

Sampling the exact electron distribution by diffusion quantum Monte Carlo

Allan L. L. East,^{a),b)} Stuart M. Rothstein,^{a)} and Jan Vrbik^{a),c)}
Brock University, St. Catharines, Ontario L2S 3A1, Canada

(Received 28 September 1987; accepted 8 July 1988)

By the accumulation of branching factors in diffusion quantum Monte Carlo (DQMC) and their use as statistical weights, instead of the standard deletion and replication of configurations, we can estimate the averages of (nondifferential) operators taken over the exact electron distribution. This requires only a trivial modification of existing DQMC codes. We illustrate our algorithm by computing ground-state properties for H₂ and LiH which are related to the interelectron distance. We also estimate the dipole moment of LiH.

I. INTRODUCTION

Diffusion quantum Monte Carlo (DQMC) is the simplest of various quantum Monte Carlo techniques available to solve the Schrödinger equation; for a recent review of quantum Monte Carlo see Ref. 1. Several DQMC applications have been reported; a reasonably complete list of citations appears in recent papers published by Garmer and Anderson,² Hammond, Reynolds, and Lester, Jr.,³ and by Rothstein, Patil, and Vrbik.⁴

First it is necessary to review some basic terminology and theory. We will solve the following modified Schrödinger equation:

$$-(1/2)\nabla^2 f(\mathbf{R},t) + [E_L(\mathbf{R}) - E_0]f(\mathbf{R},t) + \nabla \cdot [f(\mathbf{R},t)\mathbf{F}(\mathbf{R})] + \frac{\partial f(\mathbf{R},t)}{\partial t} = 0, \quad (1)$$

where \mathbf{R} represents the $3n$ Cartesian coordinates of the n electrons in the atom or molecule, and E_0 is the exact ground-state energy. The other quantities depend upon $\psi_T(\mathbf{R})$, a known approximate trial solution of the time-dependent Schrödinger equation: $\mathbf{F}(\mathbf{R}) = \nabla\psi_T/\psi_T$, and (the "local energy") $E_L(\mathbf{R}) = H\psi_T/\psi_T$. The objective of using the trial solution is to reduce the variance of the stochastic estimate of the ground-state energy.

It can be shown that the asymptotic ($t \rightarrow \infty$) solution to Eq. (1), $f(\mathbf{R})$ say, is equal to $\psi_T(\mathbf{R})\phi_0(\mathbf{R})$, where $\phi_0(\mathbf{R})$ is the exact ground-state wave function. The function $f(\mathbf{R})$ can be interpreted as a probability density function of a statistical distribution, and as such it can be approximated by computer simulation methods. In DQMC simulations one starts with an "ensemble" of n -electron "configurations" $\{\mathbf{R}_i, i = 1, N\}$, chosen arbitrarily. Then one moves the electrons from \mathbf{R} to \mathbf{R}' in a prescribed manner, corresponding to a small fixed time interval ("time step") τ . The move is designed to simulate the action of the Green's function for Eq. (1), but involves several approximations, so that $f(\mathbf{R})$ and other results become exact only when subsequently extrapolated to time step zero. After a large number of "moves" (called iterations) the distribution of configurations stabilizes (by reaching an equilibrium independent of time). The ensemble is then distributed according to $f(\mathbf{R})$

(plus a small error of the order τ) and may be viewed as a random sample drawn from this distribution.

Here we need be more specific only in the following: Simulating the contribution to the Green's function from the second term requires a process called "branching." This involves a random replication or deletion of configurations; more specifically, one makes

$$B_{i,k} = \exp\{-\tau[E_L(\mathbf{R}_{i,k}) - E_0]\} \quad (2)$$

copies of each configuration, where i denotes the configuration and k the iteration number. (We will refer to $B_{i,k}$ as the "branching factor.")

Some time ago Liu, Kalos, and Chester⁵ indicated how to sample the exact distribution, $\phi_0^2(\mathbf{R})$, using effectively the same simulation procedure. They proved that, in the expected value sense, the number of "descendants" of each configuration \mathbf{R}_i , "many" iterations later, is equal to $\phi_0(\mathbf{R}_i)/\psi_T(\mathbf{R}_i)$. Let us call this number N_i . If one then uses the set of N_i as weights when averaging some quantity $Q(\mathbf{R})$ (evaluated at each \mathbf{R}_i , distributed as $\psi_T\phi_0$):

$$Q_e = \sum_{i=1}^N Q_i N_i / \sum_{i=1}^N N_i, \quad (3)$$

one estimates the expected value of Q taken over the exact distribution of electrons, since

$$\begin{aligned} \langle Q \rangle_e &= \int Q(\mathbf{R})(\phi_0/\psi_T)(\psi_T\phi_0)d\mathbf{R} / \\ &\int (\phi_0/\psi_T)(\psi_T\phi_0)d\mathbf{R} \\ &= \int Q(\mathbf{R})\phi_0^2 d\mathbf{R} / \int \phi_0^2 d\mathbf{R}. \end{aligned} \quad (4)$$

Obviously, it is computationally rather difficult to implement Liu *et al.*'s algorithm: before taking the average of Q in the present iteration one must determine not only the number of immediate descendants of each configuration, but also the descendants of these, for many iterations in the future. Nevertheless, a program to do this has been written and applied to H₂ and N₂ by Reynolds *et al.*⁶ This is a preliminary publication; more details will appear in a followup paper.⁷ Reference 6 also reports some results for Q being replaced by a differential operator.

^{a)} Department of Chemistry.

^{b)} NSERC Undergraduate Summer Research Scholar.

^{c)} Department of Mathematics.

We describe below an algorithm which can achieve the same objective, but it requires only a *minor* modification of DQMC as it circumvents the complications introduced by having to look far into the “future.”

The theoretical basis of our work rests on the following relatively simple argument: *without* branching the simulated distribution is the “variational” one, corresponding to $\psi_T(\mathbf{R})^2$. [This may be easily derived by substituting ψ_T^2 into Eq. (1) without the second term.] On the other hand, *with* branching we obtain the mixed distribution $\psi_T(\mathbf{R})\phi_0(\mathbf{R})$, this being the standard result of DQMC. One would like to carry this a step further, to “square,” so to speak, the branching and obtain ϕ_0^2 , but this cannot be done using the standard process of deletion and replication of configurations. But, suppose instead of performing the usual branching, we simply *accumulate* the branching factors [Eq. (2)] for many iterations (denote the cumulative weight carried by the i th configuration \mathbf{R}_i by W_i). If we then make the corresponding number (W_i) of replicates of each final \mathbf{R}_i in the ensemble, the distribution of these is immediately converted from the variational to the mixed one. (This implies that the weights, like the number of “descendants,” are proportional to ϕ_0/ψ_T .) *Equivalently*, we may use the W_i 's as weights in all averages, and we must get identical results as the old procedure (in terms of the expected values of nondifferential operators). Finally, it follows that by averaging any quantity Q with the *squared* weights:

$$Q_e = \sum_{i=1}^N Q_i W_i^2 / \sum_{i=1}^N W_i^2, \quad (5)$$

one has an estimate of the exact expected value:

$$\langle Q \rangle_e = \int Q(\mathbf{R}) (\phi_0/\psi_T)^2 (\psi_T^2) d\mathbf{R} / \int (\phi_0/\psi_T)^2 (\psi_T^2) d\mathbf{R}, \quad (6)$$

where the \mathbf{R} 's remain distributed according to the “variational” distribution. Thus we have avoided the problem of following the future course of the configurations to do the exact averaging. Furthermore (as we will show below), the only significant additional programming requirement is to keep track of the local energies for sufficiently many of iterations in the past (this from a computational point of view requires only some increase in memory size).

As mentioned above, the DQMC results have a τ -related bias. This necessitates estimating the exact expectation value by extrapolating the τ -biased averages [Eq. (5) and Eqs. (10) and (11) below] to zero time step. (This must be done carefully to achieve reasonable reliability for the final estimate. We emphasize this point in Sec. II. Also, when using weights, the following problem arises: It is easy to show [see Eq. (8) below] that the *variance* of the individual weights increases proportional to M (the number of iterations). Thus, in the $M \rightarrow \infty$ limit only *one* configuration would essentially contribute to all averages. This implies that mindlessly accumulating weights throughout the *whole* simulation would totally ruin the statistical efficiency, to the extent of making this procedure useless.

We have resolved this by accumulating only a relatively small number of weights (L , say) before doing the averaging, and maintaining this number for all iterations. We can show that making L τ dependent (according to $L = A\tau^{-\alpha}$, where $1 < \alpha < 2$) will alleviate the variance problem, without introducing a zero order of τ bias to the extrapolated expected values.

Firstly (to prove no bias), it is safe to assume that stabilization occurs after A/τ iterations, so if we chose α bigger than 1, the equilibrium distribution will be guaranteed in the $\tau \rightarrow 0$ limit. The proper choice of A should result in a value of L which is more than sufficient for good equilibrium even at the largest time step.

Second (to prove the variance is under control), using Eq. (4), the total weight accumulated by the i th configuration in L consecutive iterations is given by

$$W_i = \prod_{k=1}^L B_{i,k} = \exp \left\{ -\tau \sum_{k=1}^L [E_L(\mathbf{R}_{i,k}) - E_0] \right\}. \quad (7)$$

Now

$$\begin{aligned} \text{Var}(W_i) &\approx \text{Var} \left\{ \tau \sum_{k=1}^L [E_L(\mathbf{R}_{i,k}) - E_0] + O(\tau^2) \right\} \\ &= \tau^2 L \{ \text{Var}[E_L(\mathbf{R})] \\ &\quad \times (1 + 2\rho^{(1)} + 2\rho^{(2)} + \dots) \} + O(\tau^3). \end{aligned} \quad (8)$$

The infinite series of serial correlation coefficients for the local energies sums to a quantity proportional to τ^{-1} (Ref. 8). Thus, for $L \propto \tau^{-\alpha}$,

$$\text{Var}(W_i) \approx A\tau^{1-\alpha}, \quad (9)$$

where A is a positive constant.

For α equal to 1.5 (our choice for this paper) the variance of the weights carried by the configurations is no worse than the variance due to the serial correlation of consecutive iterations (present even in simulations with no branching; Ref. 8), therefore not introducing any additional difficulties.

To summarize our algorithm: First select a *fixed* number of configurations N of the order of several hundred. The total number of iterations M and much smaller value of L are chosen, the latter proportional to $\tau^{-3/2}$. After a suitable number of iterations to equilibrate the ensemble, run the DQMC program for L iterations, and then compute W_i for each configuration from Eq. (7). (Here one may set $E_0 = 0$ since it cancels out in our estimators.) The local energies generated in the *subsequent* iteration replace those obtained in the *first* of these L iterations; those of the *next* iteration replace those obtained in the second, etc. (This may be viewed as a “push up stack” of E_L values, efficiently programmed by using a cyclic index to label the iteration number.)

After each such iteration, compute Q_e [Eq. (5)], an estimate of the exact expectation value. In addition, for the purposes of this paper, we also compute

$$Q_m = \sum_{i=1}^N Q_i W_i / \sum_{i=1}^N W_i, \quad (10)$$

$$Q_v = \sum_{i=1}^N Q_i / N, \quad (11)$$

the mixed and variational expectation value estimates, respectively.

Continue to iterate until M estimates [Eqs. (5), (10), and (11)] are obtained. The average of these provides a grand mean estimate of the respective time step biased expectation value of the quantity Q . The whole procedure needs to be repeated with several different values of τ , and all results must be extrapolated (by regression) to $\tau = 0$.

As is almost always the case in DQMC applications, we use a linearly accurate DQMC algorithm. (This means that the distributions and expectation values have a bias, the dominant term of which is proportional to τ itself.) To this accuracy we may employ the following branching factor which is less sensitive to singularities in the local energy:

$$B'_{i,k} = 1 - \tau[E_L(\mathbf{R}_{i,k}) - E_0], \quad (12)$$

with a corresponding

$$W'_i = \prod_{k=1}^L \{1 - \tau[E_L(\mathbf{R}_{i,k}) - E_0]\}, \quad (13)$$

where E_0 should be a reasonably good estimate of the ground-state energy. (The procedure is not overly sensitive to the particular choice of E_0 .)

In Sec. II we will show that expectation values for H_2 are consistently estimated using either choice of weights: Eq. (7) or (13). For the sake of convenience we will only use the "linear weights" [Eq. (13)] for simulations done on another small molecule: LiH.

II. APPLICATIONS

First we will apply our algorithm to compute various properties related to the interelectronic distances for the ground state of H_2 : powers of r_{12} and of the longitudinal (u_z) and transverse (u_c) projections of \mathbf{r}_{12} ; the u_c are by definition nonnegative. The guiding function (ψ_T) was ψ_I in the notation of Ref. 9. We use the BJDJB algorithm, one of several reported in Ref. 4, because for our guiding function it produced the best (least time step biased) estimates of the ground-state energy over a large interval of time steps.

As described in detail in Ref. 4, one iteration involves four changes in the locations of the electrons in each configuration, namely: diffusion, followed by two drifts, and another diffusion. All quantities are evaluated after each of the four moves. The average (per electron) of these four evaluations is in turn weight averaged over the ensemble of configurations, using Eqs. (5), (10), and (11).

Extreme values of $1/r_{12}$, $(1/r_{12})^2$, and $1/u_c$ were truncated. The cutoffs are designed to control the variance of estimators, without introducing a zero order (in τ) bias into their expected values. For this purpose, we obtained histograms of these quantities and chose τ^{-1} -dependent and τ^{-2} -dependent cutoffs, where appropriate, to truncate about 1% of the values at our largest time step. Theoretical justification for the cutoffs is in direct analogy to those employed by us in some relativistic calculations.¹⁴

The results of extrapolating $\langle Q \rangle(\tau)$ (the τ -biased expectation values) to zero τ appear in Table I, together with analytical values either computed or cited by Sharma and Thakkar.¹⁰ We used third order polynomial weighted regression. The ground-state energy estimate (here, only the

TABLE I. Energy and properties derived from the interelectronic distance vector for H_2 molecule ground state at $R = 1.401$ bohr, using two weighting schemes. Energy units are hartrees and distance units are bohr.

	DQMC ^a				Analytical
	Our procedure		Approximate formula ^b		
	W^c	W^d	W^c	W^d	
$\langle E_L \rangle_m$		-1.174(3)			-1.174 475 ^e
$\langle r_{12}^2 \rangle_e$	5.64(6)	5.65(7)	5.52(6)	5.55(7)	5.632 39 ^e
$\langle r_{12} \rangle_e$	2.17(1)	2.18(1)	2.15(1)	2.15(1)	2.168 95 ^e
$\langle r_{12}^{-1} \rangle_e$	0.587(3)	0.590(4)	0.588(3)	0.590(2)	0.587 366 ^e
$\langle r_{12}^{-2} \rangle_e$	0.518(4)	0.517(4)	0.516(5)	0.515(5)	0.518 27 ^f
$\langle u_z^2 \rangle_e$	2.39(3)	2.39(3)	2.32(2)	2.32(2)	2.365 2 ^e
$\langle u_z \rangle_e$	1.244(7)	1.245(8)	1.235(5)	1.234(5)	1.244 1 ^f
$\langle u_c^2 \rangle_e$	3.28(6)	3.27(6)	3.25(5)	3.24(5)	3.267 2 ^e
$\langle u_c \rangle_e$	1.57(2)	1.57(2)	1.57(1)	1.56(1)	1.569 9 ^f
$\langle u_c^{-1} \rangle_e$	1.03(1)	1.03(1)	1.03(1)	1.03(1)	1.040 4 ^f

^a Algorithm BJDJB of Ref. 4 with modifications described in the text. Simulations used ($M =$)500 configurations and ($L =$)INT(1.18 $\tau^{-3/2}$) configurations to accumulate the configuration weights. Several runs of 500 hundred iterations each were done at six equally spaced time steps in the range 0.03–0.18 a.u. Parentheses denote one standard deviation of the $\tau = 0$ intercept derived from weighted regression of the grand mean of the runs.

^b Approximate formula, Eq. (14).

^c Equation (13).

^d Equation (7).

^e Nearly exact value; for energy Ref. 12, for moments Ref. 13.

^f Derived from 36 correlated Gaussian geminals; Ref. 10.

TABLE II. Energy, dipole moment, and properties derived from the interelectronic distance vector for LiH molecule ground state at $R = 3.015$ bohr. Energy units are hartrees and distance units are bohr.

	DQMC ^a		Best CI calculation
$\langle E_L \rangle_m$	- 8.0667(6)		- 8.0690 ^b
$\langle \mu \rangle_e$	- 2.287(25)		- 2.30 ^b
	Parallel spin	Antiparallel spin	
$\langle r_{12}^2 \rangle_e$	13.2(1)	9.9(1)	
$\langle r_{12} \rangle_e$	3.42(1)	2.72(1)	
$\langle r_{12}^{-1} \rangle_e$	0.333(2)	0.66(1)	
$\langle r_{12}^{-2} \rangle_e$	0.132(1)	1.17(1)	
$\langle u_z^2 \rangle_e$	8.98(6)	5.55(4)	
$\langle u_z \rangle_e$	2.718(6)	1.86(1)	
$\langle u_c^2 \rangle_e$	4.2(1)	4.3(1)	
$\langle u_c \rangle_e$	1.70(2)	1.65(2)	
$\langle u_c^{-1} \rangle_e$	1.09(1)	1.29(2)	

^a Algorithm **BDJB-PUSH** (Ref. 16) with branching removed. Simulations used ($M =$) 600 configurations and ($L =$) INT($0.5\tau^{-3/2}$) configurations to accumulate the configuration weights [Eq. (13)]. Several runs of 600 hundred iterations each were done at seven time steps in the range 0.15–0.12 a.u. Parentheses denote one standard deviation of the zero time step intercept derived from weighted regression of the grand mean of the runs.

^b Ref. 17. The exact energy is equal to - 8.0700 a.u. (Ref. 15) and the experimental dipole moment is - 2.29 a.u. (Ref. 18).

mixed distribution is required) also appears in Table I.

We also tested the following approximate formula (first derived by Ceperley and Kalos¹¹), which is accurate to the same order as the trial function itself:

$$\langle Q \rangle_e \approx 2\langle Q \rangle_m - \langle Q \rangle_v. \quad (14)$$

The results are also quoted in Table I.

The exact-sampling results are well within statistical error of the true analytical values. The approximate formula [Eq. (14)] is reasonably accurate for some estimates (which was also the case in applications by Reynolds *et al.*⁶), but has a noticeable error for a few other properties: r_{12}^2 , u_z^2 , and $|u_z|$ in particular.

The precision of our regression intercepts is typically 1%, reasonable in light of the simplicity of our guiding function (double- ξ quality SCF with a single Jastrow correlation function) and the modest CPU invested in our calculations. One finds no apparent advantage to using W'_i over W_i in terms of the accuracy or the variance of the extrapolated values.

Next we consider the LiH molecule. Here we face the inevitable (but relatively small) error of simulating the exact electron distribution using the incorrect nodes of the trial function ψ_τ (again, a double- ξ quality SCF with a single Jastrow correlation function^{9,15}). Also, due to the electrons "overshooting" the nuclei (much more severe than in the case of H_2^4), special techniques¹⁵ must be employed to deal with the resulting time step error introduced into the local energy and all other simulated properties. This necessitates extreme care to ensure that the extrapolated intercepts are reliable. (In addition to the statistical error of the intercepts, there is a systematic error introduced by using a specific

simple regression model which can only approximately "mimic" the exact τ dependence.)

Unlike the case of H_2 , the LiH simulations were done using the **BDJB-PUSH**^{15,16} algorithm. For each configuration in one iteration there are now *three* changes in the locations of the electrons: diffusion, drift, and then an extra "push." Again unlike H_2 , evaluations taken after each location change were kept *separate*; ultimately we obtain *three* distinct $\langle Q \rangle(\tau)$ vs τ curves for each property.

We fit the three curves by weighted regression using first the simplest "plausible" polynomial model (by plausible we mean a set of fits with good agreement between the *three* $\tau = 0$ intercepts—the true values of these intercepts must be *identical*) and then again, using the next highest polynomial, as a way of estimating the "model bias."¹⁵ We quote as our final estimate the average of all these, together with an error which we believe is sufficient to account for both the statistical and systematic uncertainty. (It would be very difficult trying to separate the two.)

The results appear in Table II. Both the energy and dipole moment estimates are in excellent agreement with the true values. We are not aware of any analytical values for the remaining properties which appear in the table.

ACKNOWLEDGMENTS

One of us (S.M.R.) is happy to acknowledge the suggestion of Ajit J. Thakkar to apply DQMC to the estimation of the properties listed in Table I. Our research was supported by grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

- ¹D. Ceperley and B. Alder, *Science* **231**, 555 (1986).
- ²D. R. Garmer and J. B. Anderson, *J. Chem. Phys.* **86**, 4025 (1987).
- ³B. L. Hammond, P. J. Reynolds, and W. A. Lester, Jr., *J. Chem. Phys.* **87**, 1130 (1987).
- ⁴S. M. Rothstein, N. Patil, and J. Vrbik, *J. Comput. Chem.* **8**, 412 (1987).
- ⁵K. S. Liu, M. H. Kalos, and G. V. Chester, *Phys. Rev. A* **10**, 303 (1974).
- ⁶P. J. Reynolds, R. N. Barnett, B. L. Hammond, and W. A. Lester, Jr., *J. Stat. Phys.* **43**, 1017 (1986).
- ⁷R. N. Barnett, P. J. Reynolds, and W. A. Lester, Jr. (preprint cited in Ref. 6).
- ⁸S. M. Rothstein and J. Vrbik, *J. Comput. Phys.* **74**, 127 (1988).
- ⁹P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, Jr., *J. Chem. Phys.* **77**, 5593 (1982).
- ¹⁰B. S. Sharma and A. J. Thakkar, *J. Phys. B* **17**, 3405 (1984).
- ¹¹D. M. Ceperley and M. H. Kalos, in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer, Berlin, 1979), pp. 145–194.
- ¹²W. Kolos and L. Wolniewicz, *J. Chem. Phys.* **49**, 404 (1968).
- ¹³W. Kolos and L. Wolniewicz, *J. Chem. Phys.* **43**, 2429 (1965).
- ¹⁴J. Vrbik, M. F. Depasquale, and S. M. Rothstein, *J. Chem. Phys.* **88**, 3784 (1988).
- ¹⁵M. F. Depasquale, S. M. Rothstein, and J. Vrbik, *J. Chem. Phys.* **89**, 3629 (1988).
- ¹⁶J. Vrbik and S. M. Rothstein, *J. Comput. Phys.* **63**, 130 (1986).
- ¹⁷N. C. Handy, R. J. Harrison, P. J. Knowles, and H. F. Schaefer III, *J. Phys. Chem.* **88**, 4852 (1984).
- ¹⁸L. Wharton, L. Gold, and W. Klemperer, *J. Chem. Phys.* **37**, 2149 (1962).