Chaotic versus Nonchaotic Stochastic Dynamics in Monte Carlo Simulations: A Route for Accurate Energy Differences in N-Body Systems

Roland Assaraf,[1] Michel Caffarel,[2] and A. C. Kollias[3]

[1]*Laboratoire de Chimie Théorique, CNRS-UMR 7616 et Université Pierre et Marie Curie, 75252 Paris Cedex, France*
[2]*Laboratoire de Chimie et Physique Quantiques, CNRS-UMR 5626, IRSAMC et Université de Toulouse,
31062 Toulouse Cedex, France*
[3]*United States Patent and Trademark Office, 600 Dulany Street, Alexandria, Virginia 22314-5796, USA*

We present a method to efficiently evaluate small energy differences of two close N-body systems by employing stochastic processes having a stability versus chaos property. By using the same random noise, energy differences are computed from close trajectories without reweighting procedures. The approach is presented for quantum systems but can be applied to classical N-body systems as well. It is exemplified with diffusion Monte Carlo simulations for long chains of hydrogen atoms and molecules for which it is shown that the long-standing problem of computing energy derivatives is solved.

At the heart of the N-body quantum problem in physics and chemistry is the evaluation of excitation energies corresponding to small energy differences associated with two close Hamiltonians. Evaluating such excitations allows detailed knowledge of both the nature of the ground state and of the low-energy properties of the system. Denoting H as the reference N-body Hamiltonian defined over N degrees of freedom and H_λ some small perturbation of it (including eventually a slight change in N), where λ is a small parameter connecting both Hamiltonians, we are interested in the excitation energies defined as the difference $\Delta \equiv E_0^\lambda - E_0$ in ground-state energies. Depending on the domain of application (nuclear physics, quantum liquids, solid-state physics, quantum chemistry, etc.), such excitations may appear under various names such as, e.g., phonon, plasmon, spinon, or charge excitations, bound states, reaction barriers, electronic affinities, binding energies, etc. We may also be interested in evaluating *infinitesimal* energy differences such as in the important case of computing properties other than the energy via the Hellman-Feynman theorem (energy derivative) or in the case of the thermodynamic limit where energy differences (for example, a one-particle gap) usually scale at most as a tiny fraction of order $1/N$ of the total energy. Although total energies are usually computed with high accuracy in quantum Monte Carlo (QMC) calculations, it is in most cases not sufficient to compensate for the tremendous loss of relative accuracy when small differences are computed by using *independent* calculations of each energy component. Today, a number of approaches have been devised to tackle the problem of computing small differences and derivatives of energies, e.g., [1–5]. The most popular strategy employed up to now is to introduce a scheme based on the idea of correlated sampling [1,6–8]. This idea has been implemented by introducing common stochastic dynamics for the two close Hamiltonians H_λ and H and by taking

into account their difference via modified estimators including weighting factors related to the difference, i.e., reweighting techniques. Here, it is emphasized that introducing such weights is problematic since they are responsible for the occurrence of large statistical or in some case infinite fluctuations; see, for example, [3]. These fluctuations not only completely or partially cancel the benefit of the correlation but also introduce systematic errors, in particular, for the important case of fermionic systems where the change of nodal topology is difficult to take into account [1,2,9].

The purpose of this Letter is to present a method which circumvents such difficulties. In short, we rely on a stability property versus chaos which states that stochastic trajectories having different initial conditions but a common noise meet exponentially fast in time. We show that because of this property two slightly different processes sharing the same noise will produce two trajectories that will remain close forever. This situation is clearly particularly favorable for computing small energy differences efficiently, since we keep the advantage of the correlation without the drawback of reweighted estimators. Note that the observation that configurations can coalesce after some time in a Monte Carlo scheme when a common set of random numbers is used is not new [10–13]. The ideas will be presented here in the case of two popular QMC approaches, namely, the variational Monte Carlo and the fixed-node diffusion Monte Carlo methods. However, we emphasize that such ideas can be extended without difficulties to other types of Monte Carlo methods including *classical* Monte Carlo simulations (e.g., by sampling the Boltzmann density with a continuous Langevin-type process). Finally, we present a few applications of the method to long chains of hydrogen atoms and molecules.

In any Monte Carlo algorithm, one evaluates a property, for example, the energy, as the average of some function

$e(\mathbf{R})$, over a probability distribution $\pi(\mathbf{R})$ (analytically known or sampled):

$$E_0 = \langle e(\mathbf{R}) \rangle_\pi = \int e(\mathbf{R})\pi(\mathbf{R})d\mathbf{R}. \qquad (1)$$

Depending on the variant of the Monte Carlo method used, $\mathbf{R}$ may represent a point in the N-particle space (variational, diffusion, Green's function, projector Monte Carlo, etc.) or a path in a path-integral formulation (path-integral, reptation, pure diffusion Monte Carlo, etc.). The distribution $\pi(\mathbf{R})$ may represent various N-body densities or some positive functional of the path (typically, a Feynman-Kac weight). In all cases the following ground-state energy difference has to be calculated:

$$E_0^\lambda - E_0 = \langle e_\lambda(\mathbf{R}) \rangle_{\pi_\lambda} - \langle e(\mathbf{R}) \rangle_\pi, \qquad (2)$$

where λ is the (small) parameter connecting the two (close) N-body Hamiltonians, π and π_λ the respective ground-state densities, and e and e_λ the energy estimators. In a standard correlated sampling approach, we would use the same distribution for the two calculations:

$$\langle e_\lambda(\mathbf{R}) \rangle_{\pi_\lambda} - \langle e(\mathbf{R}) \rangle_\pi = \frac{\langle e_\lambda(\mathbf{R})\frac{\pi_\lambda}{\pi}(\mathbf{R}) \rangle_\pi}{\langle \frac{\pi_\lambda}{\pi}(\mathbf{R}) \rangle_\pi} - \langle e(\mathbf{R}) \rangle_\pi. \quad (3)$$

Here, to avoid the introduction of the weights $\frac{\pi_\lambda}{\pi}$, we define a stochastic process within the space of pairs $(\mathbf{R}, \mathbf{R}_\lambda)$, where $\mathbf{R}$ and $\mathbf{R}_\lambda$ sample π and π_λ, respectively. For that purpose, two close stochastic processes admitting π and π_λ as steady densities are introduced. They are correlated through the use of a common realization of the stochastic part (same pseudorandom sequence), and the energy difference reads

$$E_0^\lambda - E_0 = \langle e_\lambda(\mathbf{R}_\lambda) - e(\mathbf{R}) \rangle_{\pi(\mathbf{R}, \mathbf{R}_\lambda)}, \qquad (4)$$

where $\pi(\mathbf{R}, \mathbf{R}_\lambda)$ is the density in the space of pairs.

Condition for finite variance on the derivative.—We search to keep the fluctuations of the energy difference of order λ when $\lambda \to 0$. Writing the energy estimator difference inside the brackets in (4) to first order in λ, we get

$$e_\lambda(\mathbf{R}_\lambda) - e(\mathbf{R}) = \lambda e'[\mathbf{R}] + (\mathbf{R}_\lambda - \mathbf{R})\nabla e, \qquad (5)$$

where $e'(\mathbf{R}) = \frac{\partial e_\lambda}{\partial \lambda}|_{\lambda=0}$. It follows that the squared average of the left-hand side of Eq. (5) becomes

$$\langle (e_\lambda[\mathbf{R}_\lambda] - e[\mathbf{R}])^2/\lambda^2 \rangle = \left\langle \left(\frac{\mathbf{R}_\lambda - \mathbf{R}}{\lambda}\nabla e\right)^2 \right\rangle$$
$$+ \left\langle e'^2 + 2e'\frac{\mathbf{R}_\lambda - \mathbf{R}}{\lambda}\nabla e \right\rangle. \quad (6)$$

To see when the average of the squared term can be finite in the limit $\lambda \to 0$, let us first note that the quantities e, e', and ∇e have in general a small variance (for zero-temperature QMC methods, we even have a zero-variance property). Assuming that these quantities have a finite variance, it is straightforward to apply the Schwartz inequality and check that (6) has a finite variance in the limit $\lambda \to 0$ provided that the following stability property is fulfilled:

$$\lim_{\lambda \to 0} \left\langle \left(\frac{\mathbf{R}_\lambda - \mathbf{R}}{\lambda}\right)^2 \right\rangle_{[\mathbf{R}, \mathbf{R}_\lambda]} \text{finite.} \qquad (7)$$

Equivalence to nonchaotic behavior.—To proceed further we shall prove that the property (7) is equivalent to a stability property with respect to chaos of the stochastic dynamics. The proof is written in the context of an overdamped Langevin stochastic equation but extends to more general stochastic processes. The overdamped Langevin equation is written as

$$\mathbf{R}(t + dt) = \mathbf{R}(t) + \mathbf{b}dt + d\mathbf{W}, \qquad (8)$$

where $d\mathbf{W}$ is the standard Wiener process and $\mathbf{b}$ a drift vector leading to the stationary density π, i.e., $\mathbf{b} = \frac{1}{2}\nabla\ln\pi$. The secondary process has an analogous expression

$$\mathbf{R}_\lambda(t + dt) = \mathbf{R}_\lambda(t) + \mathbf{b}_\lambda dt + d\mathbf{W}_\lambda \qquad (9)$$

with $\mathbf{b}_\lambda$ the new drift vector associated with the density π_λ. The two Wiener processes are chosen to be identical: $d\mathbf{W}_\lambda = d\mathbf{W}$, and thus we have

$$\mathbf{R}_\lambda(t + dt) - \mathbf{R}(t + dt) = (\mathbf{b}_\lambda - \mathbf{b})dt - [\mathbf{R}_\lambda(t) - \mathbf{R}(t)]. \qquad (10)$$

Introducing the following tangent vector:

$$\mathbf{T}(t) \equiv \lim_{\lambda \to 0} \frac{\mathbf{R}_\lambda(t) - \mathbf{R}(t)}{\lambda} \qquad (11)$$

and writing Eq. (10) at first order in λ, one has

$$\frac{d\mathbf{T}}{dt}(t) = \mathcal{H}[\mathbf{R}(t)]\mathbf{T}(t) + \mathbf{b}'(t), \qquad (12)$$

where $\mathcal{H}$ is the Hessian matrix built from $\frac{1}{2}\ln\pi$ (spatial second derivatives of it) and

$$\mathbf{b}' \equiv \frac{\partial \mathbf{b}_\lambda}{\partial \lambda}\bigg|_{\lambda=0}. \qquad (13)$$

Equation (12) is a linear differential equation of first order. For a given stochastic trajectory the formal solution with $\mathbf{T} = 0$ as the initial condition [i.e., $\mathbf{R}_\lambda(t = 0) = \mathbf{R}(t = 0)$] can be written as

$$\mathbf{T}(t) = \int_0^t U(s)\mathbf{b}'(s)ds, \qquad (14)$$

where $U(s)$ is the operator obeying

$$\frac{dU}{dt} = \mathcal{H}[\mathbf{R}(t)]U, \quad \text{with } U(t = 0) = Id. \qquad (15)$$

From a physical point of view, the operator U—which depends on the trajectory—expresses the sensitivity of the walker position at time t with respect to a perturbation of the initial conditions. The expression arising in (7) which can also be written as

$$\langle \mathbf{T}^2 \rangle_{[\mathbf{R}(t), \mathbf{R}_\lambda(t)]} = \lim_{t \to \infty} \frac{1}{t}\int_0^t \mathbf{T}^2(s)ds \qquad (16)$$

is finite if $\mathbf{T}^2(s)$ is bounded as a function of time. From Eq. (14) it is seen that it is sufficient that the norm of the operator $U(t)$ decreases exponentially fast. This property tells us that two trajectories with different initial conditions and the same random noise meet exponentially fast (in the L^2 norm sense). Such a result is referred to as stability versus chaos in the context of dynamical systems or synchronization in the context of stochastic processes.

Variational Monte Carlo calculations for hydrogen chains.—In Fig. 1, the fundamental aspects of our approach are presented for the linear metallic H_{120} molecule with a constant internuclear distance (1.4 a.u.). The variational energy is equal to the average of the local energy defined as $E_{L\lambda} \equiv \frac{H_\lambda \psi_{G_\lambda}}{\psi_{G_\lambda}}$ over the overdamped Langevin process with $\psi_{G_\lambda}^2$ as stationary density. Here, ψ_{G_λ} is a simple Hartree-Fock wave function and λ the magnitude of the small displacement of the first atom of the chain. On the left side of the figure, the variance of the spatial distances between correlated walkers ($\lambda = 0$ and $\neq 0$) as a function of the simulation time is shown. As seen, the fundamental synchronization or stability property [Eq. (7)] is perfectly fulfilled, a remarkable result for such a large system (120 electrons). On the right side, the finite differences of both local energy and displacement as a function of λ are displayed. For small enough values of λ, the finite differences are found to be independent on the parameter. As an important consequence, exact derivatives can be obtained with *finite* fluctuations. Comparison with the standard (reweighted) correlated method [1,2] shows that the statistical error is reduced by a factor proportional to the square root of the number of electrons over standard correlated sampling methods, making the method vastly more efficient at large sizes [14].

Diffusion Monte Carlo simulations.—We now apply the approach to the calculation of the *exact* derivatives by using a diffusion Monte Carlo (DMC) scheme.

Compared to variational Monte Carlo simulations, in DMC an additional weighting factor is introduced to recover the exact energy. This factor can be simulated via a birth-death process as in the standard DMC method [15] or "carried" by the walkers as in the pure DMC (PDMC) method [16] we shall employ here. In the correlated approach E_0 and E_0^λ are computed along the same stochastic trajectory by using identical Wiener processes and guiding functions $\Psi_{G_\lambda} = \Psi_G$. A major drawback of such an approach for computing the energy difference $E_0^\lambda - E_0$ is that the difference of local energies $E_{L\lambda} - E_L$ appearing in the estimator can have large or even infinite fluctuations (e.g., in the case of a force on a nucleus). Also, the same nodes are imposed for the two close systems, thus preventing one from accounting for the nodal dependence of ψ_{G_λ} as a function of λ. All proposals presented so far [1,2,17] thus rely on the quality of an analytical guess for the fixed-node function and its derivative, which introduces an uncontrolled and unsatisfactory approximation. We first consider the problem of computing the internuclear force for the H_2 molecule, a system for which the force curve is known with high accuracy. The ground-state wave function being nodeless, there is no fixed-node approximation and the PDMC energy yields the exact energy regardless the choice of the (positive) wave function Ψ_G. Here, we choose a standard simple form [15]. Upon displacement λ, the wave function Ψ_{G_λ} is chosen to keep the same form with the same parameters, except that the center of atomic orbitals (nuclei) are displaced.

On the left part of Fig. 2, several curves are presented representing the difference between the finite-difference derivative of the PDMC energy, $F(t)$, and the exact force value F_{exact} as a function of the internucleus distance. Each curve corresponds to a different value of the projection time t. At zero projection time, the curve gives the difference between the derivatives of the variational and exact
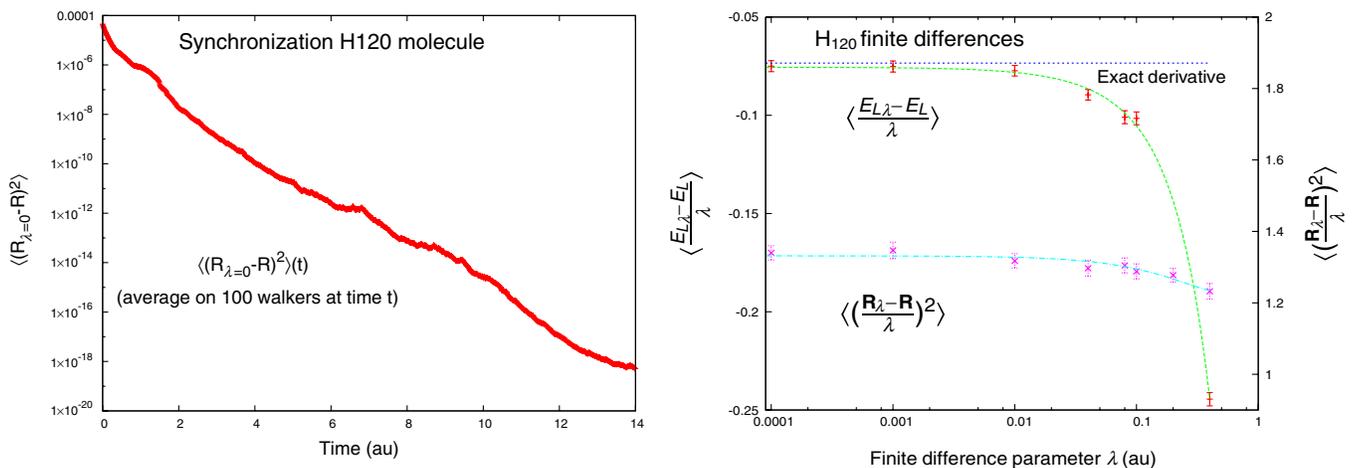
FIG. 1 (color online). H_{120} molecule. On the left, square-averaged distance between correlated walkers having different initial conditions as a function of time. On the right, finite-difference energy and displacement derivatives as a function of λ, displacement of the first atom in the chain.
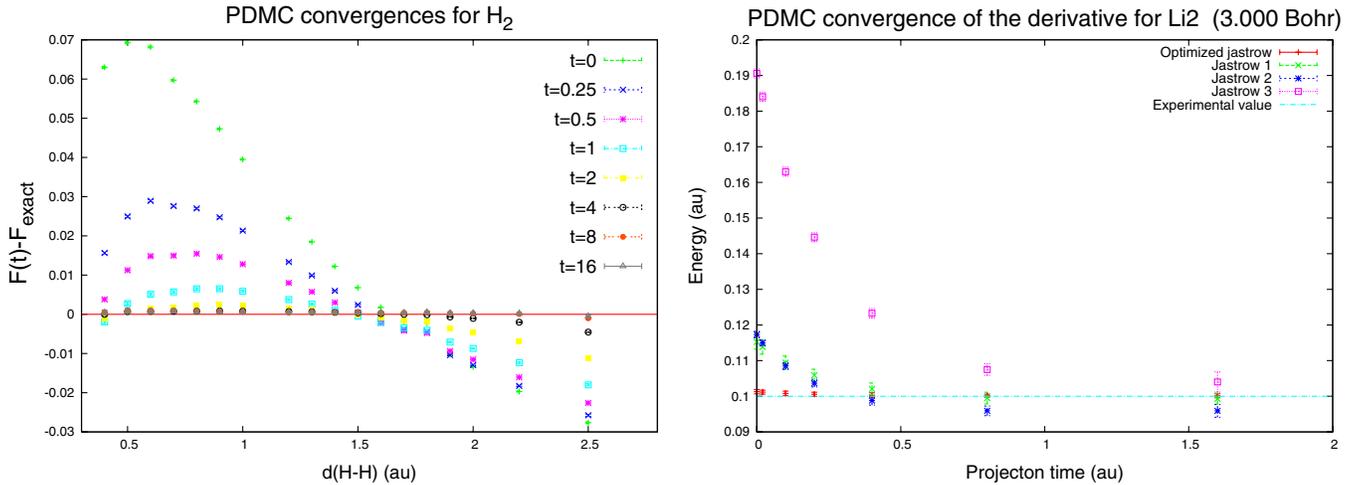
FIG. 2 (color online).   On the left, PDMC energy derivatives for the H_2 molecule as a function of the internucleus distance and computed at different projection times t. On the right, convergence of the derivatives of PDMC energies computed with different Jastrow factors for the Li_2 molecule.

energies. As seen, the curve is far away from being equal to zero, thus illustrating the poor quality of the guiding function used here. However, in spite of this, the exact energy derivative is recovered without difficulties at large projection times. Let us emphasize that, as desired, the result depends neither on the chosen trial wave function nor on any analytical approximation of its derivative. Now, let us consider the important case where the trial wave function displays a nodal pattern depending on the parameter λ (here, change of nuclei positions). We consider the Li_2 molecule described by a standard Jastrow-Slater form. As in the previous cases, the stability condition [Eq. (7)] is found to be perfectly fulfilled. In the right part of Fig. 2, the derivative of the fixed-node PDMC energy as a function of the projection time t is displayed. The different curves correspond to different choices of the parameters of the Jastrow factors. As observed, the long-time behavior of the derivative does not depend on these different choices, thus showing that the derivative of the fixed-node energy can be recovered for this system.

As a last remark, we would like to emphasize that, although the approach presented here has been shown to be particularly efficient, we still have to deal with the problem of the validity of the stability condition which is not a trivial result. It has been found fulfilled for rather long chains of hydrogen atoms with a restricted Hartree-Fock guiding function (up to 120 atoms tested) and for the Li_2 molecule. However, the stability of the "naive" over-damped Langevin process does not hold in general; for example, we have found that applied to larger Li clusters a weak instability may develop, even if it requires such large times to amplify that a transient estimate of the energy differences is still possible in practice. This work motivates further the interest of looking for stable stochastic dynamics in QMC and, more generally, in Monte Carlo methods.

[1] C. Filippi and C. J. Umrigar, Phys. Rev. B **61**, R16 291 (2000).

[2] R. Assaraf and M. Caffarel, J. Chem. Phys. **119**, 10 536 (2003).

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

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

[5] M. W. Lee, M. Mella, and A. M. Rappe, J. Chem. Phys. **122**, 244103 (2005).

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

[7] P. Belohorec, S. M. Rothstein, and J. Vrbik, J. Chem. Phys. **98**, 6401 (1993).

[8] P. Langfelder, S. M. Rothstein, and J. Vrbik, J. Chem. Phys. **107**, 8525 (1997).

[9] A. Badinski, P. D. Haynes, J. R. Trail, and R. J. Needs, J. Phys. Condens. Matter **22**, 074202 (2010).

[10] S. Fahy and D. R. Hamann, Phys. Rev. Lett. **69**, 761 (1992).

[11] J. G. Propp and D. B. Wilson, Random Struct. Algorithms **9**, 223 (1996).

[12] B. Derrida and G. Weisbuch, Europhys. Lett. **4**, 657 (1987).

[13] C. Chanal and W. Krauth, Phys. Rev. Lett. **100**, 060601 (2008).

[14] D. Domin and R. Assaraf (unpublished).

[15] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, Jr., J. Chem. Phys. **77**, 5593 (1982).

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

[17] A. Badinski, P. D. Haynes, and R. J. Needs, Phys. Rev. B **77**, 085111 (2008).