Hellman-Feynman Operator Sampling in Diffusion Monte Carlo Calculations

R. Gaudoin[1] and J. M. Pitarke[2,3]

[1]*Donostia International Physics Center (DIPC), Manuel de Lardizabal Pasealekua, E-20018 Donostia, Basque Country, Spain*
[2]*CIC nanoGUNE Consolider, Mikeletegi Pasealekua 56, E-2009 Donostia, Basque Country, Spain*
[3]*Materia Kondentsatuaren Fisika Saila, UPV/EHU, and Unidad Física Materiales CSIC-UPV/EHU, 644 Posta kutxatila, E-48080 Bilbo, Basque Country, Spain*

(Received 23 January 2007; published 20 September 2007)

Diffusion Monte Carlo (DMC) calculations typically yield highly accurate results in solid-state and quantum-chemical calculations. However, operators that do not commute with the Hamiltonian are at best sampled correctly up to second order in the error of the underlying trial wave function once simple corrections have been applied. This error is of the same order as that for the energy in variational calculations. Operators that suffer from these problems include potential energies and the density. This Letter presents a new method, based on the Hellman-Feynman theorem, for the correct DMC sampling of all operators diagonal in real space. Our method is easy to implement in any standard DMC code.

DOI: 10.1103/PhysRevLett.99.126406 PACS numbers: 71.10.Ca, 71.15.−m

Diffusion Monte Carlo (DMC) calculations are widely used for the computation of properties of solids and molecules [1]. Frequently, it is used as a check on other methods [2] or even as an input [3]. It is therefore very important that DMC calculations be as accurate as possible. However, other than for the total energy, standard DMC calculations are not as definitive as one would hope, since operators that do not commute with the Hamiltonian cannot be sampled exactly within standard DMC calculations. Here we present a simple yet effective addition to standard DMC calculations that plugs that gap and is easy to implement.

DMC calculations by construction yield the normalized expectation value $\langle\hat{O}\rangle_{\mathrm{DMC}} = \langle\Psi_T|\hat{O}|\Psi_0^{\mathrm{fn}}\rangle/\langle\Psi_T|\Psi_0^{\mathrm{fn}}\rangle$, which is generally not the true ground-state expectation value $\langle\hat{O}\rangle = \langle\Psi_0|\hat{O}|\Psi_0\rangle/\langle\Psi_0|\Psi_0\rangle$. In fact, it is not even $\langle\hat{O}\rangle_{\mathrm{fn}} = \langle\Psi_0^{\mathrm{fn}}|\hat{O}|\Psi_0^{\mathrm{fn}}\rangle/\langle\Psi_0^{\mathrm{fn}}|\Psi_0^{\mathrm{fn}}\rangle$, the ground-state expectation value constrained by the nodal structure of the fermionic many-body wave function that is given by Ψ_T. Ψ_T is a trial wave function that approximates the generally unknown ground-state wave function Ψ_0 and is real. In its basic form, Ψ_0^{fn} is the ground state for a fixed nodal structure given by that of Ψ_T. In addition, operators that do not commute with the Hamiltonian are generally subject to a further error, the leading term of which is linear in the difference between Ψ_T and Ψ_0^{fn}. In conjunction with variational Monte Carlo (VMC) calculations, this error can be reduced by one order [4] by using $\langle\hat{O}\rangle_{c\mathrm{DMC}} = 2\langle\hat{O}\rangle_{\mathrm{DMC}} - \langle\hat{O}\rangle_{\mathrm{VMC}}$. Correct sampling can be achieved, e.g., by using forward-walking [5], reptation Monte Carlo [6] calculations, and other methods [7]. Many of these methods aim to sample $\Psi_0^{\mathrm{fn}}\Psi_0^{\mathrm{fn}}$, rather than the usual DMC distribution $\Psi_T\Psi_0^{\mathrm{fn}}$. They are therefore not straightforward additions to the DMC algorithm. Alternatively, the Virial theorem or the related Hellman-Feynman (HF) theorem [8] can be used to evaluate operator expectation values [9] which in the case of DMC calculations, however, involves numerical derivatives of noisy data.

In this Letter, we present a method based on the application of the HF theorem to the DMC algorithm directly. Our method—Hellman-Feynman sampling (HFS)—can be tagged onto the usual sampling of operators with nearly no extra computational overhead. The aim is to maintain the basic DMC algorithm that samples $\Psi_T\Psi_0^{\mathrm{fn}}$. The total energy is evaluated correctly within standard DMC calculations, and crucially operator expectation values can be cast as HF derivatives of the total energy. Keeping in mind that ultimately the DMC algorithm is nothing but a large sum that yields the total energy, we see the HF derivative can be applied without problem to the algorithm itself. One advantage over numerical derivatives is that the resulting formula can handle several operators simultaneously in a single DMC calculation and maintaining orbital occupancy for perturbed Hamiltonians ceases to be a problem. The DMC algorithm only involves numbers, so noncommutability of operators—the source of the difficulties—is no issue, either. Writing down the DMC algorithm as a mathematical formula and applying the HF derivative to it yields an object that when sampled using standard DMC calculations produces the exact operator expectation value. It has to by construction.

In the following, we present a schematic overview of the DMC algorithm, which, however, is sufficient to derive the relevant formulas. The basic idea is to split the imaginary-time propagator $\exp(-\Delta t\hat{H}) \approx \exp(-\Delta t\hat{T})\exp(-\Delta t\hat{V})$ for sufficiently small time intervals Δt into a kinetic and a potential term and then to iterate it. This ultimately [10] gives rise to a real-space drift-diffusion process sampled using the Monte Carlo (MC) method, augmented by an exponential growth term whereby N_w so-called walkers are propagated in parallel. Courtesy of this growth term, at each propagation or (imaginary) time step i the walker j acquires a multiplicative weight: $e^{-\Delta t(E_{i,j}^L - \tilde{E}_i^0)}$, where

$E^L_{i,j} = \hat{H}\Psi_T/\Psi_T$ evaluated at the real-space position of walker j at time step i and $\tilde{E}^0_i$ is an estimate for the ground-state energy also at time step i. The total weight of walker j at time step i becomes

$$\omega_{i,j} = \prod_{k=1}^{i} e^{-\Delta t(E^L_{k,j} - E^0_i)}, \quad \text{where } E^0_i = \frac{1}{i}\sum_{l=1}^{i} \tilde{E}^0_l \tag{1}$$

and the presence of E^0_i ensures normalization. At time step i the estimator for an operator that a DMC code yields is

$$\overline{O^L_i} = \sum_j^{N_w} \omega_{i,j} O^L_{i,j}, \tag{2}$$

where $O^L_{i,j} = \hat{O}\Psi_T/\Psi_T$ and the wave function Ψ_T is evaluated for walker j at time step i. For brevity, we use this bar average $\overline{O^L_i}$ where applicable and note that $\overline{O^L_i}$ has to be averaged over all i to yield the final DMC estimate $\langle\hat{O}\rangle_{\mathrm{DMC}}$. Since the ground-state energy is not known, an estimate chosen such that Eq. (1) remains normalized has to be used. This is the growth estimator E^0_i [11] and is updated at each step, hence the index i. Note that E^0_i is independent of j; i.e., it is the same for every walker and thus a property of the DMC process as a whole. For reasons of numerical stability, DMC implementations often allow walkers to die or multiply such that the walker's survival probability optionally augmented by residual weights corresponds to Eq. (1).

Given a perturbed Hamiltonian $\hat{H}(\alpha) = \hat{H} + \alpha\hat{O}$ and the associated fixed-node ground-state energy $E_{\mathrm{fn}}(\alpha) = \langle\hat{H}\rangle_{\mathrm{DMC}}$, first-order perturbation theory for Ψ^{fn}_0 yields a fixed-node equivalent of the HF theorem [12]

$$\langle O\rangle_{\mathrm{fn}} = \frac{\partial E_{\mathrm{fn}}(\alpha)}{\partial\alpha}\bigg|_{\alpha=0}, \tag{3}$$

where $\langle O\rangle_{\mathrm{fn}}$ converges to the correct ground-state expectation value as the nodes of Ψ_T become exact though Ψ_T itself need not. Note that while $\langle\hat{H}\rangle_{\mathrm{DMC}} = \langle\hat{H}\rangle_{\mathrm{fn}}$ we have $\langle\hat{O}\rangle_{\mathrm{DMC}} \neq \langle\hat{O}\rangle_{\mathrm{fn}}$, unless $[\hat{O}, \hat{H}] = 0$, so Eq. (3) is not trivial. The energy $E_{\mathrm{fn}}(\alpha)$ is accessible exactly within standard DMC calculations as the Hamiltonian $\hat{H}(\alpha)$ commutes with itself. Analytic operator estimators can then be derived by applying the HF theorem to the formula expressing the DMC algorithm Eq. (2). Using Eqs. (1) and (2) the expectation value at time step i becomes

$$E_i(\alpha) = \sum_j^{N_w} E^L_{i,j}(\alpha) \prod_{k=1}^{i} e^{-\Delta t[E^L_{k,j}(\alpha) - E^0_i(\alpha)]}. \tag{4}$$

Here, $E^L_{i,j}(\alpha) = E^L_{i,j} + \alpha O^L_{i,j}$ and $E^0_i(\alpha) = E^0_i + \Delta E^0_i(\alpha)$, so the weight of the wave function is

$$\Omega_i = \sum_j^{N_w} \underbrace{\exp\{-\Delta t \sum_{k=1}^{i}(E^L_{k,j} - E^0_i)\}}_{\omega_{i,j}} \times \exp\{-\Delta t \sum_{k=1}^{i}[\alpha O^L_{k,j} - \Delta E^0_i(\alpha)]\} \tag{5}$$

$$= \overline{\exp(-t\alpha X_i)}\exp[t\Delta E^0_i(\alpha)], \tag{6}$$

where $X_{i,j} = \frac{1}{i}\sum_{k=1}^{i} O^L_{k,j}$ and $t = i\Delta t$, and we have made use of the fact that the growth estimator $\Delta E^0_i(\alpha)$ is independent of the index j. $\Delta E^0_i(\alpha)$ ensures that $\Omega_i(\alpha) = 1$, hence $\Delta E^0_i(\alpha) = -\frac{1}{t}\log[\overline{\exp(-t\alpha X_i)}]$. Equation (4) then becomes

$$E_i(\alpha) = \frac{\overline{E^L_i(\alpha)e^{-t\alpha X_i}}}{\overline{e^{-t\alpha X_i}}}. \tag{7}$$

Evaluating ΔE^0_i to first order gives the growth estimator of an operator:

$$O^{\mathrm{GR}}_i = \frac{\partial E^0_i(\alpha)}{\partial\alpha}\bigg|_{\alpha=0} = \frac{\partial\Delta E^0_i(\alpha)}{\partial\alpha}\bigg|_{\alpha=0} = \overline{X_i} \tag{8}$$

In other words, the DMC sampling of $X_{i,j}$ by virtue of the HF theorem yields a growth estimator of the true expectation value of $\hat{O}$. Interestingly, the growth estimator, if the residual weights are chosen to be zero, appears to be similar to Eq. (13) of Ref. [7]. Applying the HF theorem to $E_i(\alpha)$ in Eq. (7) yields a second estimator

$$O^E_i = \frac{\partial E_i(\alpha)}{\partial\alpha}\bigg|_{\alpha=0} = \overline{O^L_i} - t(\overline{E^L_i X_i} - \overline{E^L_i}\cdot\overline{X_i}). \tag{9}$$

Equations (8) and (9) are of course evaluated at $\alpha = 0$ and are therefore accessible in a regular DMC calculation. We see that for O^E_i the standard estimator $\overline{O^L_i}$ is augmented by a correction term $\Delta O^E_i = -t(\overline{E^L_i X_i} - \overline{E^L_i}\cdot\overline{X_i})$. Several observations can be made. First, in the case of the Ψ_T being the ground state Ψ^{fn}_0 for a given nodal structure, the correction term is zero ($E^L_{i,j}$ is a constant) and only $\overline{O^L_i}$ contributes as it should. Furthermore, the new estimator O^E_i and the usual one $\overline{O^L_i}$ sample an observable and are both independent of the auxiliary DMC parameter t. It follows that $\overline{E^L_i X_i} - \overline{E^L_i}\cdot\overline{X_i} \sim \frac{1}{t}$. Third, since the growth estimator Eq. (8) is derived from the "averaged" quantity E^0_i rather than $\tilde{E}^0_i$, Eq. (8) is itself already averaged over i and therefore the final estimate at i. This is in contrast to Eq. (9) which still has to be averaged over all i to yield the final DMC estimate. Using $\tilde{E}^0_i$ yields an estimator $\tilde{O}^{\mathrm{GR}}_i$ which when averaged over i gives O^{GR}_i. Finally, within the fixed-node approximation the correction term in Eq. (9) can be viewed as a direct measure of the error of the trial wave function with respect to a certain operator. In the remainder of the Letter we will only discuss the direct estimator Eq. (9).

An important question is which operators are admissible and can be sampled using the HF estimator Eq. (9) or, for that matter, the growth estimator Eq. (8)? Looking at the definition of the DMC algorithm one sees that it is based on splitting the Hamiltonian into a kinetic energy kernel that gives rise to the diffusion part of the algorithm and a potential energy term that has to be diagonal in real space. The diffusion part always being the same it follows that $\Delta\hat{H} = \alpha\hat{O}$ has to be diagonal in real space, too. Using, for example, $O^L = T^L = \frac{\hat{T}\Psi_T}{\Psi_T}$ therefore actually corresponds to sampling the real space many-body operator given by

the function T^L, rather than the kinetic energy. The result using Eq. (9) is $\int \frac{\hat{T}\Psi_T}{\Psi_T}[\Psi_0^{\rm fn}]^2 dV$ which in general is not the desired expectation value $\langle \hat{T} \rangle_{\rm fn} = \int \frac{\hat{T}\Psi_0^{\rm fn}}{\Psi_0^{\rm fn}}[\Psi_0^{\rm fn}]^2 dV$. Nevertheless, $\langle \hat{T} \rangle_{\rm fn}$ is accessible within DMC calculations by using $\langle \hat{T} \rangle_{\rm fn} = \langle \hat{H} \rangle_{\rm fn} - \langle \hat{V} \rangle_{\rm fn}$ since the last two quantities can be sampled using standard DMC calculations and HFS, respectively.

In the following, we give a few examples to demonstrate the applicability of HFS. We apply the method to sample (i) the density of helium and (ii) the Ewald energy of a homogeneous electron gas with and without interactions. All data are given in atomic units and we used the CASINO [13] package. The target for the number of walkers was between 200 and 400 and the residual weights were allowed to fluctuate between 0.5 and 2. While we did not perform extensive studies it seems the algorithm works with and without residual weights. The only modification to the code consisted of adding a variable X to each walker, updating X, and applying Eq. (9). Other than that we used the code as-is in a standard setup.

Figure 1 shows the electron density (arbitrary units) of He, as obtained from standard DMC calculations and from our new HFS method. When the well-converged (i.e., $\Delta O_i^E \ll O_i^E$) correlated wave function supplied with CASINO is used both calculations yield essentially the same result (solid line); when an ''incorrect'' trial wave function (which we have chosen to be the same as the ''correct'' one but with the radial term heavily skewed) is used, only our new method (dotted line) recovers the correct density, albeit the noise in the data is larger. Equally, the interaction energy is also recovered (DMC calculations correct wave function: 0.947, incorrect wave function: 0.791, incorrect wave function HFS: 0.958). We have also performed DMC calculations of the hydrogen density, where we systematically deformed the known exact wave function. Suffice it to say, as for He we again see confirmation of our algorithm. An interesting point to add here regards the extent to which the wave function could be skewed. It turns out—rather plausibly—that if the wave function ceases to actually sample certain parts of phase space HFS cannot recover the true form. Nevertheless it seems capable of correcting relatively strong errors in the wave function (viz. the rarely sampled asymptotically decaying part of the wave function in Fig. 1), but the details are clearly a topic for further investigation.

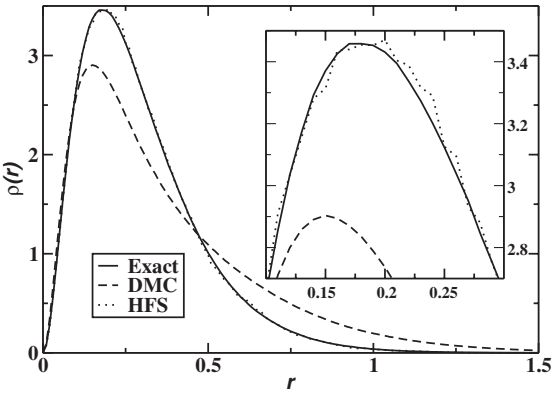

FIG. 1. The ''exact'' helium density (solid line) was derived using the well-optimized wave function provided by the CASINO package: the difference (not shown) between standard DMC sampling and HFS is essentially zero ($\Delta O_i^E \ll O_i^E$). In addition, results using a trial wave function with wrong radial function are presented. Standard DMC calculations yield a smooth but rather poor density. HFS, while noisier (see inset), follows the correct density even in the asymptotic region far from the nucleus where despite little information HFS corrects for the wrong behavior.

As in standard DMC sampling, the worse the trial wave function Ψ_T, the larger the noise when using HFS. However, when looking at the raw data before averaging over i (not shown) we observed that the noise in the HFS data rises during the progression of the sampling; hence, standard error estimation does not work. The source can be traced to sampling over histories X_i. Limiting their depth results in a constant noise term though also reintroduces a systematic bias. Also, in a recent paper [14] Warren and Hinde observe that using the forward-walking method in DMC calculations necessitates a rapidly growing number of walkers as the dimensionality of the quantum system is increased. These two issues then lead us to the question as to whether HFS works for larger systems. We have therefore looked at an unpolarized homogeneous electron gas at $r_s = 1$. We used a finite simulation cell (periodic boundary condition) with 54 electrons. The data we plot show the Ewald interaction energy with no additional finite size corrections. We show in Fig. 2 results for a fully interacting system that we have obtained by using trial wave functions with either no Jastrow factor, a partially optimized Jatrow factor, or a fully optimized one. We show the mixed DMC estimate $\langle \hat{O} \rangle_{\rm DMC}$, the corrected estimate $\langle \hat{O} \rangle_{c\rm DMC} = 2\langle \hat{O} \rangle_{\rm DMC} - \langle \hat{O} \rangle_{\rm VMC}$ which contains a second-order error,

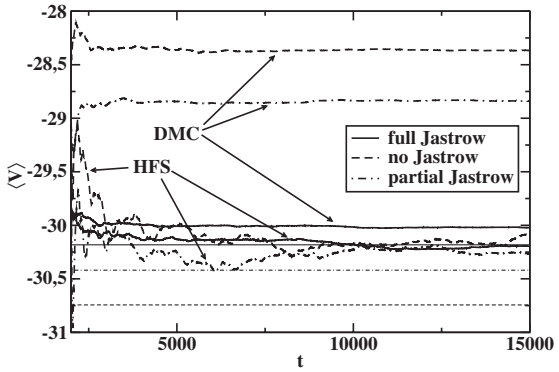

FIG. 2. Results for the Ewald energy of an unpolarized homogeneous electron gas ($r_s = 1$) with 54 electrons. Standard DMC sampling, $\langle \hat{O} \rangle_{\rm DMC}$, yields the relatively smooth curves at the top (see arrows). The noisier data below use HFS (see arrows) and the thin straight lines correspond to $\langle \hat{O} \rangle_{c\rm DMC}$ at the end of the run. The partial Jastrow factor contains a correlation term without cusp.

and the results for HFS. The DMC calculations start at 0 with a short equilibration phase and we start sampling at time step 2000. The corrected estimate using the fully optimized Jastrow factor ought to give the best result. Clearly all three HFS estimates are very close but especially the nonoptimized wave functions yield quite noisy data. Nevertheless, even in that case the results are a lot better than using the standard DMC output for the best wave function. However, they are all also better than the corrected $\langle \hat{O} \rangle_{c\mathrm{DMC}}$ results of the partially or nonoptimized wave function. Regarding the noise one also has to keep in mind the difficulty of the task: the interaction energy is dominated by the region where the electrons get close to each other but that is where the error of the nonoptimized wave functions is largest. HFS essentially has to build a cusp from scratch.

Figure 3 repeats the same analysis for a noninteraction Hamiltonian where the Slater determinant (no Jastrow factor) is the exact solution, whence the HFS data and the standard DMC data in that case being identical. This is of course consistent with Eq. (9) and proves that given the correct nodes, HFS yields the correct answer. Apart from that Fig. 3 is essentially a mirror image of Fig. 2. In general, we see that unless the wave function is well optimized the HFS estimate is considerably better, despite the noise in the data. Such situations might occur when the system is dominated by the bulk while we are interested in sampling data in the surface region. Optimization based on the total energy or variance would result in a suboptimal wave function away from the bulk and hence erroneous standard sampling.

In conclusion, by applying the HF theorem directly to the DMC algorithm we have introduced a new method to sample a large class of operators exactly within standard DMC calculations. Our method works for both small and large systems and is easy to add to standard DMC calculations, enabling the sampling of a large class of operators (densities, interaction energies, etc.): only one extra variable per operator ($X_{i,j}$) needs be added to the walkers, involving no more than an extra summation step during sampling; simple algebra [Eqs. (8) and (9)] does the rest. Future work is needed to better understand, estimate, and deal with the noise and its slow increase with simulation time. This is currently under investigation. Similarly, the effect of residual weights needs to be looked at in more detail. A promising line of research already under way is to look at the second derivative. This might allow efficient DMC sampling of the fixed-node density-response function and related quantities, the study of which is currently not feasible due to being numerically too demanding.

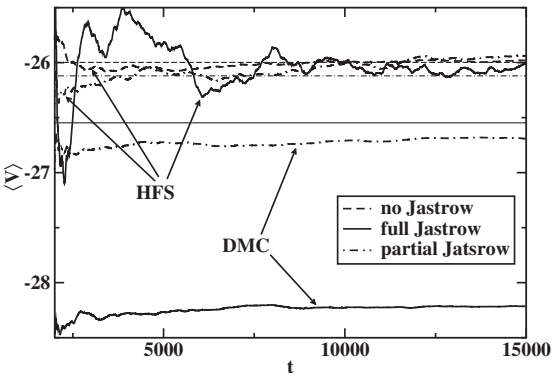

FIG. 3. As in Fig. 2, but for an interaction-free Hamiltonian. The noisier data at the top have been sampled using the HFS estimator (see arrows). Below follow the relatively smooth standard DMC results (see arrows) sampling $\langle \hat{O} \rangle_{\mathrm{DMC}}$, except for the no-Jastrow case where the two estimators yield the same data as $\Delta O_i^E = 0$. The thin straight lines correspond to $\langle \hat{O} \rangle_{c\mathrm{DMC}}$ at the end of the run, except in the case of no Jastrow factor where the thin line gives the essentially exact VMC value.

R. G. would like to thank W. M. C. Foulkes, B. Wood, and N. Hine of Imperial College for helpful discussions. The authors acknowledge partial support by the University of the Basque Country, the Basque Unibertsitate eta Ikerketa Saila, the MCyT, and the EC 6th framework Network of Excellence NANOQUANTA (Grant No. NMP4-CT-2004-500198).

[1] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[2] M. Nekovee and J. M. Pitarke, Comput. Phys. Commun. **137**, 123 (2001).
[3] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980); see also R. M. Dreizler and E. K. U. Gross, *Density-Functional Theory: An Approach to the Quantum Many-Body Problem* (Springer, New York, 1990).
[4] P. A. Whitlock, D. M. Ceperley, G. V. Chester, and M. H. Kalos, Phys. Rev. B **19**, 5598 (1979).
[5] K. S. Liu, M. K. Kalos, and G. V. Chester, Phys. Rev. A **10**, 303 (1974).
[6] S. Baroni and S. Moroni, Phys. Rev. Lett. **82**, 4745 (1999).
[7] J. Casulleras and J. Boronat, Phys. Rev. B **52**, 3654 (1995).
[8] R. P. Feynmann, Phys. Rev. **56**, 340 (1939).
[9] G. Ortiz and P. Ballone, Phys. Rev. B **50**, 1391 (1994).
[10] In practice one uses importance sampling [R. C. Grimm and R. G. Storer, J. Comput. Phys. **7**, 134 (1971)], which changes the kinetic diffusion kernel into drift diffusion and replaces the potential energy by the much smoother local energy $E_L = [\hat{H}\Psi_T]/\Psi_T$. The process then samples $\Psi_T \Psi_0^{\mathrm{fn}}$ rather than just Ψ_0^{fn}.
[11] Inverting $E_i^0 = \frac{1}{i}\sum_{l=1}^{i} \tilde{E}_l^0$ then gives $\tilde{E}_i^0 = iE_i^0 - (i-1)E_{i-1}^0$.
[12] When using the HF theorem in conjunction with numerical derivatives, problems appear if the underlying Hilbert space changes with α. This is not the case here.
[13] R. J. Needs, M. Towler, N. Drummond, and P. Kent, *CASINO Version 1.7 User Manual* (University of Cambridge, Cambridge, England, 2004).
[14] G. L. Warren and R. J. Hinde, Phys. Rev. E **73**, 056706 (2006).