Gamma Distribution Model To Provide a Direct Assessment of the Overall Quality of Quantum Monte Carlo-Generated Electron Distributions†

Braden Coles, Paul Vrbik, Robert D. Giacometti, and Stuart M. Rothstein*

Department of Chemistry, Brock University, St. Catharines, Ontario L2S 3A1 Canada

Received: July 23, 2007; In Final Form: November 19, 2007

Our objective is to assess the accuracy of simulated quantum Monte Carlo electron distributions of atoms and molecules. Our approach is first to model the exact electron distribution by a linear combination of gamma distribution functions, with parameters chosen to exactly reproduce highly accurate literature values for a number of selected moments for the system of interest. In application to the ground-state electron distributions of helium and dihydrogen, a high level of accuracy of the model was confirmed upon comparing its predicted moments, not used in the model's parametrization, to those calculated from high-level theory. Next, we generated electron−electron and electron−nucleus distributions for dihydrogen from electron positions outputted from a variety of quantum Monte Carlo algorithms. Upon juxtaposition of the simulated distributions with the putatively exact one that we derived from the model, we quantified the error in simulated distributions. The most accurate distributions were obtained from no-compromise reptation quantum Monte Carlo, a recently developed algorithm designed to ameliorate the distributions' time-step bias. Marginally less accurate distributions were generated from fixed-node diffusion Monte Carlo with descendant counting and detailed balance.

Introduction

Decades of research devoted to developing quantum Monte Carlo methods have firmly established the advantages of this approach to electronic structure calculations: rapid convergence with basis-set size, favorable scaling with the number of electrons, no calculation and storage of large numbers of integrals, and codes that are naturally suited for parallel computation. Monographs[1,2] and recent reviews, *e.g.*, ref 3, have been devoted to these issues.

Benchmark calculations suggest that the most commonly employed variant, fixed-node diffusion Monte Carlo[4] (FNDMC), estimates the energy with accuracy on par with CCSD(T), although not yet to within chemical accuracy.[5] FNDMC samples the mixed distribution, $\Psi\Phi_0$, where Φ_0 is "exact", except for the incorrect nodes[6] imposed by the importance sampling function, Ψ. This so-called "nodal error" introduces a positive bias in the simulated energy.[7]

In addition to this, there is yet another significant source of error that arises in both FNDMC and the more-recently developed reptation quantum Monte Carlo[8,9] (RQMC) approach: "time-step bias". This error is a result of mathematical moves being made in accordance with the so-called "short-time Green's function", G_0G_b. This quantity is exact only in the limit of zero time step: a small interval of imaginary simulation time.[10−12] Although in practice there are elegant ways to improve the sampling, *e.g.*, refs 13 and 14, and thus reduce the bias, in principle the time step, and consequently the bias associated with its use, cannot be reduced to zero. Alternatively, of course, one may simply reduce the value used for the time step, but this adversely affects the efficiency of the simulation.[15]

Normally, expectation values of operators that do no commute with the Hamiltonian are biased by the inputted importance sampling function. Nevertheless, both FNDMC and variational Monte Carlo (VMC) recover expectation values as if they had been drawn from the "exact" distribution, Φ_0^2, by employing methods of "descendant counting", *e.g.*, refs 16 and 17, and by averaging variationally distributed quantities with accumulated past and future branching factors, *e.g.*, refs 18 and 19, respectively. On the other hand, RQMC algorithms directly sample the "exact" distribution, at the middle of the reptile. Here and above we qualified the word "exact", as all of these distributions suffer from time-step and nodal error.

We are presently concerned with assessing the accuracy of simulated quantum Monte Carlo electron−electron and electron−nucleus distributions. The normal practice, comparing the simulated energy and a selection of other properties with accurate determinations in the literature, is indirect and monitors only isolated regions of the distributions. Instead, our objective is to provide a direct assessment of the overall quality of the simulated distributions. Our approach is first to model the unknown, truly exact electron distributions by linear combinations of gamma distributions. The model is parametrized to reproduce literature results for a broad range of moments, and its efficacy verified by its accurately reproducing literature values for moments not used in its parametrization. Next, curves for quantum Monte Carlo-generated distributions are constructed by spline-fitting histograms of simulated electron positions. Finally, errors in the simulated distributions are revealed upon juxtaposition of the simulated distributions with the putatively exact ones that are derived from the model.

This paper is organized as follows: In the following section we introduce the gamma distribution model and apply it to electron−nucleus and electron−electron distributions of ground-

† Part of the "William A. Lester, Jr., Festschrift".

* Corresponding author. Department of Physics. Phone: 905-688-5550 x3401. Fax: 905-682-9020. E-mail: srothste@brocku.ca.

10.1021/jp075790e CCC: $40.75
Published on Web 02/06/2008

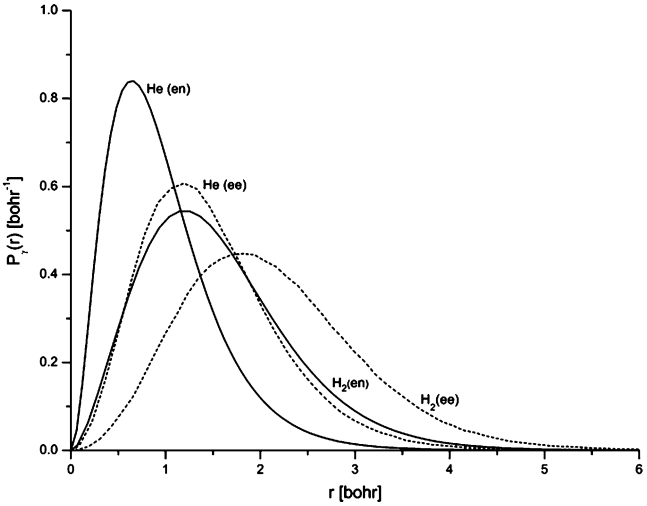

Figure 1. Gamma electron−nucleus (en, solid lines) and electron−electron (ee, dashed lines) distribution model for ground-state helium and dihydrogen.

state helium and dihydrogen. Next, we briefly describe quantum Monte Carlo algorithms whose electron distributions for dihydrogen are assessed by comparison with those of the model. We conclude with a discussion of our results and suggestions for further work.

Gamma Distribution Model for Electron Distributions

A random variable X has a gamma distribution if its probability density function is given by[20]

$$\Gamma(x;\gamma,\beta) = \frac{x^{\gamma-1}\exp(-x/\beta)}{\beta^{\gamma}\Gamma(\gamma)} \qquad \text{for } x > 0, \text{ where } \gamma, \beta > 0 \quad (1)$$

and for any positive integer γ, $\Gamma(\gamma) = (\gamma - 1)!$.

Moments of the gamma distribution are available analytically as

$$\langle x^n \rangle = \beta^n \Gamma(\gamma + n)/\Gamma(\gamma) \qquad n \geq 1 \quad (2)$$

This formula also holds for the first and second inverse moments.

The gamma distribution is useful for describing positively skewed, non-negative data, such as the electron−nucleus and electron−electron probability distributions that concern us here. This is immediately obvious for the electron distribution of 1s hydrogen, where $\gamma = 3$ and scale parameter, β, is equal to 2^{-1} au. Similarly, for the 2p states, $\gamma = 5$ and $\beta = 1.0$ au. The 2s state electron distribution is a linear combination of gamma distributions, with $\gamma = 3$, 4 and 5, and $\beta = 1.0$ au.

By extension, a model for the electron−electron and electron−nucleus distributions for ground states of He and H_2 is as follows:

$$P_{\gamma}(r) = \sum_{k=3}^{7} c_k \Gamma(r;k,\beta) \quad (3)$$

where r is a shorthand notation for the electron−nucleus and electron−electron distance variable; each case is considered separately. The coefficients, $\{c_k\}$, are chosen to provide a *perfect fit* of gamma distribution moments to highly accurate literature values of $\langle r^n \rangle$.

$$\langle r^n \rangle = \sum_{k=3}^{7} c_k \beta^n \Gamma(k+n)/\Gamma(k) \quad (4)$$

where

$$\langle r^0 \rangle = \sum_{k=3}^{7} c_k = 1 \quad (5)$$

For both the electron−nucleus and electron−electron distributions of He, we parametrized the model using literature values for the following moments: $n = -1, 0, ..., 3$. For the electron−nucleus distribution of H_2, $n = -1, 0, ..., 3$, and for the electron−electron distribution, $n = -2, -1, 0, ..., 2$. These choices reflect the availability of suitably accurate determinations in the literature.

For the electron−nucleus distributions, the remaining parameter, β, was set equal to half the inverse of the orbital exponent reported in the literature: ref 21 for He and ref 22, with a slight adjustment, for H_2.

Because there is no "orbital exponent" to draw on for the electron−electron distributions, we selected β to fit the delta function.

$$\delta(\mathbf{r}) = c_3/(8\pi\beta^3) \quad (6)$$

For example of application of this formula, the ground-state hydrogen atom is represented by a single gamma distribution ($c_3 = 1$) with $\beta = 0.5$. Its delta function equals to π^{-1}.

Model parameters obtained in this manner are reported in Table 1. Asterisks in Tables 2 and 3 precede literature values for moments used to parametrize the model for He and H_2, respectively; the model gives an exact fit to these values. The remaining low- and high-order moments appearing in these tables were calculated from the gamma distribution model (eqs 4 and 6). These satisfactorily agree with the literature values, suggesting that we have captured adequately for our purposes the underlying exact, mutually consistent marginal distributions, despite their not having been derived from some underlying parent wave function.

We plotted the electron distributions in Figure 1. Each distribution is unimodal, the dihydrogen distributions (at the equilibrium geometry) are more diffuse than its Helium analogs, as is the electron−electron distribution relative to the electron−nucleus one. For a chemist, these are not surprising. Our model, again from which the distributions were independently derived, reflects these fundamentals quite well.

The model can be extended to account for overlapping electron shells, for example to first row atoms by taking two linear combinations of gamma distributions, each with different scale parameters:[23]

$$F(r) = \sum_{k} c_k \Gamma(k,\beta) + \sum_{k'} c'_k \Gamma(k',\beta') \quad (7)$$

Quantum Monte Carlo Algorithms

Variational Monte Carlo[1] (VMC). In VMC one moves electrons mathematically (not physically) in imaginary time by the processes of drift and diffusion. The size of the moves, performed for a set of configurations (or walkers), depends on the time step, τ, as follows:

TABLE 1: Model Parameters[a] (au) for Electron−Nucleus (en) and Electron−Electron (ee) Distributions of Ground-State Helium and Dihyrodgen

system (distribution)	c_3	c_4	c_5	c_6	c_7	β
He (en)	1.177 350	−0.613 920	0.616 241 3	−0.237 051 5	0.057 380 4	0.296 30
He (ee)	0.096 586 05	0.383 196 69	0.672 770 02	−0.182 801 10	0.030 248 34	0.330 612
H_2 (en)	0.369 871	−0.077 795 0	1.289 321 1	−0.754 359 1	0.172 961 8	0.394
H_2 (ee)	0.021 479 8	0.095 962 4	0.120 431 2	0.515 799 0	0.246 337 7	0.369 525

[a] Equation 3.

TABLE 2: Calculated and Literature Values of Moments for Electron−Nuclear (en) and Electron−Electron (ee) Distributions for Ground-State He Atom[a]

	en (this work[b])	en (literature)	ee (this work[b])	ee (literature)
$\langle\delta(\mathbf{r})\rangle$	1.800 822	1.810 429 29	0.106 345 4	0.106 345 374
$\langle r^{-2}\rangle$	6.011 462	6.017 408 8662	1.464 641	1.467 709 2349
$\langle r^{-1}\rangle$	*	1.688 316 800 717	*	0.945 818 448 800
$\langle r^{1}\rangle$	*	0.929 472 294 874	*	1.422 070 255 566
$\langle r^{2}\rangle$	*	1.193 482 995 021	*	2.516 439 312 836
$\langle r^{3}\rangle$	*	1.967 948 106 70	*	5.308 009 640 84
$\langle r^{4}\rangle$	3.975 588	3.973 564 9319	12.981 729	12.981 271 359

[a] An asterisk means the literature value was used to parametrize the gamma function model, eq 3. The literature values are moments derived from variationally optimized, explicitly correlated wave functions; ref 32. All entries are in atomic units. [b] Equations 4 and 6.

TABLE 3: Calculated and Literature Values of Moments for Electron−Nuclear (en) and Electron−Electron (ee) Distributions for Ground-State Hydrogen Molecule[a]

	en (this work[b])	en (literature)	ee (this work[b])	ee (literature)
$\langle\delta(\mathbf{r})\rangle$	0.240 615	0.2300(2.5)[c]	0.0163	0.016 84(3)[c]
$\langle r^{-2}\rangle$	1.594 09	1.608(5)[d]	*	0.5244(8)[d]
$\langle r^{-1}\rangle$	*	0.911 902[e]	*	0.587 365 91[f]
$\langle r^{1}\rangle$	*	1.548 269[e]	*	2.168 952 8[f]
$\langle r^{2}\rangle$	*	3.037 186[e]	*	5.632 389 5[f]
$\langle r^{3}\rangle$	*	7.175 786[e]	16.93155	16.885(8)[d]
$\langle r^{4}\rangle$	19.8668	19.867 215 5[e]		

[a] An asterisk means the literature value was used to parametrize the gamma function model, eq 3. All entries are in atomic units. [b] Equations 4 and 6. [c] Variational Monte Carlo; ref 33. [d] Variational Monte Carlo; ref 22. [e] Full CI; ref 34. [f] Variational theorem with explicitly correlated wave function; ref 35.

$$\mathbf{R}_{i+1,j} = \mathbf{R}_{i,j} + \tau\nabla\Psi(\mathbf{R}_{i,j})/\Psi(\mathbf{R}_{i,j}) + \sqrt{\tau}\,\chi \tag{8}$$

where $\mathbf{R}_{i,j}$ is the positions of the n electrons for the jth configuration on the ith iteration, ∇ implies partial derivatives with respect to all the coordinates, Ψ is a specified trial function, and χ is a $3n$-dimensional vector with independent components drawn from a $3n$-dimensional standard normal distribution.

These moves simulate the Green's function:

$$G_0(\mathbf{R}_{i,j}\rightarrow\mathbf{R}_{i+1,j},\tau) = (2\pi\tau)^{-3n/2}\exp(-[\mathbf{R}_{i+1,j} - \mathbf{R}_{i,j} - \tau\nabla\Psi(\mathbf{R}_{i,j})/\Psi(\mathbf{R}_{i,j})]^2/2\tau) \tag{9}$$

generating the "variational" distribution of walkers, Ψ^2, albeit with a time-step bias.

Fixed-Node Diffusion Monte Carlo with Physical Branching[4] (FNDMC). To simulate the remaining factor in the short-time Green's function, G_b, during each iteration we either replicate or delete the moved configurations with probability proportional to their so-called "branching factors":

$$b_{i,j} = \mathrm{e}^{-(E^{\mathrm{loc}}(\mathbf{R}_{i,j})-E_{\mathrm{T}})\tau} \tag{10}$$

where E_{T} is a constant, which is approximately equal to the exact energy, and

$$E^{\mathrm{loc}}(\mathbf{R}_{i,j}) = \frac{\hat{H}\Psi(\mathbf{R}_{i,j})}{\Psi(\mathbf{R}_{i,j})} \tag{11}$$

This algorithm samples the mixed distribution, $\Phi_0\Psi$, and is biased both by the incorrect nodes imposed on the distribution by Ψ and by the time step.

By evoking detailed balance (DB), a Metropolis-type decision, in principle, one may ameliorate, but not completely eliminate, the time-step bias. Here one accepts a move given by (8) with probability

$$A(\mathbf{R}_{i,j}\rightarrow\mathbf{R}_{i+1,\mathrm{j}}) = \min(1, W(\mathbf{R}_{i+1,j},\mathbf{R}_{i,j})) \tag{12}$$

where

$$W(\mathbf{R}_{i+1,j},\mathbf{R}_{i,j}) = \frac{|\Psi(\mathbf{R}_{i+1,j})|^2 G_0(\mathbf{R}_{i+1,j}\rightarrow\mathbf{R}_{i,j},\tau)}{|\Psi(\mathbf{R}_{i,j})|^2 G_0(\mathbf{R}_{i,j}\rightarrow\mathbf{R}_{i+1,j},\tau)} \tag{13}$$

Fixed-Node Diffusion Monte Carlo with Descendent Counting[16,17] (FNDMC-DC). To sample the desired "exact" distribution, Φ_0^2, albeit still with time-step and nodal error, one may engage in descendent counting. Here one counts the number of descendents in the far-future for each configuration in the logical recent iteration. This quantity is proportional to Φ_0/Ψ.[24] When it is used as a weight for averaging quantities of interest in the logical recent iteration, one can estimate expected values had they been drawn from the "exact" distribution.

Variational Monte Carlo with Past−Future Weighting[18] (VMC-W). One may perform VMC and accumulate branching factors from distant-past (p) through to far-future (f) iterations, for use as weights when averaging in the logical-present ith iteration. The accumulated branching factors are given by

$$w_{i,j}^{(\mathrm{p,f})} = \prod_{k=i-L+1}^{i} b_{k,j} \prod_{k=i}^{i+L-1} b_{k,j} \tag{14}$$

with L chosen proportional to τ^{-1}.

This accumulation of branching factors is proportional to Φ_0^2/Ψ^2. Performing a weighted-average of variationally distributed quantities in the present iteration estimates expectation values had they been drawn from the exact distribution, Φ_0^2. This algorithm may be employed with detailed balance (12).

VMC-W is akin to appending future-walking to the so-called "pure diffusion Monte Carlo" algorithm,[25] with the advantage of allowing one to use a limited number of weights to control

any numerical instabilities that may arise when crude trial functions are employed.[26]

Reptation Quantum Monte Carlo[8,9] (RQMC). One may again perform VMC but now employ accumulated past- and future-branching factors in a Metropolis-type decision to accept or reject moves taken in the distant past or far future. This is the basis of the so-called "reptation quantum Monte Carlo" algorithm.

A set of consecutive VMC iterations, *e.g.*, N of them, describes a path through imaginary simulation time. This defines "reptile" X_j. One accumulates the local energies as follows:

$$S(X_j) = \tau\left(\frac{1}{2}E^{\mathrm{loc}}(\mathbf{R}_{1,j}) + \sum_{i=2}^{N-1} E^{\mathrm{loc}}(\mathbf{R}_{\mathrm{i,j}}) + \frac{1}{2}E^{\mathrm{loc}}(\mathbf{R}_{N,j})\right) \qquad (15)$$

With probability $^1/_2$, a new reptile is generated by removing M iterations from the original reptile's head and adding M new ones to its tail. Alternatively, again with probability $^1/_2$, M iterations are removed from the tail and replaced by the same number of new ones that are added to its head.

One accepts the new reptile, Y_j, with probability

$$A(X_j \rightarrow Y_j) = \min(1,\exp(-S(Y_j))/\exp(-S(X_j))) \qquad (16)$$

This essentially amounts to performing a Metropolis-type decision done on the ratio of VMC-W past−future weights (14).

Described above is a typical Metropolis−Hastings type algorithm,[9] here denoted by RQMC-MH. It generates the same distribution as its well-studied RQMC counterpart,[8] which is not reversible yet still converges to the intended distribution.

For our purposes it is only the middle of a reptile that is important. At this position, analogous to the logical-present in the VMC-W algorithm, we are able to sample from the exact distribution. However, unlike VMC-W, RQMC-MH allows us to directly sample this distribution, rather than doing so indirectly, *via* weighted averages of the quantities of interest.

Unfortunately, the algorithm's failure to meet the assumed criterion of microscopic reversibility causes the time-step bias to accumulate at the middle of the reptile. To alleviate this problem, there is an alternative algorithm, "no-compromise reptation quantum Monte Carlo"[9] (RQMC-NC). When the assumption of microscopic reversibility is relaxed, this algorithm is designed to stabilize the middle of each reptile. This involves modifying the acceptance probabilities as follows:

As before, with probability $^1/_2$, generate the new reptile by removing M iterations from the head with the same number of new ones to be added to the tail. The new reptile is accepted with probability

$$A^0(X_j \rightarrow Y_j) = \min\left(1, \frac{G_0(x_{N/2+1} \rightarrow x_{N/2};\tau)\ldots G_0(x_{N/2+M} \rightarrow x_{N/2+M-1};\tau)\Psi(x_{N/2+M})^2 \exp(-S(Y_j))}{G_0(x_{N/2+M-1} \rightarrow x_{N/2+M};\tau)\ldots G_0(x_{N/2} \rightarrow x_{N/2+1};\tau)\Psi(x_{N/2})^2 \exp(-S(X_j))}\right)$$

Otherwise, remove M iterations from the reptile's tail, replaced by M new ones added to its head. This proposed reptile is accepted with probability

$$A^1(X_j \rightarrow Y_j) = \min\left(1, \frac{G_0(x_{N/2-1} \rightarrow x_{N/2};\tau)\ldots G_0(x_{N/2-M} \rightarrow x_{N/2-M+1};\tau)\Psi(x_{N/2-M})^2 \exp(-S(Y_j))}{G_0(x_{N/2-M+1} \rightarrow x_{N/2-M};\tau)\ldots G_0(x_{N/2} \rightarrow x_{N/2-1};\tau)\Psi(x_{N/2})^2 \exp(-S(X_j))}\right)$$

Both RQMC-MH and RQMC-NC converge very quickly in the following application to dihydrogen, and consequently there is no need to randomize the choice of M as in the original RQMC algorithm.[8]

In practice, one observes reptiles that grow "stale", failing to change or move over a large number of iterations. To ameliorate this, we improved the sampling by incorporating a "bounce".[27] During the first iteration with probability $^1/_2$ one assigns to each reptile the location at which a new segment will be added, *i.e.*, its head or tail. That location is retained for further iterations until a proposed reptile is rejected, at which point the opposite location is used during the next iteration and all subsequent ones until a move is again rejected.

Errors in Simulated Electron Distributions for Dihydrogen

We performed ten independent quantum Monte Carlo simulations on the ground state of dihydrogen, a nodeless system. The form of the trial function employed here was energy- and geometry-optimized Ψ_3, taken from ref 28, a high-quality, explicitly correlated LCAO-MO type wave function, whose variational energy accounted for 93% of the electron correlation energy. Simulations were performed at a small value of the time step, $\tau = 0.05$ au, which is still large enough to expect a bias in the simulated electron distributions to be visible. Electron distributions were constructed by spline-fitting histograms ($^1/_{18}$ au-sized bins) of simulated electron positions, weighted by the number of descendants or by past−future weights (14), for FNDMC-DC and VMC-W, respectively. Runs with these algorithms were done both with and without detailed balance (12). Distributions were also constructed from middle-of-the-reptile electron positions for RQMC-MH and -NC, positions where Φ_0^2 is sampled. In addition, VMC runs, which sample Ψ^2, here with a time-step bias, for the sake of comparison. (VMC simulations commonly reported in the literature employ Metropolis sampling, thereby sample Ψ^2 without a time-step bias.)

All of the simulated electron distributions suffer from the use of a finite time step as well as some further errors, which are mentioned in the discussion. To visualize these, we compared the simulated electron−nucleus and electron−electron distributions (P_{MC}) with their gamma distribution counterparts (P_γ), taking the gamma distributions to be truly exact; Figure 2. In Table 4 we report the integrated absolute error in the simulated distributions: $\int |P_{\mathrm{MC}} - P_\gamma|\, \mathrm{d}r$.

As is to be expected, errors in the variational (VMC) distributions are relatively large, particularly for the electron−electron distribution, due to its time-step bias and the error that is introduced by the trial function itself. The variational error is reduced in the other algorithms by virtue of physical- or weight-branching and better sampling of configuration space; the more the error is reduced, the more effective are these algorithmic details.

Employing detailed balance (DB) with FNDMC-DC dramatically reduces the error both in the simulated electron−nucleus and in electron−electron distributions but does not improve upon the VMC-W distributions. Notably, for both distributions the no-compromise reptation algorithm is more accurate than its Metropolis−Hastings analog, closely followed in accuracy by RNDMC-DB-DC. Overall, the distributions listed in order of decreasing accuracy are as follows: RQMC-NC, FNDMC-DB-DC, RQMC-MH, VMC-W, VMC-W-DB, FNDMC-DC, and VMC.

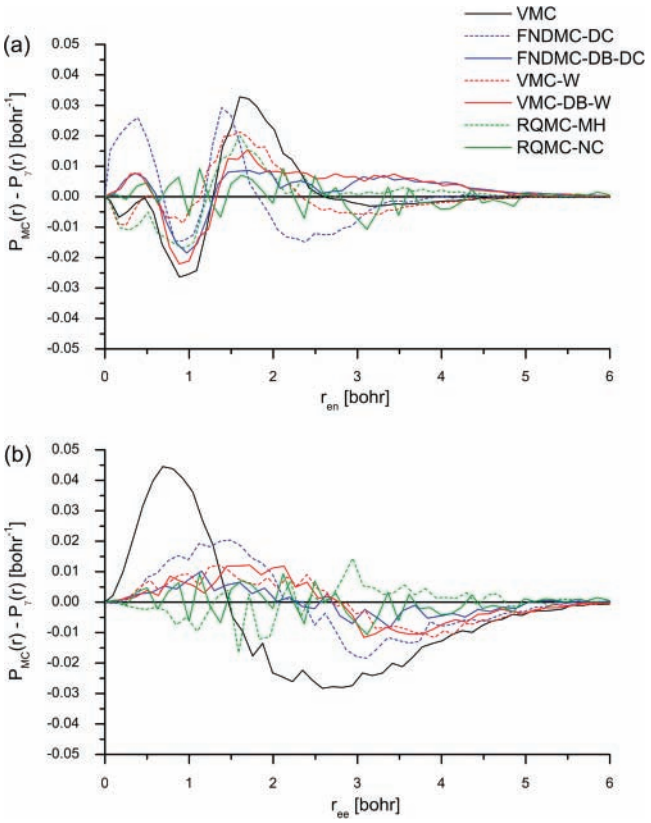

Figure 2. Difference of simulated distributions and gamma distribution model (Figure 1): dihydrogen electron−nucleus distribution (a) and electron−electron distribution (b). Results are displayed for variational Monte Carlo (VMC); fixed-node diffusion Monte Carlo with descendent counting (FNDMC-DC); fixed-node diffusion Monte Carlo with detailed balance and descendent counting (FNDMC-DB-DC); variational Monte Carlo with past−future weighting (VMC-W); variational Monte Carlo with detailed balance and past−future weighting (VMC-DB-W); Metropolis−Hastings reptation quantum Monte Carlo (RQMC-MH); no compromise reptation quantum Monte Carlo (RQMC-NC).

TABLE 4: Integrated Absolute Error in Simulated Electron Distributions for the Ground State of Dihydrogen[a]

algorithm	electron−nucleus	electron−electron
VMC	0.041	0.100
FNDMC−DB	0.043	0.050
FNDMC−DB−DC	0.028	0.019
VMC−W	0.028	0.034
VMC−W−DB	0.035	0.033
RQMC-MH	0.027	0.019
RQMC-NC	0.026	0.016

[a] Algorithm acronyms are as follows: variational Monte Carlo (VMC); fixed-node diffusion Monte Carlo with descendent counting (FNDMC-DC); fixed-node diffusion Monte Carlo with detailed balance and descendent counting (FNDMC-DB-DC); variational Monte Carlo with past−future weighting (VMC-W); variational Monte Carlo with detailed balance and past−future weighting (VMC-DB-W); Metropolis−Hastings reptation quantum Monte Carlo (RQMC-MH); no compromise reptation quantum Monte Carlo (RQMC-NC).

Discussion and Conclusions

The commonly used procedure of comparing the simulated energy and other physical properties with their accurate values, determined from either experiment or high-level theory, provides an indirect assessment of the quality of quantum Monte Carlo-simulated distributions. To our knowledge, the present paper is the first attempt to directly ferret out their quality.

Ground-state dihydrogen has no electron exchange nodes, for which the fixed-node approximation introduces an uncontrolled approximation.[29] However, there are other algorithmic errors affecting the quality of the generated electron distribution, albeit smaller ones than the time-step bias, such as those arising from the use of a finite number of walkers (or reptiles) and method (or lack) of population control. Furthermore, our Hamiltonian is in error by its fixing the positions of the nuclei and ignoring relativistic effects, in common practice with standard quantum chemistry approaches.

Although of a compact form, our trial function is of high quality, representative of that used in the literature nowadays. Algorithms more sophisticated than VMC, namely those described above, are known to improve the variational energy and other physical properties. Indirectly, this suggests an improvement of the underlying electron distributions. Focusing here on them directly, we quantified the extent to which the simulated distributions were improved over that of a drift-diffusion algorithm (time-step-biased VMC), on an algorithm-by-algorithm basis. The nominally best choice, no compromise reptation quantum Monte Carlo, reduces the variational electron−nuclear distribution error by 40%, and the electron−electron distribution error by 80%. Marginally less accurate distributions were generated from fixed-node diffusion Monte Carlo with descendant counting and detailed balance, which thus can be taken as a viable alternative.

It should be noted that, although extensive, our choice of algorithms was not exhaustive. Those not considered include bilinear quantum Monte Carlo[30] and path integral quantum Monte Carlo methods (*e.g.*, ref 31).

In addition, employing a gamma distribution model employed to assess the overall accuracy of quantum Monte Carlo-generated distributions, as we did in this paper, is an approach that may aid in the development and testing of improved density functionals. Unfortunately, applications to more complex systems rests on the availability of highly accurate moments, and these are not routinely reported in the literature. We hope our paper will inspire theoreticians to remedy this deficiency.

Acknowledgment. This work was supported, in part, by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC). We gratefully acknowledge Oliver Bass's technical assistance with preparing the figures.

References and Notes

(1) *Monte Carlo Methods in Ab Initio Quantum Chemistry*; Hammond, B. L., Lester, W. A., Jr., Reynolds, P. J., Eds.; World Scientific: Singapore, 1994.

(2) *Quantum Monte Carlo: Origins, Development, Applications*; Anderson, J. B., Ed.; Oxford: New York, 2007.

(3) Aspuru-Guzik, A.; Lester, W. A., Jr. In *Handbook of Numerical Analysis: Computational Chemistry, Vol. 10*; Ciarlet, P. G., LeBris, C., Eds.; North Holland: Amsterdam, 2003; p 485.

(4) Reynolds, P. J.; Ceperley, D. M.; Alder, B. J.; Lester, W. A., Jr. *J. Chem. Phys.* **1982**. *77*, 5593.

(5) Grossman, J. C. *J. Chem. Phys.* **2002**, *117*, 1434.

(6) Anderson, J. B. *J. Chem. Phys.* **1975**, *63*, 1499.

(7) Klein, D. J.; Pickett, H. M. *J. Chem. Phys.* **1976**, *64*, 4811.

(8) Baroni, S; Moroni, S. *Phys. Rev. Lett.* **1999**, *82*, 4745.

(9) Yuen, W. K.; Farrar, T. J.; Rothstein, S. M. *J. Phys. A: Math. Theor.* **2007**, *40*, F639.

(10) Rothstein, S. M.; Vrbik, J. *J. Chem. Phys.* **1987**, *87*, 1902.

(11) Anderson, J. B.; Garmer, D. R. *J. Chem. Phys.* **1987**, *87*, 1903.

(12) Reynolds, P. J.; Owen, R. K.; Lester, W. A., Jr. *J. Chem. Phys.* **1987**, *87*, 1905.

(13) Rothstein, S. M.; Patil, N.; Vrbik, J. *J. Comput. Chem.* **1987**, *8*, 412.

(14) Umrigar, C. J.; Nightingale, M. P.; Runge, K. J. *J. Chem. Phys.* **1993**, *99*, 2685.

(15) Rothstein, S. M.; Vrbik, J *J. Comput. Phys.* **1988**, *74*, 127.

(16) Reynolds, P. J.; Barnett, R. N.; Hammond, B. L.; Lester, W. A., Jr. *J. Stat. Phys.* **1986**, *43*, 1017.

(17) Barnett, R. N.; Reynolds, P. J.; Lester, W. A., Jr. *J. Chem. Phys.* **1992**, *96*, 2141.
(18) Langfelder, P.; Rothstein, S. M.; Vrbik J. *J. Chem. Phys.* **1997**, *107*, 8525.
(19) Hornik, M.; Rothstein, S. M. In *Recent Advances in Quantum Monte Carlo Methods; Part II*; Lester, W. A., Jr., Rothstein, S. M., Tanaka, S., Eds.; World Scientific: Singapore, 2002; p 71.
(20) *Kendall's Advanced Theory of Statistics, Vol. 1*; Stuart, A., Ord, J. K., Eds.; Oxford: New York, 1987; Chapter 3.
(21) Clementi, E.; Raimondi, D. L. *J. Chem. Phys.* **1963**, *38*, 2686.
(22) Snajdr, M., Rothstein, S. M. *J. Chem. Phys.* **2000**, *112*, 4935.
(23) Coles, B. Unpublished work.
(24) Liu, K. S.; Kalos, M. H.; Chester, G. V. *Phys. Rev. A* **1974**, *10*, 303.
(25) Caffarel, M.; Claverie, P. *J. Chem. Phys.* **1988**, *88*, 1100.
(26) Bosá, I.; Rothstein, S. M. *J. Chem. Phys.* **2004**, *121*, 4486.
(27) Pierleoni, C.; Ceperley, D. M. *Chem. Phys. Chem.* **2005**, *6*, 1.
(28) Snajdr, M.; Dwyer, J. R.; Rothstein, S. M. *J. Chem. Phys.* **1999**, *111*, 9971.
(29) Kalos, M. H.; Pederiva, F. In *Advances in Quantum Monte Carlo*; Anderson, J. B., Rothstein, S. M., Eds.; ACS Symposium Series 953; American Chemical Society: Washington, DC, 2007; p 93.
(30) de Saavedra, A.; Kalos, M. H. *Phys. Rev. E.* **2003**, *67*, 026708.
(31) Sarsa, A.; Schmidt, K. E.; Magro, W. R. *J. Chem. Phys.* **2000**, *113*, 1366.
(32) Frolov, A. M. *Phys. Rev. A* **1998**, *57*, 2436.
(33) Alexander, S. A.; Coldwell, R. L.; Aissing, G.; Thakkar, A. J. *Int. J. Quantum Chem.* **1992**, *S26*, 213.
(34) Sarasola, C.; Elorza, J. M.; Ugalde, J. M. *J. Math. Chem.* **1998**, *23*, 405.
(35) Kolos, W.; Wolniewicz, L. *J. Chem. Phys.* **1965**, *43*, 2429.