can be re-written under the final form

$$\tilde{O} \equiv O + \frac{(H-E_L)\tilde{\psi}}{\psi_T} + 2(E_L - E_v)\frac{\tilde{\psi}}{\psi_T}, \qquad (32)$$

where the pair of functions $(\psi_T, \tilde{\psi})$ is the current guess for the exact solution (ψ_0, ψ_0').

Let us now turn our attention to the problem of optimizing the two trial functions $(\psi_T, \tilde{\psi})$. Regarding ψ_T we know that the standard procedure consists in minimizing the variance of the local energy with respect to the parameters of the trial function. Quite remarkably, we have here a similar result for $\tilde{\psi}$: the best choice is obtained by minimizing the variance of the renormalized operator $\tilde{O}$ with respect to parameters of $\tilde{\psi}$.

To prove this property it is sufficient to show that the zero-variance (or zero-fluctuations) equations for E_L and $\tilde{O}$:

$$E_L = \langle E_L \rangle_{\psi_T^2},$$
$$\tilde{O} = \langle \tilde{O} \rangle_{\psi_T^2}, \qquad (33)$$

are equivalent to the equations defining ψ_0 and ψ_0', namely

$$(H-E_0)\psi_0 = 0,$$
$$(H-E_0)\psi_0' + (O - \langle O \rangle_{\psi_0^2})\psi_0 = 0. \qquad (34)$$

In these formulas, the first equation is just the ordinary Schrödinger equation. The second one is obtained by deriving the Schrödinger equation:

$$H(\lambda)\psi(\lambda) = E_0(\lambda)\psi(\lambda) \qquad (35)$$

with respect to λ at $\lambda = 0$. Note that Eq. (34) determines an unique solution, $(\psi_0, \psi_0', E_0, \langle O \rangle_{\psi_0^2})$, as soon as H has a nondegenerate ground state. Now, using Eqs. (19) and (2) for the definitions of $\tilde{O}$ and E_L, respectively, the system of equations (33) can be rewritten under the form

$$(H-E_v)\psi_T = 0, \qquad (36)$$

$$(H-E_v)\psi_T' + (O - \langle O \rangle_{\psi_T^2})\psi_T = 0, \qquad (37)$$

which are nothing but Eq. (34) with $(\psi_T, \psi_T') = (\psi_0, \psi_0')$. Accordingly, the zero-variance equations (33) admits this latter pair of functions as unique solution.

In practical calculations, different strategies of optimization can be employed. A first approach consists of minimizing *separately* the variance of the local energy with respect to the wave function ψ_T and the variance of $\tilde{O}$ with respect to $\tilde{\psi}$. In this way, we get an optimal trial wave function ψ_T for the energy and the best derivative at fixed ψ_T. However, let us emphasize that this approach is not the most general: we can also minimize both variances simultaneously with respect to the two independent functions. Another remark is that the second equation of system (34) can be viewed as an ordinary first-order perturbation equation. This is expected

since, when λO is considered as a perturbation of the Hamiltonian H, ψ_0' is nothing but the first-order correction to the ground state and $\langle O \rangle_{\psi_0^2}$ the first-order correction to the energy.

Finally, let us end this section by commenting in more detail on the various terms entering expression (32) of the improved operator. Three different contributions can be distinguished.

(i) The ordinary bare estimator O corresponding to $\tilde{\psi} = 0$.

(ii) A second contribution given by $(H-E_L)\tilde{\psi}/\psi_T$. It is easy to verify that this contribution has a zero average over the density ψ_T^2,

$$\langle (H-E_L)\tilde{\psi}/\psi_T \rangle_{\psi_T^2} = 0. \qquad (38)$$

Accordingly, its role is to lower the variance of the improved estimator without changing the average of the observable (no influence on the systematic error). Note that for applications where the stationary density is known and can be exactly sampled (that is, there is no systematic error in the average) the use of contributions (i) and (ii) is sufficient. Important examples include all "classical" Monte Carlo simulations based on the METROPOLIS algorithm or one of its variants. Such a possibility was the subject of a previous work.[23]

(iii) A third term given by $2(E_L - E_v)(\tilde{\psi}_T/\psi_T)$. This contribution has a very small impact on the statistical fluctuations since the variance of $(E_L - E_v)$ is of order two in the trial wave function error for any choice of $\tilde{\psi}$. Its main effect is to take into account the change of stationary density under the external perturbation defined by the observable and, therefore, to lower the systematic error in the expectation value of the observable. Note that in the limit $\psi_T = \psi_0$, this contribution reduces to zero and, therefore, the average of this term can be understood as a correction to the Hellmann–Feynman formula when ψ_T is not the exact ground state [note that similar corrections to the HF formula exist also in more traditional *ab initio* calculations, e.g., the "Pulay force" [36] resulting from approximate Hartree–Fock (or local density approximation) orbitals in self-consistent schemes].

B. More improved estimators: Use of coordinate transformations

In this section it is shown how to generalize further our renormalized operators. The basic idea of the generalization is based on an original idea proposed by Umrigar[28] and recently extended by Filippi and Umrigar.[33] Working in a finite difference formalism the forces are computed as a small but finite difference of energies for two close enough geometries. In order to minimize the fluctuations a correlated sampling method, in which a common Monte Carlo density (the so-called primary one) is used for the two close geometries, is employed. Written within our notations and taking the limit of the two geometries infinitely close ($\delta R \rightarrow 0$ is equivalent to $\lambda \rightarrow 0$) it means that the variational energy is written under the form

J. Chem. Phys., Vol. 119, No. 20, 22 November 2003

Observables in QMC   10541

$$E_v(\lambda) = \frac{\left\langle E_L(\lambda) \frac{\psi_T^2(\lambda)}{\psi_T^2} \right\rangle_{\psi_T^2}}{\left\langle \frac{\psi_T^2(\lambda)}{\psi_T^2} \right\rangle_{\psi_T^2}}, \tag{39}$$

where $\psi_T(\lambda)$ is the trial wave function chosen for a parameter λ and ψ_T is the reference (primary) trial wave function.

The price to pay when doing that is the introduction of some additional fluctuations associated with the weight $\psi_T^2(\lambda)/\psi_T^2$. The remedy they propose to deal with this problem is to use a specific coordinate transformation (space-warp transformation) based on physical motivations: The transformation is built so that the electrons close to a given nucleus move almost rigidly with that nucleus when the geometry is changed. Here, we generalize this idea: coordinate transformation can help to minimize the relative fluctuations when varying the external parameter λ. As a physical consequence, estimators built from the derivative are expected to have smaller fluctuations and smaller systematic errors.

Let us write a general coordinate transformation as follows:

$$\vec{y} = \vec{T}(\lambda, \vec{x}), \tag{40}$$

where the vector $\vec{x}$ (or $\vec{y}$) denotes the set of the $3n_{elec}$ electronic coordinates. Using this transformation the variational energy at a given λ can be written as

$$E_v(\lambda) = \frac{\left\langle E_L[\lambda, \vec{T}(\lambda, \vec{x})] J(\lambda, \vec{x}) \frac{\psi_T^2[\lambda, \vec{T}(\lambda, \vec{x})]}{\psi_T^2(\vec{x})} \right\rangle_{\psi_T^2}}{\left\langle J(\lambda, \vec{x}) \frac{\psi_T^2[\lambda, \vec{T}(\lambda, \vec{x})]}{\psi_T^2(\vec{x})} \right\rangle_{\psi_T^2}}, \tag{41}$$

where $J(\lambda, \vec{x})$ is the Jacobian of the transformation. Introducing the vector field $\vec{v}$ such that at first order in λ we have

$$\vec{T}(\lambda, \vec{x}) = \vec{x} + \lambda \vec{v}(\vec{x}) + O(\lambda^2), \tag{42}$$

we can compute the derivative of the variational energy with respect to λ at $\lambda = 0$. After some simple but tedious algebra we get the following equality:

$$\left. \frac{dE_v(\lambda)}{d\lambda} \right|_{\lambda=0} = \langle \tilde{O} \rangle_{\psi_T^2}, \tag{43}$$

where $\tilde{O}$ is a new renormalized operator given by

$$\tilde{O} \equiv O + \frac{(H - E_L)\psi_T'}{\psi_T} + 2(E_L - E_v)\frac{\psi_T'}{\psi_T} + \frac{\vec{\nabla}[(E_L - E_v)\psi_T^2 \vec{v}]}{\psi_T^2}. \tag{44}$$

To derive this expression we have used the fact that the Jacobian defined as

$$J(\lambda, \vec{x}) = \det \left[\frac{\partial T_i(\lambda, \vec{x})}{\partial x_j} \right] \tag{45}$$

has the following small-λ expression:

$$J(\lambda, \vec{x}) = \det \left[\delta_{ij} + \lambda \frac{\partial v_i}{\partial x_j} \right] + O(\lambda^2) \tag{46}$$

and, therefore,

$$J(0, \vec{x}) = 1, \tag{47}$$

$$\frac{\partial J}{\partial \lambda}(0, \vec{x}) = \vec{\nabla} \cdot \vec{v}. \tag{48}$$

This more general operator is identical to the operator derived in Sec. II A plus a new contribution resulting from the derivative of the coordinate transformation. This new term has a zero average over the VMC distribution ψ_T^2. Accordingly, its main role is to reduce further the statistical error. However, it is important to emphasize that, when the trial function $\tilde{\psi}$ and the vector field $\vec{v}$ are optimized simultaneously it has also an influence on the magnitude of the systematic error.

III. BEYOND VARIATIONAL MONTE CARLO

In Sec. II we have shown how to construct improved observables, $\tilde{O}$, associated with accurate expectation values:

$$\frac{\langle \psi_T | \tilde{O} | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle} = \frac{\langle \psi_0 | O | \psi_0 \rangle}{\langle \psi_0 | \psi_0 \rangle} + O[(\psi_T - \psi_0)(\tilde{\psi} - \psi_0')]. \tag{49}$$

When the error $\delta\psi' = \tilde{\psi} - \psi_0'$ in the trial function for the derivative is comparable to the error in the trial function for the ground state, $\delta\psi = \psi_T - \psi_0$, the accuracy reached with the preceding variational estimate, Eq. (49), can be comparable to the very good accuracy usually obtained for total energies. However, despite this remarkable improvment, we are still left with some small residual systematic error associated with approximate ψ_T and $\tilde{\psi}$. In the energy case it is known that this error can be entirely suppressed (at least for systems with no nodes or known nodes) by averaging the local energy over the mixed DMC probability distribution, $\pi_{DMC} \sim \psi_T \psi_0$ instead of the VMC distribution, $\pi_{VMC} \sim \psi_T^2$. Unfortunately, we have no such result for the improved observables defined here. However, as we shall see now, we can still define some approximate way for recovering most of the error.

A natural way of defining an exact extimator for the observable is to consider the derivative of the exact DMC energy estimator instead of the VMC one,

$$E_0(\lambda) \equiv \langle E_L(\lambda) \rangle_{\psi_T(\lambda)\psi_0(\lambda)} = \left\langle \frac{H(\lambda)\psi_T(\lambda)}{\psi_T(\lambda)} \right\rangle_{\psi_T(\lambda)\psi_0(\lambda)}. \tag{50}$$

Making the derivative and rewriting the result as an ordinary average we get

$$\left. \frac{dE_0(\lambda)}{d\lambda} \right|_{\lambda=0} = \langle \tilde{O} \rangle_{\psi_T \psi_0}, \tag{51}$$

where $\tilde{O}$ is written as

$$\tilde{O} \equiv O + \frac{(H-E_L)\psi_T'}{\psi_T} + (E_L - E_0)\left(\frac{\psi_T'}{\psi_T} + \frac{\psi_0'}{\psi_0}\right). \quad (52)$$

Of course, written under the above-given form, this exact estimator is useless since the exact wave function is not known. Here, we propose to make the following natural approximation:[15,37]

$$\frac{\psi_0'}{\psi_0} = \frac{\psi_T'}{\psi_T}. \quad (53)$$

Therefore, our final approximate DMC estimator is written as

$$\tilde{O}_{\text{DMC}} \equiv O + \frac{(H-E_L)\tilde{\psi}}{\psi_T} + 2(E_L - E_0)\frac{\tilde{\psi}}{\psi_T}, \quad (54)$$

where $\tilde{\psi}$ is as usual our trial function for the derivative of the exact wave function. Note that this estimator is very similar to the VMC one, Eq. (32). The only difference lies in the value of the average energy, $E_0 = \langle E_L \rangle$, entering the definition of $\tilde{O}_{\text{DMC}}$. More precisely, we have

$$\tilde{O}_{\text{DMC}} - \tilde{O}_{\text{VMC}} = 2(E_v - E_0)\frac{\tilde{\psi}}{\psi_T}. \quad (55)$$

Now, in order to further reduce the error let us show that we can generalize the usual "hybrid formula" $\langle O \rangle_{\text{hybrid}} \equiv 2\langle O \rangle_{\text{DMC}} - \langle O \rangle_{\text{VMC}}$ defined for bare observables to the case of improved observables. To do that, let us evaluate the quantity $\langle \delta\psi | \tilde{O}_{\text{DMC}} | \delta\psi \rangle$ where $\delta\psi = \psi_T - \psi_0$,

$$\langle \delta\psi | \tilde{O}_{\text{DMC}} | \delta\psi \rangle = \langle \psi_T | \tilde{O}_{\text{DMC}} | \psi_T \rangle - 2\langle \psi_T | \tilde{O}_{\text{DMC}} | \psi_0 \rangle$$
$$+ \langle \psi_0 | \tilde{O}_{\text{DMC}} | \psi_0 \rangle, \quad (56)$$

which leads to

$$2\langle \psi_T | \tilde{O}_{\text{DMC}} | \psi_0 \rangle - \langle \psi_T | \tilde{O}_{\text{DMC}} | \psi_T \rangle$$
$$= \langle \psi_0 | O | \psi_0 \rangle + A + O(\delta\psi^2),$$

where the intermediate quantity A is defined as

$$A \equiv \langle \psi_0 | \frac{(H-E_L)\tilde{\psi}}{\psi_T} + 2(E_L - E_0)\frac{\tilde{\psi}}{\psi_T} | \psi_0 \rangle.$$

Expanding A in terms of $\delta\psi = \psi_T - \psi_0$ we get

$$A = 2\langle \psi_T | E_L - E_0 | \tilde{\psi} \rangle - 2\langle \delta\psi | (H-E_L) | \tilde{\psi} \rangle$$
$$- 4\langle \delta\psi | E_L - E_0 | \tilde{\psi} \rangle + O(\delta\psi^2).$$

Using now the equality

$$E_L - E_0 = \frac{(H-E_0)\delta\psi}{\psi_T}$$

we obtain

$$A = O(\delta\psi^2).$$

This latter result shows that the error in the hybrid estimator is of order two in $\delta\psi$,

$$2\langle \tilde{O}_{\text{DMC}} \rangle_{\psi_T\psi_0} - \langle \tilde{O}_{\text{DMC}} \rangle_{\psi_T^2} = \langle \psi_0 | O | \psi_0 \rangle + O(\delta\psi^2), \quad (57)$$

thus generalizing the standard result for the bare observable. Note that we can use either $\tilde{O}_{\text{DMC}}$ or $\tilde{O}_{\text{VMC}}$ [Eq. (32)] in the latter formula since the difference between the two renormalized operators is proportional to $E_v - E_0$, Eq. (55), and, therefore, is also of order two in $\delta\psi$ [Eq. (7)]. Finally, note that for Fermi systems treated within the fixed-node approximation (DMC density given by $\psi_T\psi_0^{\text{FN}}$) expression (57) is slightly modified, we have

$$2\langle \tilde{O}_{\text{DMC}} \rangle_{\psi_T\psi_0^{\text{FN}}} - \langle \tilde{O}_{\text{DMC}} \rangle_{\psi_T^2}$$
$$= \langle \psi_0 | O | \psi_0 \rangle + O(\delta\psi^2) + O[(\psi_0^{\text{FN}} - \psi_0)(\tilde{\psi} - \psi_0')]. \quad (58)$$

The latter formula shows that the dependence of the systematic error on the quality of $\tilde{\psi}$ is quite small as soon as the nodes of the trial wave function are reasonably good. When using coordinate transformation we have similar results. The exact DMC estimator is found to be

$$\tilde{O} \equiv O + \frac{(H-E_L)\psi_T'}{\psi_T} + (E_L - E_0)\left(\frac{\psi_T'}{\psi_T} + \frac{\psi_0'}{\psi_0}\right)$$
$$+ \frac{\vec{\nabla}[(E_L - E_0)\psi_T\psi_0\vec{v}]}{\psi_T\psi_0} \quad (59)$$

and we propose to use the following approximate form:

$$\tilde{O}_{\text{DMC}} \equiv O + \frac{(H-E_L)\tilde{\psi}}{\psi_T} + 2(E_L - \langle E_L \rangle)\frac{\tilde{\psi}}{\psi_T}$$
$$+ \frac{\vec{\nabla}[(E_L - \langle E_L \rangle)\psi_T^2\vec{v}]}{\psi_T^2}. \quad (60)$$

Because the difference $(E_L - \langle E_L \rangle)$ is of order $\delta\psi$ it is easy to verify that the error in the hydrid estimator given by Eq. (57) remains here also of order two.

Before ending this section let us emphasize that it is possible to write a closed computable expression for the exact estimator of the observable, Eq. (52), by expressing the unknown quantity ψ_0'/ψ_0 as a computable stochastic average. Choosing a λ-independent trial wave function ψ_T we can write[18,38,39]

$$\psi_0(\lambda,x) = \psi_T(x) \lim_{T\to+\infty}$$
$$\times \left\langle\!\!\left\langle \exp\left(-\int_0^T ds\, E_L[\lambda,x(s)]\right)\right\rangle\!\!\right\rangle_{x(0)=x}, \quad (61)$$

where x denotes an arbitrary point in configuration space and $\langle\!\langle \ldots \rangle\!\rangle_{x(0)=x}$ denotes the sum over all drifted random walks of length T starting at x as obtained in a pure diffusion Monte Carlo (PDMC) scheme (DMC without branching).[18] Of course, a similar formula can also be obtained in a DMC scheme.[17] Now, using formula (61) we get

J. Chem. Phys., Vol. 119, No. 20, 22 November 2003

Observables in QMC   10543

$$\frac{\psi_0'}{\psi_0} = \lim_{T \to +\infty} \int_0^T dt$$

$$\times \frac{\langle\!\langle O[x(t)]\exp(-\int_0^T ds\, E_L[\lambda,x(s)])\rangle\!\rangle_{x(0)=x}}{\langle\!\langle \exp(-\int_0^T ds\, E_L[\lambda,x(s)])\rangle\!\rangle_{x(0)=x}} \quad (62)$$

and, therefore, the exact estimator can be written in terms of a standard part plus a time integral of the two-point correlation function between the local energy and the observable

$$\tilde{O} = O + \frac{(H-E_L)\tilde{\psi}}{\psi_T} + (E_L - \langle E_L\rangle)\frac{\tilde{\psi}}{\psi_T} + \lim_{T \to +\infty}\int_0^T dt$$

$$\times \frac{\langle\!\langle (E_L - \langle E_L\rangle)[x(0)]O[x(t)]\exp(-\int_0^T ds\, E_L)\rangle\!\rangle_{x(0)=x}}{\langle\!\langle \exp(-\int_0^T ds\, E_L)\rangle\!\rangle_{x(0)=x}}.$$

$$(63)$$

It is important to emphasize that this latter estimator is exact: averaged over the mixed DMC distribution it leads to an unbiased estimate of the exact average. However, the correlator can only be obtained within a forward walking scheme and, therefore, the stability in time is not guaranteed. In this work, we shall not use this expression, its implementation will be presented in a forthcoming work.

IV. APPLICATION TO FORCES

The average force between atoms in a molecular system is defined as

$$\bar{F}_{q_i} \equiv -\frac{\partial E_0(\mathbf{q})}{\partial q_i}, \quad (64)$$

where $E_0(\mathbf{q})$ is the total electronic ground-state energy for a given nuclear configuration; $\mathbf{q}$ represents the $3N_{\text{nucl}}$ nuclear coordinates (N_{nucl}, number of nuclei) and q_i the particular force component in which we are interested.

Defining the local force as follows:

$$F_{q_i}(\mathbf{x},\mathbf{q}) \equiv -\frac{\partial V(\mathbf{x},\mathbf{q})}{\partial q_i}, \quad (65)$$

where $\mathbf{x}$ represents the $3n_{\text{elec}}$ electronic coordinates (n_{elec}, number of electrons) and V the total potential energy operator, and making use of the HF theorem the average force can be rewritten as the statistical average of the local force over the exact distribution $\psi_0^2(\mathbf{x})$:

$$\bar{F}_{q_i} = \langle F_{q_i}(\mathbf{x},\mathbf{q})\rangle_{\psi_0^2(\mathbf{x})}. \quad (66)$$

Written under this form the various proposals presented in the preceding sections can be applied to the calculation of the average force. It is important to emphasize that for approximate probability densities (VMC or DMC) the HF theorem is no longer valid and a systematic error in the statistical average $\langle F_{q_i}(\mathbf{x},\mathbf{q})\rangle$ is introduced. However, it is not a problem here since it is the purpose of this work to show that, by using suitable improved estimators, this error can be reduced and even suppressed in the zero-bias limit.

In order to discuss the various aspects of the method we shall restrict ourselves to the case of diatomic molecules. Let us consider a diatomic molecule AB with atom A located at

$(R,0,0)$ and atom B located at the origin. The only nonzero component of the local force acting on the nucleus A is the x component given by

$$F = -\frac{\partial V}{\partial R} = \frac{Z_A Z_B}{R^2} - Z_A \sum_{i=1}^{n_{\text{elec}}} \frac{(x_i - R)}{|\mathbf{r}_i - \mathbf{R}|^3}. \quad (67)$$

In this work we present a number of VMC and fixed-node DMC calculations for the diatomic molecules H_2, LiH, and Li_2. Implementation of the quantum Monte Carlo methods is well known and will not be discussed here. For the H_2 molecule the trial wave function used has the following simple form:

$$\psi_T = (1s_A 1s_B + 1s_B 1s_A) + c(1s_A 1s_A + 1s_B 1s_B), \quad (68)$$

where $1s_M$ is a $1s$-Slater function centered at nucleus $M = A,B$ with exponent μ and c a parameter describing the amount of ionic contribution into the wave function. Of course, much more accurate trial wave functions can be constructed for H_2. However, our purpose here is to show that such a simple form for ψ_T is already sufficient to get accurate values of the force.

For LiH and Li_2 we have employed two types of trial wave function. Our main choice is standard in QMC calculations for molecules. The trial wave function is made of a determinant of single-particle orbitals multiplied by a Jastrow factor. The determinantal part is obtained from a restricted Hartree–Fock (RHF) calculation and only the Jastrow factor is optimized. As we shall see in the following, we have also used valence-bond (VB)-type wave functions consisting of a number of determinants multiplied by a Jastrow factor. We have used such a multideterminantal description to reproduce correctly the large interatomic distance regime (dissociation limit). In the case of LiH the determinantal part consists of three determinants corresponding to the covalent VB resonating structure: $(1s_{\text{Li}})^2[2s_{\text{Li}}1s_{\text{H}} + 1s_{\text{H}}2s_{\text{Li}}]$ ($\{1s_{\text{Li}},2s_{\text{Li}},1s_{\text{H}}\}$ optimized *atomic* orbitals for the Li and H atoms, orbitals occupied by electrons α and β antisymmetrized separately), and one ionic VB structure: $(1\tilde{s}_{\text{Li}})^2(1\tilde{s}_{\text{H}})^2$ ($1\tilde{s}_{\text{Li}},1\tilde{s}_{\text{H}}$ optimized atomic orbitals for the Li^+ and H^- ions). In the case of Li_2 we have considered a six-determinant representation consisting of the three covalent VB structures describing the resonance between atomic orbitals $(2s_A,2s_B)$, $(2p_{y_A},2p_{y_B})$, and $(2p_{z_A},2p_{z_B})$. This latter trial wave function reproduces not only the dissociation limit but also a major part of the $2s-2p$ near-degeneracy.

In Figs. 1–3 the energy curves obtained for H_2, LiH, and Li_2 are presented. Upper curves are the VMC curves (open squares joined by a dotted line). For H_2 the two parameters c and μ have been optimized for each interatomic distance. For LiH and Li_2 the Jastrow-RHF wave function (one determinant) has been used. All the parameters entering the Jastrow factor have been optimized for all distances. Optimizations have been performed by minimizing the variance of the local energy using the correlated sampling method of Umrigar *et al.*[40] The first important observation is that, except for H_2, VMC curves are not smooth as a function of R.
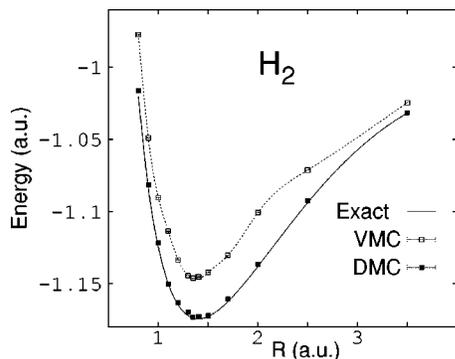
FIG. 1. H_2 molecule. Variational Monte Carlo (VMC) energies (open squares), diffusion Monte Carlo (DMC) energies (closed squares), and exact nonrelativistic curve (solid line). The dotted line between VMC results is a simple linear interpolation to guide the eye.

Such a result is not surprising: It is typical of a situation where an approximate trial wave function is optimized *independently* for different values of an external parameter (here, R) with respect to a large number of variables (for LiH and Li_2 we have used about 30 independent variational parameters). Depending on the initial conditions (which are themselves very dependent on R) the algorithm used for minimizing the variance can be trapped within one of the various *local* minima. As a consequence, the actual value obtained for the variance (and the corresponding energy) can vary abruptly even when the external parameter is changed smoothly. Of course, this problem can be solved in principle by making very careful optimizations on very large samples. Indeed, the functional form of the trial wave function being identical at all distances a smooth curve must be obtained when the correct lowest minimum of the variance is obtained at each distance. Here, this is the case for H_2 whose trial wave function contains only two variational parameters. However, for large systems including a much larger number of variational parameters and nuclear degrees of freedom, the possibility of fully optimizing the trial wave function is just unrealistic. As an important consequence, let us emphasize that, in practice, there is no hope of obtaining meaningful forces by making straight finite differences of optimized variational energies without using some sort of correlated
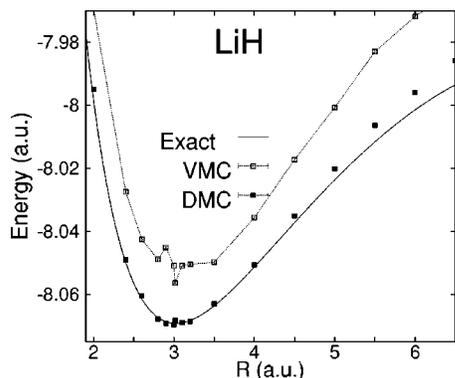
FIG. 2. LiH molecule. Variational Monte Carlo (VMC) energies (open squares), fixed-node diffusion Monte Carlo (DMC) energies (closed squares), and exact nonrelativistic curve (solid line). The dotted line between VMC results is a simple linear interpolation to guide the eye.
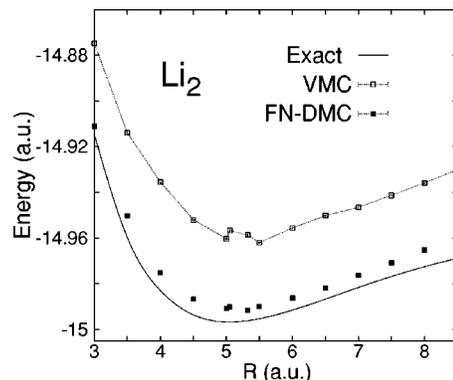
FIG. 3. Li_2 molecule. Variational Monte Carlo (VMC) energies (open squares), fixed-node diffusion Monte Carlo (DMC) energies (closed squares), and exact nonrelativistic curve (solid line). The dotted line between VMC results is a simple linear interpolation to guide the eye.

sampling scheme. This is a good illustration of how difficult the calculation of forces is within a QMC framework.

Intermediate points (closed squares) are the DMC results obtained from fixed-node calculations using the optimized VMC trial wave functions. In sharp contrast with VMC, the DMC curves are now regular. This is so because, unlike VMC, fixed-node DMC averages do not depend on the particular form of the trial function used, except for the nodal structure. Here, the nodal hypersurfaces vary smoothly as a function of the distance and, therefore, the corresponding DMC energy curves are smooth within error bars. The second important observation is related to the global shape of the curves. Ideally, we are interested in having energy curves which differ from the exact curve only by a constant independent of the distance. This is indeed the condition to obtain accurate derivatives. Here, it is not the case for the VMC curves of LiH and Li_2; the difference between the VMC and exact energy curves is an increasing function of the distance. It is not a surprise since in both cases the trial wave function is built from a single RHF determinant based on delocalized molecular orbitals which leads to a wrong description of the dissociation limit. However, and very interestingly, fixed-node DMC results have a much better behavior at large distances. As a consequence, one may expect at this stage to obtain accurate forces from the derivative of the fixed-node energy curve even when relatively crude wave functions are used. Finally, let us note that the quality of the fixed-node calculations for the molecules considered here is quite good. To give an example, at the equilibrium distance of the Li_2 molecule, the total energy obtained is $-14.9901(6)$ to be compared with the exact nonrelativistic value of $E_0 = -14.9954$. The amount of correlation energy recovered within the fixed-node approximation is about 95.7%. A similar quality is obtained for other distances and also for the LiH molecule. In the case of the nodeless H_2 molecule (no fixed-node approximation), the DMC energies agree perfectly well with the exact ones. For the H_2 molecule the variance of the local energy varies between 0.3 at $R=0.8$ and 0.02 at $R=3.5$; for LiH the variance is about 0.07, and for Li_2 it varies between 0.09 and 0.2.

The crucial point when implementing the various formulas presented in the preceding section is the choice of the
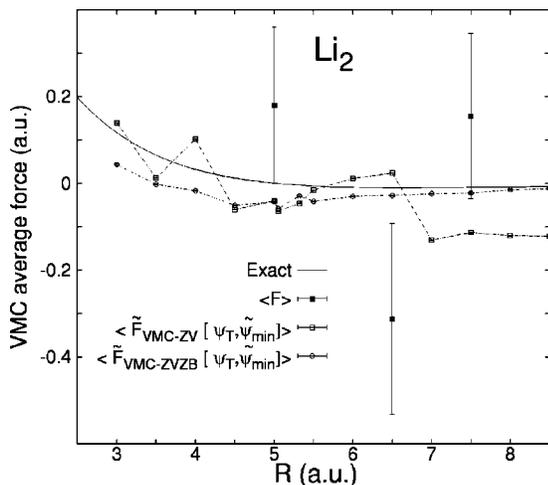
FIG. 4. Various VMC average forces for Li_2. Closed squares with large error bars: $\langle F \rangle$, Eq. (67). Open squares joined by the dashed line: $\langle \widetilde{F}_{\text{VMC-ZV}}[\psi_T, \widetilde{\psi}_{\min}] \rangle$, Eq. (72). Circles joined with the dotted line: $\langle \widetilde{F}_{\text{VMC-ZVZB}}[\psi_T, \widetilde{\psi}_{\min}] \rangle$, Eq. (73). Solid line: exact nonrelativistic force curve.

trial function $\widetilde{\psi}$ for the derivative. In our previous study on forces[24] where we focused our attention on the reduction of statistical fluctuations only, we proposed to employ the minimal form leading to a finite variance of the renormalized local force. As can be viewed from Eq. (67), at short electron–nucleus distance r the local force behaves as $F \sim 1/r^2$ and, therefore, the variance $\langle F^2 \rangle - \langle F \rangle^2$ is infinite. This well-known problem has been discussed in different places (see, e.g., Ref. 15, Chap. 8.2 or Ref. 31). Here, the "minimal" form removing the singular part responsible for the infinite variance is written as

$$\widetilde{\psi}_{\min}(\mathbf{x}) = Q \psi_T, \tag{69}$$

where Q is given by

$$Q = Z_A \sum_{i=1}^{n_{\text{elec}}} \frac{(x_i - R)}{|\mathbf{r}_i - \mathbf{R}|}. \tag{70}$$

To see this, we just need to compute the following quantity:

$$\frac{(H - E_L)\widetilde{\psi}_{\min}}{\psi_T} = Z_A \sum_{i=1}^{n_{\text{elec}}} \frac{(x_i - R)}{|\mathbf{r}_i - \mathbf{R}|^3} - \nabla Q \cdot \nabla \psi_T / \psi_T. \tag{71}$$

By adding this latter quantity to the bare local force, Eq. (67), the singular part is exactly removed, the remaining contribution having a finite variance. In what follows, $\widetilde{\psi}_{\min}$ will be referred to as the minimal form for $\widetilde{\psi}$.

In Fig. 4 we present various VMC calculations of the average force for the Li_2 molecule as a function of the interatomic distance. A first set of points (closed squares with very large error bars at $R = 5$, $R = 6.5$, and $R = 7.5$) are results obtained from the ordinary bare estimator, Eq. (67). Open squares (with small error bars) joined by the dashed curve correspond to results obtained by using $\widetilde{\psi}_{\min}$ as trial function for the derivative

$$\widetilde{F}_{\text{VMC-ZV}}[\psi_T, \widetilde{\psi}_{\min}] = F + \frac{(H - E_L)\widetilde{\psi}_{\min}}{\psi_T}. \tag{72}$$

This estimator can be viewed as the simplest improved estimator we can think of having a finite variance; it corresponds to the form employed in our previous study.[24] The subscript ZV (zero variance) is used here to emphasize that the improved estimator is built to decrease the statistical error only. Circles joined by a dotted line are results obtained from the ZVZB improved estimator derived in the preceding section, Eq. (32):

$$\widetilde{F}_{\text{VMC-ZVZB}}[\psi_T, \widetilde{\psi}_{\min}] = F + \frac{(H - E_L)\widetilde{\psi}_{\min}}{\psi_T}$$

$$+ 2(E_L - \langle E_L \rangle)\frac{\widetilde{\psi}_{\min}}{\psi_T}. \tag{73}$$

Note the use of the subscript ZVZB to emphasize the two aspects: reduction of statistical and systematic errors. Finally, the solid line represents the "exact" nonrelativistic force curve for Li_2.

A first important observation is that using improved estimators is extremely efficient in reducing the statistical error. This can be seen by comparing the magnitude of the error bars on data obtained from the ordinary bare estimator (closed squares at $R = 5$, 6.5, and 7.5) with those corresponding to other calculations based on improved estimators. A reduction of at least two orders of magnitude is observed. As already discussed this remarkable result is a direct consequence of the fact that the infinite variance of the bare estimator has been reduced to a finite value. At the scale of the figure error bars associated with improved estimators are almost not visible. A more quantitative analysis will be given later (see Table II).

Now, regarding systematic errors, results are much more disappointing. Using the pure zero-variance (ZV) renormalized estimator, Eq. (72), the behavior of the average force (open squares joined by the dashed line in Fig. 4) as a function of R appears erratic. This can be easily understood since the term added to the bare force in Eq. (72) has a zero average and, therefore, the erratic behavior is a direct consequence of the irregular VMC energy curve presented in Fig. 3. When adding the term correcting the average, the results are improved. As seen in the figure the behavior of $\langle \widetilde{F}_{\text{VMC-ZVZB}}[\psi_T, \widetilde{\psi}_{\min}] \rangle$, Eq. (73), as a function of R, is much less irregular, thus illustrating the important role played by the zero-bias additional contribution [third term on the right-hand side of Eq. (73)] to correct the error due to the approximate trial wave function. Despite of that, the resulting curve is far from being satisfactory. To weaken the role played by ψ_T we can think of going beyond VMC calculations. In Fig. 5 we present such calculations for Li_2 using the DMC-ZVZB improved estimator $\widetilde{F}_{\text{DMC-ZVZB}}[\psi_T, \widetilde{\psi}_{\min}]$ written as

$$\widetilde{F}_{\text{DMC-ZVZB}}[\psi_T, \widetilde{\psi}_{\min}] = F + \frac{(H - E_L)\widetilde{\psi}_{\min}}{\psi_T}$$

$$+ 2(E_L - \langle E_L \rangle)\frac{\widetilde{\psi}_{\min}}{\psi_T}, \tag{74}$$
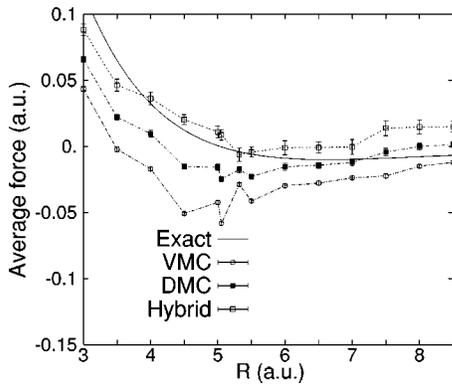
FIG. 5. Li_2 molecule. Average forces using $\widetilde{F}_{ZVZB}(\widetilde{\psi}_T, \widetilde{\psi}_{min})$, Eqs. (73)–(75). VMC average: lowest curve with open circles. DMC average: intermediate curve with closed squares. Hybrid average: highest curve with open squares. Solid line: exact nonrelativistic force curve. Dashed lines between QMC results are a simple linear interpolation to guide the eye.

where the energy average is a fixed-node DMC average, and, also, results obtained by using the generalized hybrid formula, Eq. (57),

$$\overline{F} \simeq 2\langle \widetilde{F}_{DMC\text{-}ZVZB}\rangle_{\psi_T\psi_0} - \langle \widetilde{F}_{VMC\text{-}ZVZB}\rangle_{\psi_T^2}. \tag{75}$$

A clear improvement is observed when going from VMC (open circles) to DMC (closed squares) and, then, to hybrid calculations (open squares): The systematic error present in VMC calculations is reduced. However, the resulting curves are still not satisfactory. Extracting from them a meaningful equilibrium distance or first derivative of the force curve (calculation of ω_e) is impossible. Very similar behaviors have been obtained for H_2 and LiH. They do not need to be reproduced here.

The main reason for the poor results just presented is the low quality of the trial function $\widetilde{\psi}_{min}$ used. According to our general presentation of Sec. II we know that a good trial function $\widetilde{\psi}$ must be close to the derivative of the exact ground-state wave function with respect to R. Here, this is only true when an electron approaches the nucleus A (note that nucleus B has been fixed at the origin and, thus, has no pathological contribution). In that case the nonvanishing part of the exact wave function is expected to behave as

$$\psi_0 \sim_{\mathbf{r}_i \to \mathbf{R}} \exp(-Z_A|\mathbf{r}_i - \mathbf{R}|), \tag{76}$$

which leads to

$$\frac{\partial \psi_0}{\partial R} \sim_{\mathbf{r}_i \to \mathbf{R}} -Z_A \frac{(x_i - R)}{|\mathbf{r}_i - \mathbf{R}|} \psi_0, \tag{77}$$

which is nothing but (up to a minus sign) the minimal form for $\widetilde{\psi}$ given above, Eqs. (69) and (70).

In order to improve our trial wave function $\widetilde{\psi}$ we propose to use the following finite-difference form:

$$\widetilde{\psi}_{deriv} = \frac{\psi_T[R+\Delta R, p(R+\Delta R)] - \psi_T[R, p(R)]}{\Delta R}. \tag{78}$$

In this expression $p(R)$ denotes the complete set of variational parameters entering the trial wave function (coefficients of molecular orbitals, basis set exponents, Jastrow pa-

rameters, etc.). The main advantage of using a finite-difference form instead of the exact derivative is practical: To estimate the derivative we only need to compute two additional local energies and, thus, we avoid deriving and programming the lengthy expressions resulting from the explicit derivative. Note also that using an approximate finite-difference representation is not a problem here: In any case $\widetilde{\psi}_{deriv}$ must be considered as an approximate trial function for the exact derivative and ΔR can always be interpreted as a new additional variational parameter for $\widetilde{\psi}$. In practical calculations, the complete set of parameters we use for minimizing the fluctuations of the various improved estimators consists of $\{p(R), p(R+\Delta R), \text{and } \Delta R\}$.

At the VMC level we consider the following form for the improved estimator

$$\widetilde{F}_{VMC\text{-}ZVZB}[\psi_T, \widetilde{\psi}_{deriv}, \vec{v}]$$

$$= F + \frac{(H - E_L)\widetilde{\psi}_{deriv}}{\psi_T} + 2(E_L - \langle E_L\rangle)\frac{\widetilde{\psi}_{deriv}}{\psi_T}$$

$$+ \frac{\vec{\nabla}[(E_L - \langle E_L\rangle)\psi_T^2\vec{v}]}{\psi_T^2}. \tag{79}$$

At the DMC level the expression used is very similar, see Eq. (60). In this expression the vector field $\vec{v}$ associated with the coordinate transformation is chosen as follows:

$$\vec{v} = \sum_{i=1}^{n_{elec}} e^{-\alpha r_{iA} - \beta r_{iA}^2} \vec{u}_x, \tag{80}$$

where $\vec{u}_x$ is the unit vector along the x axis. The vector field depends on two parameters α and β, which are optimized to lower the variance of $\widetilde{F}_{VMC\text{-}ZVZB}$. The vector field is built so that electrons close to the nucleus A translate with the nucleus, while electrons far away do not move. In practice, we compute the additional term associated with the coordinate transformation using a finite-difference scheme along the direction defined by the vector $\vec{v}$,

$$\frac{\vec{\nabla}[(E_L - \langle E_L\rangle)\psi_T^2\vec{v}]}{\psi_T^2}$$

$$= (E_L - \langle E_L\rangle)\nabla.\vec{v} + \left[(E_L - \langle E_L\rangle)(\vec{x} + \epsilon\vec{v})\frac{\psi_T^2(\vec{x} + \epsilon\vec{v})}{\psi_T^2(\vec{x})}\right.$$

$$\left. - (E_L - \langle E_L\rangle)(\vec{x})\right] \bigg/ \epsilon, \tag{81}$$

where $\vec{x}$ represents the electronic coordinates and ϵ a small positive quantity whose magnitude can also be optimized.

In Fig. 6 we present VMC calculations for Li_2 using $\widetilde{F}_{VMC\text{-}ZVZB}[\psi_T, \widetilde{\psi}_{min}]$ and $\widetilde{F}_{VMC\text{-}ZVZB}[\psi_T, \widetilde{\psi}_{deriv}, \vec{v}]$ as improved estimators. The two estimators have been evaluated on the same Monte Carlo samples. There are two striking differences when using the second estimator $\widetilde{F}_{VMC\text{-}ZVZB}[\psi_T, \widetilde{\psi}_{deriv}, \vec{v}]$. First, the gain in statistical error is spectacular (about one order of magnitude, for a quantitative analysis see the following discussion, Table II). Second, the
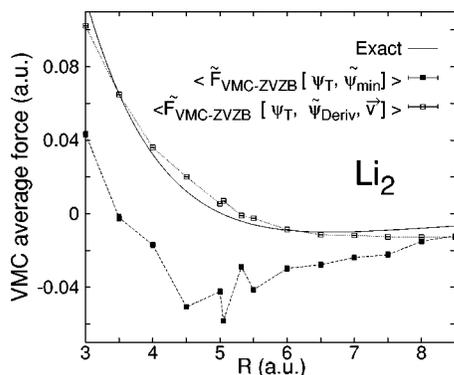
FIG. 6. VMC force for Li_2. Lowest irregular curve with closed squares: $\langle \tilde{F}_{VMC-ZV}[\psi_T, \tilde{\psi}_{min}]\rangle$, Eq. (72). Upper curve with open squares: $\langle \tilde{F}_{VMC-ZV}[\psi_T, \tilde{\psi}_{deriv}, \vec{v}]\rangle$, Eq. (79). Solid line: exact nonrelativistic force curve.

FIG. 8. DMC force for Li_2. Open squares: average fixed-node DMC forces using the Jastrow-RHF one-determinant wave function. Open circles: average fixed-node DMC forces using the Jastrow-VB six-determinant wave function. Solid line: exact nonrelativistic force curve.

curve is much more regular and closer to the exact result (solid line). The VMC results are very satisfactory at small distances (between $R=3$ and $R=4$). However, at larger interatomic distances, the VMC curve begins to separate from the exact one. This is due to the fact that the wave function is built from a RHF calculation and, therefore, the dissociation limit is not correctly described. To address this problem we have considered a more sophisticated trial wave function consisting of a product of a Jastrow factor and a six-determinant one-particle part (for a more detailed description, see above). This VB-type trial wave function has been used only for the largest distances: $R=7$, $R=7.5$, $R=8$, and $R=8.5$. In Figs. 7 and 8 the comparison between results obtained with the Jastrow-RHF (one determinant) and the Jastrow-VB (six determinants) wave functions is presented. At the VMC level (Fig. 7), the improvement resulting from the multideterminant wave function is clearly seen, the forces computed are much closer to the exact curve than in the one-determinant case. At the DMC level (Fig. 8) we could expect that this error disappears even with the Jastrow-RHF (one determinant) wave function since the DMC results depend only on the nodal structure of the wave function. However, it is not true. The difference between the DMC curve and the exact one is still important at large distances

like in the VMC case. This result takes its origin in the approximation made for the exact derivative of the wave function in the DMC estimator, Eq. (53) (ψ_0'/ψ_0 is replaced by ψ_T'/ψ_T). When using the Jastrow-VB wave function the DMC results obtained are much better.

We are now in a position to present our final curves for the three molecules obtained with our best fully optimized estimator $\tilde{F}_{ZVZB}[\psi_T, \tilde{\psi}_{deriv}, \vec{v}]$ and the hybrid formula. Results for the molecules H_2, LiH, and Li_2 are presented in Figs. 9–11, respectively. As seen from these figures the overall agreement between the exact curves (solid lines) and QMC results (open squares) is very good. To be more quantitative we have extracted from these curves an estimate of the spectroscopic constants R_e (equilibrium distance) and ω_e (harmonic frequency). To do that, the data have been fitted with a functional form given by the derivative of a Morse potential curve $E(R) = D[\exp{-2\beta(R-R_e)} - 2\exp{-2\beta(R-R_e)}]$ over some interval of distances around the equilibrium geometry (R between 1.1 and 2 for H_2, between 2.6 and 4 for LiH, and between 4 and 6 for Li_2). Parameters D, β, and R_e have been determined via a generalized least-squares fit. Our results at the VMC, DMC, and hybrid levels are presented in Table I and compared to experimental values.[41] As seen from Table I results for the equilibrium distances are excellent. The largest systematic errors are obtained at the VMC level (relative errors of 4.3%, 3.3%, and 5.7% for

FIG. 7. VMC force for Li_2. Open squares: average VMC forces from estimator (79) using the Jastrow-RHF one-determinant wave function. Open circles: average VMC forces from estimator (79) using the Jastrow-VB six-determinant wave function. Solid line: exact nonrelativistic force curve.

FIG. 9. Hybrid force for LiH. Solid line: exact nonrelativistic force curve.

10548   J. Chem. Phys., Vol. 119, No. 20, 22 November 2003

R. Assaraf and M. Caffarel

FIG. 10. Hybrid force for LiH. Solid line: exact nonrelativistic force curve.

TABLE I. VMC, DMC, and hybrid estimates of the equilibrium geometry R_e (a.u.) and harmonic frequency ω_e (cm^{-1}). The atomic isotopic masses taken[a] are 1.007 825 035 amu for ^{1}H and 7.016 003 0 amu for ^{7}Li.

	H_2	LiH	Li_2
R_e (VMC)	1.463(12)	3.111(17)	5.346(27)
R_e (DMC)	1.426(13)	3.056(6)	5.200(16)
R_e (hybrid)	1.395(15)	3.001(15)	5.068(27)
R_e (Expt.)[b]	1.401	3.015	5.051
ω_e (VMC)	4194(130)	1559(40)	366(9)
ω_e (DMC)	4432(165)	1549(22)	373(5)
ω_e (hybrid)	4662(205)	1519(31)	387(8)
ω_e (Expt.)[b]	4395.2	1405.65	351.4

[a]Reference 42.
[b]Reference 41.

H_2,LiH, and Li_2, respectively). A reduction of a factor of about 2 is gained when DMC calculations are performed. Finally, using the hybrid formula, the exact equilibrium distances are recovered within statistical errors (the relative statistical errors being 1.1%, 0.3%, and 0.5% for H_2,LiH, and Li_2, respectively). In contrast, results for the harmonic frequencies are less accurate but still satisfactory. For the H_2 molecule, the exact experimental result is almost recovered within statistical error at the VMC, DMC, and hybrid levels, the relative statistical error being between 3% and 4%. For LiH and Li_2 the relative statistical errors are of the same order of magnitude. However, a non-negligible systematic error of about 10% is found for these molecules. This result illustrates that obtaining accurate harmonic frequencies is more difficult than obtaining accurate equilibrium geometries.

Now, we would like to present a more quantitative discussion of the performance of the various force estimators introduced in this work. This will allow us to summarize the
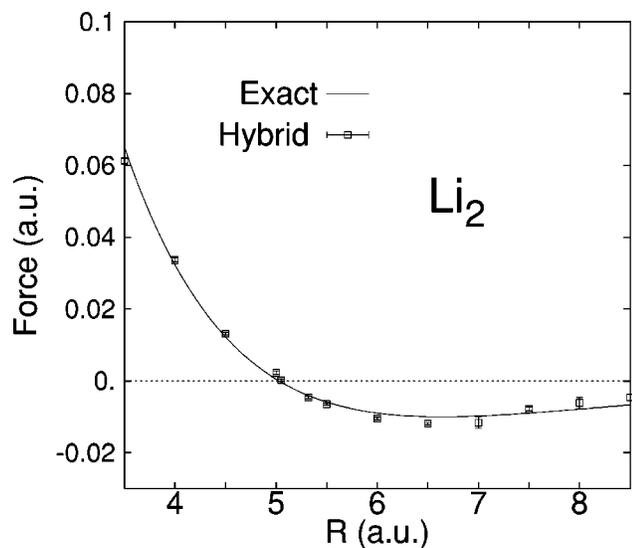
FIG. 11. Hybrid force for Li_2. Solid line: exact nonrelativistic force curve.

various aspects of the method, to present some comparisons with the recent results obtained from the improved estimator implicitly used by Casalegno and collaborators,[35] and also to emphasize some important *quantitative* issues. In Table II the systematic and statistical errors associated with the various force estimators at the VMC, DMC, and hybrid levels of calculation are presented. The results shown are for the Li_2 molecule at the equilibrium geometry ($R = 5.051$ a.u.) where the exact force average, denoted as $\langle F \rangle_{ex}$, is equal to zero. To allow direct comparisons between force estimators all averages have been computed in a common Monte Carlo calculation (identical MC samples). To compare with, we also give the systematic and statistical errors on the total energy. To give a measure of the fluctuations of each estimator the corresponding variances at the VMC level, σ^2(VMC), are reported. To facilitate comparisons between data all averages (except for the VMC variances) are given with five significant figures after the decimal point and all statistical errors are given on the fifth decimal place (magnitude 10^{-5}). The first estimator presented is the bare estimator, F [Eq. (67)]. As already pointed out, this estimator, which has an infinite variance, displays very large statistical fluctuations (between two and three orders of magnitude with respect to the improved estimators to follow) and is, therefore, not at all suitable for practical calculations. The second estimator, $\tilde{F}_{ZV}[\psi_T, \tilde{\psi}_{min}]$ [Eq. (72)], introduced in our previous work on forces[24] is the simplest estimator having a finite variance. However, as explained earlier, when using such an estimator no control on the systematic error exists. The third estimator presented, $\tilde{F}_{ZVZB}[\psi_T, \tilde{\psi}_{min}]$, is the simplest estimator having the ZVZB property. We can see that the introduction of the contribution associated with the ZB property, $2(E_L - \langle E_L \rangle)(\tilde{\psi}_{min}/\psi_T)$ is efficient in reducing the bias (the DMC and hybrid errors are roughly divided by a factor 2). However, as already discussed, the derivative of the trial wave function is not correctly reproduced as a function of the interatomic distance and the corresponding force curve is not smooth (see, Figs. 4 and 5). To get accurate and well-behaved (as a function of R) values of the force it is important to introduce an auxiliary function close to the exact derivative of the wave function. The most simple estimator based on this idea and having a finite variance can be constructed by using the minimal form $\tilde{\psi}_{min}$ [Eqs. (69) and (70)]

TABLE II. VMC, DMC, and hybrid systematic (bias) and statistical errors for the total energy and various force estimators for Li_2 at $R=5.051$ a.u. The VMC variances, σ^2 (VMC), are also given. $\langle E_L \rangle_{ex}$ and $\langle F \rangle_{ex}$ denote the exact total energy, $\langle E_L \rangle_{ex} = -14.9954$ a.u. and exact force $\langle F \rangle_{ex} = 0$ (equilibrium geometry), respectively. To facilitate comparisons between energy and force results all averages are given with five significant figures after the decimal point and statistical errors are given on the fifth decimal place (magnitude 10^{-5}). Statistical errors on VMC variances are on the last digit.

Estimator	VMC average	DMC average	Hybrid	σ^2 (VMC)
$E_L - \langle E_L \rangle_{ex}$	0.038 71(32)	0.005 31(50)	$\cdots$	0.113(5)
$F - \langle F \rangle_{ex}$ [a]	0.182 17(23216)	0.154 62(12293)	0.127 07(33185)	$+\infty$
$\widetilde{F}_{ZV}[\psi_T, \widetilde{\psi}_{min}] - \langle F \rangle_{ex}$ [b]	$-0.063\,52(84)$	$-0.040\,03(151)$	$-0.016\,54(313)$	1.27(5)
$\widetilde{F}_{ZVZB}[\psi_T, \widetilde{\psi}_{min}] - \langle F \rangle_{ex}$ [c]	$-0.058\,02(104)$	$-0.024\,84(184)$	0.008 34(382)	1.3(2)
$\widetilde{F}[\psi_T, \widetilde{\psi}_{min} | \widetilde{\psi}_{deriv}] - \langle F \rangle_{ex}$ [d]	0.006 19(109)	0.029 93(187)	0.053 67(390)	2.8(1)
$\widetilde{F}_{ZVZB}[\psi_T, \widetilde{\psi}_{deriv}, \vec{v}=0] - \langle F \rangle_{ex}$ [e]	0.008 71(218)	0.004 74(147)	0.000 77(366)	14(3)
$\widetilde{F}_{ZVZB}[\psi_T, \widetilde{\psi}_{deriv}, \vec{v}] - \langle F \rangle_{ex}$ [e]	0.006 92(9)	0.003 58(19)	0.000 24(39)	0.016(1)

[a]Equation (67).   [d]Equation (82).
[b]Equation (72).   [e]Equation (79).
[c]Equation (73).

for the zero-variance part and $\widetilde{\psi}_{deriv}$ [Eq. (78)], for the zero-bias part. Such an estimator is written as

$$\widetilde{F}[\psi_T, \widetilde{\psi}_{min} | \widetilde{\psi}_{deriv}] \equiv F + \frac{(H - E_L)\widetilde{\psi}_{min}}{\psi_T}$$

$$+ 2(E_L - \langle E_L \rangle)\frac{\widetilde{\psi}_{deriv}}{\psi_T}. \qquad (82)$$

Written with our notations this is in fact the estimator implicitly used by Casalegno and collaborators in their very recent work.[35] In contrast with the other estimators presented here this "mixed" estimator (different trial functions $\widetilde{\psi}$ are used for the ZV and ZB parts) has no ZVZB property. Accordingly, our general optimization procedure based on the minimization of the improved-estimator variance is no longer meaningful here. Nevertheless, as pointed out by Casalegno et al., to optimize estimator (82) we still have the possibility of optimizing the parameters of the trial wave function ψ_T via energy minimization. Such a procedure is justified because fully optimized trial wave functions are known to verify the Hellmann–Feynman theorem. Statistical errors associated with this estimator are reasonable and roughly similar to those obtained with the two previous estimators. Systematic errors are also comparable. However, in contrast with all ZVZB-improved-estimators introduced in this work, the DMC (and hybrid) calculations do not improve the results and, as seen in the table the hybrid results can even be bad, despite the fact that VMC results are reasonable. Regarding the dependence of the results on the interatomic distance we have been able to recover a relatively smooth force curve for the smallest molecules H_2 and LiH but not for Li_2 (results not shown here). In this latter case, the systematic error is found to be much too sensitive to the quality of the optimization of the trial wave function to lead to reliable results. The last improved estimator presented in Table II is our best proposal for the force estimator, Eq. (79). We report results with $(\vec{v} \neq 0)$ and without $(\vec{v}=0)$ to enlighten the role of the coordinate-transformation term. As seen, the introduction of the $\vec{v}$ term is extremely efficient in reducing the statistical error. For example, at the VMC level

the statistical error without this term is 218×10^{-5}, while the optimized improved estimator using the $\vec{v}$ term is decreased down to 9×10^{-5}. The reduction gained in statistical error is more than one order of magnitude. This remarkable result is general: It is valid for all molecules and all distances treated here. Another most important point is that our best improved estimator (79) is the only estimator presented in this work whose statistical error (here, 9×10^{-5}) is (much) smaller than the energy one (here, 32×10^{-5}). Note that it is also true for the systematic error (whatever the level of calculation). Such a result is particularly important since it has been found that a precise control of the magnitude of the systematic error through variance minimization of the improved force estimators is possible only when such a condition is verified. In contrast, when the statistical error on the force is larger than the energy error, the variance minimization can lead to various results and to get a smooth force curve is very difficult. Actually, we would like to emphasize that obtaining results of the quality presented in Fig. 11 for the Li_2 molecule has only been possible with the improved estimator (79). Using other estimators we have not been capable of constructing a reasonable force curve (smooth and accurate) for this molecule.

Finally, let us say a word about the dependence of our results on the optimization process (determination of the optimal parameters entering $\widetilde{\psi}$, ψ_T, and $\vec{v}$ by minimization of the variance of the improved estimator). Clearly, the method presented in this work is useful only if the results obtained do not depend too much on the way the optimization is performed and on which particular minimum has been found for the variance (as already emphasized, when a large number of parameters are considered the location of such a minimum can depend very crucially on the initial conditions and/or on the random numbers series used). To quantify this aspect we have made nine independent optimizations over nine independent sets of 2000 walkers for the Li_2 molecule at the equilibrium geometry with our best estimator. Results show that the VMC average force results may vary in a significant way for the different sets of optimized parameters found. In this case, the domain of variation is about twenty times the

magnitude of the statistical error. The domain of variation of the DMC average is also important but two times smaller than in the VMC case. However, it has been observed that the hybrid averages are much less sensitive on the optimized parameters. The error on the hybrid average force due to an incomplete optimization has been found to be of the order of magnitude of the statistical error which is rather small.

V. SUMMARY AND CONCLUSIONS

In this work we have shown how to construct improved VMC or DMC estimators for observables. By improved it is meant that, compared to the standard bare estimator O, the new estimators $\tilde{O}$ have a lower variance and a reduced systematic error when averaged over the approximate VMC or (fixed-node) DMC probability densities.

At the variational Monte Carlo level the most general form we propose for $\tilde{O}$ is given by

$$\tilde{O}_{\text{VMC}}[\psi_T,\tilde{\psi},\vec{v}] \equiv O + \frac{(H-E_L)\tilde{\psi}}{\psi_T} + 2(E_L - \langle E_L \rangle)\frac{\tilde{\psi}}{\psi_T}$$
$$+ \frac{\vec{\nabla}[(E_L - \langle E_L \rangle)\psi_T^2\vec{v}]}{\psi_T^2}, \tag{83}$$

where averages are defined over the VMC distribution. At the diffusion Monte Carlo level the expression proposed is essentially similar, except that the average of the local energy entering the definition of $\tilde{O}_{\text{DMC}}$ is defined over the DMC distribution, Eq. (60). The various terms defining the improved observables have a well-defined physical origin: The first three contributions result from the change of the energy average when the magnitude of the observable considered as an external field is varied, while the last contribution comes from the use of a coordinate transformation correlating electron displacements and change of the external field. The functions ψ_T and $\tilde{\psi}$ appearing in the improved observables play the role of trial functions: ψ_T is the ordinary trial wave function for the ground state of H and $\tilde{\psi}$ is a guess for the derivative of the exact ground-state wave function, $\psi_0(\lambda)$, of the perturbed Hamiltonian

$$H(\lambda) \equiv H + \lambda O$$

with respect to λ at $\lambda = 0$. When the trial functions are exact, $(\psi_T,\tilde{\psi}) = (\psi_0, d\psi_0(\lambda)/d\lambda|_{\lambda=0})$, the improved estimator reduces to a constant, namely the exact average for the observable. In that case both statistical and systematic errors vanish. We have called this remarkable property "zero-variance zero-bias property" (ZVZB). In the neighborhood of the exact solution, a local expansion of the various quantities obtained from the approximate guess $(\psi_T,\tilde{\psi})$ can be done. It is found that there is a *quadratic* behavior in the errors $\delta\psi = \psi_T - \psi_0$ and $\delta\psi' = \tilde{\psi} - \psi_0'$. At the VMC level it reads

$$\sigma^2(\tilde{O}) \equiv \langle (\tilde{O} - \langle \tilde{O} \rangle_{\psi_T^2})^2 \rangle_{\psi_T^2} \sim O[\delta\psi\delta\psi'] \tag{84}$$

and

$$\Delta_{\tilde{O}} \equiv \langle \tilde{O} \rangle_{\psi_T^2} - \langle O \rangle_{\psi_0^2} \sim O[\delta\psi\delta\psi'], \tag{85}$$

with a similar result in the DMC case. This important result generalizes the well-known quadratic zero-variance zero-bias property of the energy where the local energy, $E_L = H\psi_T/\psi_T$, plays the role of the improved estimator:

$$\sigma^2(E_L) \sim O[\delta\psi^2] \tag{86}$$

and

$$\Delta_{E_L} \sim O[\delta\psi^2]. \tag{87}$$

In the case of the energy we can write a zero-variance zero-bias equation defining the optimal trial wave function by imposing that the local energy reduces to a constant, namely the exact energy

$$\frac{H\psi_T}{\psi_T} = E_0. \tag{88}$$

Of course, this equation is nothing but the Schrödinger equation. Here, the zero-variance zero-bias equation for the observable is obtained by imposing that the improved observable $\tilde{O}$ reduces to the exact average

$$\tilde{O} = \langle O \rangle_{\psi_0^2}. \tag{89}$$

By optimizing the three quantities $(\psi_T,\tilde{\psi},\vec{v})$ so that fluctuations of $\tilde{O}$ are minimal we can obtain the optimal improved estimator for the observable. In practice, it is done in two steps. First, functional forms for the trial functions ψ_T and $\tilde{\psi}$ are chosen in order to reproduce the best as possible the exact solution of the zero-variance equations. The choice of the vector field $\vec{v}$ is done on physical grounds: It corresponds to the electron-coordinate transformation, $\vec{y} = \vec{x} + \lambda\vec{v}(\vec{x}) + O(\lambda^2)$, correlating as much as possible the electron displacements and the change of density associated with the external field defined by the bare observable. Second, the various parameters entering the three quantities are optimized by minimizing the fluctuations of $\tilde{O}$ over a large but finite number of configurations (typically, several thousand) drawn according to the VMC or DMC distributions.

It is important to emphasize that by using the improved estimators presented here it is possible to get an accuracy on expectation values of observables which is comparable to the very good one obtained for total energies. As can be seen from Eqs. (84)–(87) this is true when we are able to reduce the error on the derivative of the wave function at the level of the error on the wave function itself, that is $\delta\psi \sim \delta\psi'$.

Another fact worth pointing out is that there is not an unique way of constructing improved estimators. Here, we have built our estimators by considering the derivative of the variational, Eq. (18), or the exact DMC energy average, Eq. (51). We have also considered the possibility of making a coordinate transformation before making the derivative, Eqs. (41) and (43). Of course, we can think of many other choices and/or transformations. Ultimately, the better strategy will depend very much on the specific problem considered.

J. Chem. Phys., Vol. 119, No. 20, 22 November 2003

Observables in QMC   10551

Finally, in order to go beyond VMC or DMC calculations, we have shown that the reduction of error of one order in $\delta\psi$ associated with the popular "hybrid"(or "second-order") formula mixing DMC and VMC averages can be generalized to the case of our improved estimators:

$$2\langle \tilde{O}_{\text{DMC}}\rangle_{\psi_T\psi_0} - \langle \tilde{O}_{\text{VMC}}\rangle_{\psi_T^2} \sim \langle \psi_0|O|\psi_0\rangle. \tag{90}$$

As an important application we have applied our formalism to the case of the computation of forces for some simple diatomic molecules. In our preceding work on forces[24] we have focused our attention only on the zero-variance part of the problem. More precisely, we have employed (i) a simplified version of the renormalized force, Eq. (72), (ii) the minimal expression for $\tilde{\psi}$ leading to a finite variance, Eqs. (69) and (70), and (iii) the hybrid formula mixing VMC and DMC calculations, Eq. (90). Results obtained for the vanishing force at the equilibrium distance for a number of small diatomic molecules were reasonably good. Here, we have illustrated that such a strategy is in fact not valid for describing the global shape of the force curve. It has been shown that results depend very much on the trial wave function used and, particularly, on the quality of the optimization process of the numerous parameters of the trial wave function. As a result, the force curves obtained are not regular as a function of the interatomic distance and important spectroscopic quantities such as the equilibrium distance R_e and the harmonic frequency ω_e cannot be obtained reliably. To get accurate curves we need not only to have a small amount of statistical fluctuations but also a control of the systematic error. By exploiting the general ZVZB principle presented in this work it has been shown that obtaining accurate curves is now possible. The basic ingredients are: (i) the use of a trial wave function for the derivative, $\tilde{\psi}$, built as a finite difference of the trial wave function with respect to the nuclear coordinate, (ii) the use of a coordinate transformation in the spirit of the works by Umrigar[28] and Filippi and Umrigar[33] and, finally, (iii) the systematic minimization of the variance of the improved estimator with respect to all the parameters entering the two trial functions ($\psi_T, \tilde{\psi}$), and the vector field, $\vec{v}$, associated with the coordinate transformation. Let us emphasize that to get a well-balanced optimization of the two trial functions (leading to smooth curves for the forces), it is essential to reduce the variance of the improved estimator for the force at the level of the variance of the local energy. Although such a condition may appear as very difficult to fulfill (local energies have usually very small variances), we have shown that it is in fact possible thanks to the coordinate-transformation term. Such a result is remarkable and is certainly one of most important practical aspects of the approach proposed in this work.

Finally, let us remark that the price to pay with respect to the minimal scheme presented in our previous work[24] lies in the need of computing about $3N_{\text{nucl}}$ local energies to calculate the various components of the force. However, we do not think it represents a major difficulty for the realistic applications to come. Indeed, the few-percent accuracy needed on average forces will be obtained with relatively small statistics and, therefore, it will not be necessary to compute the force vector at each Monte Carlo step (the expensive $3N_{\text{nucl}}$-local energy-calculation step will be done rarely). In addition to this, in applications where the nuclear geometry is varied during the simulation (molecular dynamics-type applications) it should also be possible to use suitable re-actualization schemes to avoid recomputing entirely the $3N_{\text{nucl}}$ local energies for close nuclear configurations. Of course, the validity of these various strategies as well as the quality of the improved estimators presented in this work need now to be checked for realistic applications involving many nuclear degrees of freedom.

ACKNOWLEDGMENTS

[1] A. J. Williamson, J. C. Grossman, R. Q. Hood, A. Puzder, and G. Galli, Phys. Rev. Lett. **89**, 196803 (2002).

[2] A. Puzder, A. J. Williamson, J. C. Grossman, and G. Galli, J. Chem. Phys. **117**, 6721 (2002).

[3] A. Puzder, A. J. Williamson, J. C. Grossman, and G. Galli, Phys. Rev. Lett. **88**, 097401 (2002).

[4] G. Belomoin, E. Rogozhina, J. Therrien, P. V. Braun, L. Abuhassan, M. H. Nayfeh, L. Wagner, and L. Mitas, Phys. Rev. B **65**, 193406 (2002).

[5] C. Filippi, S. B. Healy, P. Kratzer, E. Pehlke, and M. Scheffler, Phys. Rev. Lett. **89**, 166102 (2002).

[6] A. R. Porter, M. D. Towler, and R. J. Needs, Phys. Rev. B **64**, 035320 (2001).

[7] L. Mitas, J. Therrien, R. Twesten, G. Belomoin, and M. H. Nayfeh, Appl. Phys. Lett. **78**, 1918 (2001).

[8] I. V. Ovcharenko, W. A. Lester, Jr., C. Xiao, and F. Hagelberg, J. Chem. Phys. **114**, 9028 (2001).

[9] A. J. Williamson, R. Q. Hood, and J. C. Grossman, Phys. Rev. Lett. **87**, 246406 (2001).

[10] S. B. Healy, C. Filippi, P. Kratzer, E. Penev, and M. Scheffler, Phys. Rev. Lett. **87**, 016105 (2001).

[11] P. R. C. Kent, M. D. Towler, R. J. Needs, and G. Rajagopal, Phys. Rev. B **62**, 15394 (2000).

[12] T. Torelli and L. Mitas, Phys. Rev. Lett. **85**, 1702 (2000).

[13] L. Mitas, J. C. Grossman, I. Stich, and J. Tobik, Phys. Rev. Lett. **84**, 1479 (2000).

[14] Y. Shlyakhter, S. Sokolova, A. Lüchow, and J. B. Anderson, J. Chem. Phys. **110**, 10725 (1999).

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

[16] D. M. Ceperley and M. H. Kalos, in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer, Berlin, 1979), pp. 145–194.

[17] K. S. Liu, M. H. Kalos, and G. V. Chester, Phys. Rev. A **10**, 303 (1974).

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

[19] J. Casulleras and J. Boronat, Phys. Rev. B **52**, 3654 (1995).

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

[21] R. N. Barnett, P. J. Reynolds, and W. A. Lester, Jr., J. Comput. Phys. **96**, 259 (1991).

[22] R. N. Barnett, P. J. Reynolds, and W. A. Lester, Jr., J. Chem. Phys. **96**, 2141 (1992).

[23] R. Assaraf and M. Caffarel, Phys. Rev. Lett. **83**, 4682 (1999).

[24] R. Assaraf and M. Caffarel, J. Chem. Phys. **113**, 4028 (2000).

[25] B. H. Wells, Chem. Phys. Lett. **115**, 89 (1985).

[26] P. J. Reynolds, R. N. Barnett, B. L. Hammond, R. M. Grimes, and W. A. Lester, Jr., Int. J. Quantum Chem. **29**, 589 (1986).

[27] C. A. Traynor and J. B. Anderson, Chem. Phys. Lett. **147**, 389 (1988).

[28] C. J. Umrigar, Int. J. Quantum Chem. **23**, 217 (1989).

[29] Z. Sun, W. A. Lester, Jr., and B. L. Hammond, J. Chem. Phys. **97**, 7585 (1992).

[30] J. Vrbik, D. A. Lagare, and S. M. Rothstein, J. Chem. Phys. **92**, 1221 (1990).

[31] J. Vrbik and S. M. Rothstein, J. Chem. Phys. **96**, 2071 (1992).

[32] P. Belohorec, S. M. Rothstein, and J. Vrbik, J. Chem. Phys. **98**, 6401 (1993).
[33] C. Filippi and C. J. Umrigar, Phys. Rev. B **61**, R16291 (2000).
[34] R. Baer, J. Chem. Phys. **113**, 473 (2000).
[35] M. Casalegno, M. Mella, and A. M. Rappe, J. Chem. Phys. **118**, 7193 (2003).
[36] P. Pulay, Mol. Phys. **17**, 197 (1969).
[37] P. J. Reynolds, R. N. Barnett, B. L. Hammond, and W. A. Lester, J. Stat. Phys. **43**, 1017 (1986).

[38] E. L. Pollock and D. M. Ceperley, Phys. Rev. B **30**, 2555 (1984).
[39] M. Caffarel and P. Claverie, J. Stat. Phys. **43**, 797 (1986).
[40] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988).
[41] K. P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure IV, Constants of Diatomic Molecules* (van Nostrand Reinhold, New York, 1979).
[42] A. H. Wapstra and G. Audi, Nucl. Phys. A **432**, 1 (1985).