JOURNAL OF COMPUTATIONAL PHYSICS **61**, 483–498 (1985)

A Pairing Method Which Improves Convergence in Monte-Carlo Estimation of Quantum Mechanical Expectation Values

KATHERINE A. WILSON* AND ROBERT L. COLDWELL

Physics Department, University of Florida, Gainesville, Florida 32611

Received December 20, 1983; revised February 5, 1985

The "pairing" method reduces the variance in Monte-Carlo estimates of expectation values by combining the locations of the largest and smallest contributors to these estimates. The number of paired locations and hence the variance reduction increase with the number of wave function evaluations N_e. This allows Monte-Carlo-like calculations to converge at rates faster than $N_e^{-1/2}$. In particular, an evaluation of the expectation values of the potential, $\langle V \rangle$, for H_2^+ has a standard deviation which decreases as $N_e^{-0.72}$ and an evaluation of $\langle V \rangle$ for spin-aligned H_2 at the minimum in its potential well has a standard deviation which decreases as $N_e^{-0.62}$.

INTRODUCTION

An n-dimensional integral ready for Monte-Carlo integration can have the form

$$I = \int_0^1 \cdots \int_0^1 f(\mathbf{x})\, dx_1 \cdots dx_n = \langle f \rangle \tag{1}$$

where $\mathbf{x}$ is an n-dimensional vector. A Monte-Carlo estimate of I using N random $\mathbf{x}_i$'s is

$$I_{\mathrm{MC}} = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i) \tag{2}$$

where $\mathbf{x}_i$ is the ith choice of the n components of $\mathbf{x}$ with each component being chosen uniformly and randomly between 0 and 1. The standard deviation of the mean value of I_{MC} is σ_{MC}, where

$$\sigma_{\mathrm{MC}}^2 = \frac{\langle (\langle f \rangle - f)^2 \rangle}{N}. \tag{3}$$

* Present address: Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, California 94550.

0021-9991/85 $3.00

For our purposes it is necessary to allow the argument of f to be outside the region 0 to 1. Defining $f(\mathbf{x}+\mathbf{l})=f(\mathbf{x})$ where $\mathbf{l}$ is an n-dimensional vector of integers makes f explicitly periodic so that

$$I_{\mathrm{MC}}=\frac{1}{N}\sum_{i=1}^{N} f(\mathbf{x}_i+\mathbf{a}) \tag{4}$$

where $\mathbf{a}$ is any n-dimensional vector. This allows the introduction of a set of M vectors $\mathbf{a}_q$ similar to the antithetic variates of Hammersly and Hanscomb [1], which can be held constant while $\mathbf{x}_p$ is chosen randomly to yield

$$f_{\mathrm{A}}(\mathbf{x}_p)=\frac{1}{M}\sum_{q=1}^{M} f(\mathbf{x}_p+\mathbf{a}_q). \tag{5}$$

This allows us to define

$$I_{\mathrm{A}}=\frac{1}{N}\sum_{p=1}^{N} f_{\mathrm{A}}(\mathbf{x}_p) \tag{6}$$

with standard deviation

$$\sigma_{\mathrm{A}}^2=\frac{\langle(\langle f\rangle-f_{\mathrm{A}})^2\rangle}{N}. \tag{7}$$

In Eq. (7) the fact that $\langle f_{\mathrm{A}}\rangle=\langle f\rangle$ has been utilized and it should be noted that $\langle f_{\mathrm{A}}^2\rangle$ is not equal to $\langle f^2\rangle$. Only N independent random vectors enter Eqs. (6) and (7) but the number of function evaluations N_e is N times M. Pairing is preferable to unbiased Monte-Carlo if $M\sigma_{\mathrm{A}}^2<\sigma_{\mathrm{MC}}^2$. In this work, a method for systematically finding new $\mathbf{a}_q$'s for which σ_{A}^2 decreases faster than $1/M$ will be described.

A scheme for finding these $\mathbf{a}_q$'s that essentially finds different pseudo-random number generators in each dimension and minimizes σ^2 for a class of "worst" functions has been developed by Haselgrove [2]. The scheme is supposed to yield $\sigma^2=\alpha/N^2$ and indeed for some functions this appears to be achieved. The method was extended by Burdick [3] to include the use of biased selection methods, quantum mechanical expectation values, and standard deviation estimates similar to Eq. (7). The method has $\sigma_{\mathrm{A}}^2=\alpha/N^2$ for small values of N, but with $\sigma_{\mathrm{A}}^2>\sigma_{\mathrm{MC}}^2$. For N on the order of the optimization set σ_{A}^2 becomes smaller than σ_{MC}^2, but after about a factor of 2 gain changes slope and becomes proportional to $1/N$.

The new ingredient in this work is abandonment of the notion of a universal set of $\mathbf{a}_q$'s. Indeed σ_{A}^2 for other functions often rises while that for the function of interest drops. The method is simple, intuitive, and easy to use. It probably works because it utilizes in a small way the locations of the Monte-Carlo function evaluations, information which is usually discarded.

The Method

The largest and smallest values of $f(\mathbf{x}_i)$ contribute the largest amounts to $N\sigma^2_{\mathrm{MC}}$ in Eq. (3). If they could be paired so that they can only be found for a single $\mathbf{x}_i$ value in Eq. (5), the resulting $f_{\mathrm{A}}(\mathbf{x})$ which contains these two values contributes much less to σ^2_{A} in Eq. (7) while contributing exactly the same amount to I_{A} in Eq. (6).

The pairing method utilizes a number of N point estimators of I_{A}. The first level 0 is made by selecting N vectors $\mathbf{x}_i$ uniformly and randomly and averaging as in Eq. (2). The vector $\mathbf{b}^1_1$ is made equal to the $\mathbf{x}_i$ for which $f(\mathbf{x}_i)$ is largest and the vector $\mathbf{b}^1_2$ to the $\mathbf{x}_i$ for which $f(\mathbf{x}_i)$ is smallest. Then at level 1

$$\mathbf{a}_1 = \mathbf{b}^1_1 \tag{8}$$

$$\mathbf{a}_2 = \mathbf{b}^1_2. \tag{9}$$

The second N point estimator consists of an independent set of $\mathbf{x}_p$'s used to average the $f_{\mathrm{A}}(\mathbf{x}_p)$ in Eq. (5) as in Eq. (6) with the $\mathbf{a}_q$'s given in Eqs. (8) and (9). Note that an $\mathbf{x}_p$ of zero in this set would yield

$$f_{\mathrm{A}}(0) = \tfrac{1}{2}(f(\mathbf{a}_1) + f(\mathbf{a}_2)) \tag{10}$$

which as a pair contributes less to the sum in Eq. (7) than these same terms occurring independently contribute to the sum in Eq. (3).

The $\mathbf{x}_p$'s for which $f_{\mathrm{A}}(\mathbf{x}_p)$ is extreme are saved as $\mathbf{b}^2_1$ and $\mathbf{b}^2_2$. This enables us to generate the set of $\mathbf{a}$'s for level 2 as

$$\begin{aligned} \mathbf{a}_1 &= \mathbf{b}^1_1 + \mathbf{b}^2_1 \\ \mathbf{a}_2 &= \mathbf{b}^1_1 + \mathbf{b}^2_2 \\ \mathbf{a}_3 &= \mathbf{b}^1_2 + \mathbf{b}^2_1 \\ \mathbf{a}_4 &= \mathbf{b}^1_2 + \mathbf{b}^2_2 \end{aligned} \tag{11}$$

to be used in the next N point estimator of i. Finally, the Lth set of a's to be used in the $L+1$ level estimator are

$$\mathbf{a}_j = \sum_{i=1}^{L} \mathbf{b}^i_{pj} \tag{12}$$

where pj is the jth set of the 2^L possible sets of $\{1, 2, 1, \ldots, 2\}$. Note that only $2L$ terms need be stored to generate the $M = 2^L$ vectors $\mathbf{a}_q$.

The result of these efforts is that $f_A(x_i)$ may be written

$$\begin{aligned} f_A(x_i) &= \frac{1}{M} \sum_{j=1}^{M} f(\mathbf{x}_i + \mathbf{a}_j) \\ &= \frac{1}{M} \sum_{j=1}^{2} \sum_{k=1}^{2} \cdots \sum_{l=1}^{2} f(\mathbf{x}_i + \mathbf{b}_j^1 + \mathbf{b}_k^2 + \cdots + \mathbf{b}_l^L). \end{aligned} \tag{13}$$

Note that each additional level utilizes a new set of N random $\mathbf{x}_p$'s and doubles the total number of function evaluations.

Before introducing the expectation values used in testing the pairing method, it can be illustrated with a simple $2d$ lattice function. Let space consist only of the 16 lattice points shown in Fig. 1 along with some of its periodic repeats. The function $f(x, y)$ is 4 at points (0.25, 0.25), (0.25, 0.50), (0.50, 0.25), (0.50, 0.50) and zero for the other 12. The average value $\langle f \rangle = 1$ and $\langle f^2 \rangle = 4$ which implies $N\sigma_{\mathrm{MC}}^2 = 3$. A pairing method estimation of (f) at the end of level 0 would thus have the usual Monte-Carlo standard deviation $\sigma^2 = 3/N$. In addition the code would have found $\mathbf{b}_1^1 = (0.25, 0.25)$ or its equivalent and $\mathbf{b}_2^1$ would be (0.0, 0.50) or one of the 12 other x, y value sets for which $f(x, y) = 0$. This one will be followed through a number of levels.

The vectors $\mathbf{a}_q$ at level 1 are (0.25, 0.25) and (0, 0.50), there is one x for which $f_A = 4$, which makes $b_1^2 = (0.250)$, six x's for which $f_A = 2$, and nine for which $f_A = 0$, so that b_2^2 can be picked as (0, 0.75). For this choice at level 1, $N\sigma_A^2 = \frac{3}{2}$, exactly the same as for this many random function evaluations. The vector $\mathbf{a}_q$ at level 2 is (0.50, 0.25), (0.25, 0.50) (0.25, 0), (0, 0.25). There are four x's with $f_A = 2$, eight with $f_A = 1$, and four with $f_A = 0$. The b's may be chosen to be $b_1^3 = (0, 0.25)$ and $b_2^3 = (0.5, 0.75)$. At level 2, $N\sigma_A^2 = \frac{1}{2}$, which is less than the value of $\frac{3}{4}$ for this many random evaluations. The vectors $\mathbf{a}_q$ at level 3 are (0.50, 0.50), (0.25, 0.75), (0.50, 0), (0.25, 0.25), (0, 0.50), (0.75, 0.75), (0, 0), and (0.75, 0.25) which is the checkerboard pattern covering every other square shown in Fig. 1. For all x, $f_A = 1$ and thus at level 3 $N\sigma_A^2 = 0$.

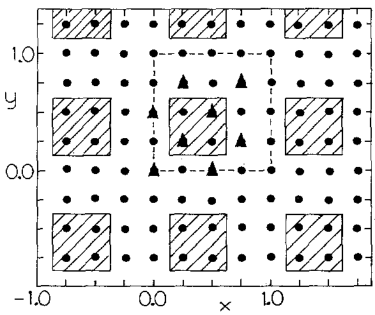

FIG. 1. A simple function $f(x, y)$ defined on the lattice points. In the shaded regions $f(x, y) = 4$, elsewhere $f(x, y) = 0$. The Δ's are the $\mathbf{a}_q$ vectors at level 3 in the example.

The example is so simple that the correct four trapezoidal rule points also give an exact answer. Furthermore only one of the multitude of ways of reaching level 3 has been followed in detail. A bit of promise, however, is present in the example.

Expectation Values

A Monte-Carlo estimate of the expectation value of f with respect to a trial wave function ψ_t may be written as

$$\langle f \rangle = \frac{\int_0^1 F(\mathbf{x})\, d\mathbf{x}}{\int_0^1 G(\mathbf{x})\, d\mathbf{x}} = \frac{1/N \sum_{i=1}^{N} F_A(\mathbf{x}_i)}{1/N \sum_{i=1}^{N} G_A(\mathbf{x}_i)} \tag{14}$$

where

$$F = f\psi_t^2, \qquad G = \psi_t^2 \tag{15}$$

and F_A and G_A are defined by Eq. (5). The same $\mathbf{x}_i$ vectors are used in both the numerator and denominator which makes the estimate more accurate than performing two independent estimates of the integrals and dividing by the result [9], though the convergence rate is still the $N^{-1/2}$ typical of Monte-Carlo methods.

The standard deviation σ in the estimate of an expectation value using (14) may be written as

$$\sigma_A^2 = \frac{1}{N\langle G_A \rangle^2} \langle (F_A - \langle f \rangle G_A)^2 \rangle \tag{16}$$

which is equivalent to Eq. (20) derived in Appendix A.

The biased selection method described in Appendix B changes a simple Monte-Carlo estimate to

$$I_{MC} = \frac{1}{N} \sum_i f(\mathbf{x}(\mathbf{q}_i))/w(\mathbf{q}_i) \tag{17}$$

where $\mathbf{q}_i$ is chosen randomly and $\mathbf{x}$ is functionally related to $\mathbf{q}$ instead of being chosen randomly as in Eq. (2). The expectation values in Eq. (14) remain the same except Eq. (15) becomes

$$F = f(\mathbf{x}(\mathbf{q}))\, \psi_t(\mathbf{x}(\mathbf{q}))/\bar{w}(\mathbf{q}), \qquad G = \psi_t^2(\mathbf{x}(\mathbf{q}))/\bar{w}(\mathbf{q}) \tag{18}$$

where $\bar{w}(\mathbf{q})$ is given explicitly for H_2^+ in Eq. (28).

The fact that for expectation values the standard deviation is proportional to $\langle (F_A - \langle f \rangle G_A)^2 \rangle$ introduces a slight complication in the method, an estimate of $\langle f \rangle$ enters into the function being extremized. A crude estimate of $\langle f \rangle$ needs to be found before the first level, then subsequently the estimate of $\langle f \rangle$ from the last completed level can be used.

RESULTS

Expectation values of the energy $\langle H\psi_t/\psi_t\rangle$, the kinetic energy $\langle T\rangle = \langle|\nabla\psi/\psi|^2\rangle$, and the potential energy $\langle V\rangle$ were estimated for H_2^+ and spin-aligned H_2 with nuclei 7.8 a_B apart using the trial wave functions described in Appendix C and the biased selection method of Appendix B combined with pairing of the randomly chosen $\mathbf{q}$ vectors. The $\mathbf{q}$ vectors for H_2^+ are four dimensional, three electron coordinates plus one extra from the biased selection method. The $\mathbf{q}$ vectors for H_2 are eight dimensional, six electron coordinates plus two extra from the biased selection method. Though in principle the H_2^+ integrals can be reduced to two dimensional and the H_2 integrals to five dimensional, owing to the fact that Monte-Carlo methods are very insensitive to such reductions no attempts were made in this direction.

Table I and Fig. 2 give the results for H_2^+ from two independent runs each involving slightly more than 4×10^6 wave function evaluations. The expectation value of $\langle V\rangle$ for both runs was calculated using Eqs. (13) and (14) and the standard deviation was calculated using Eq. (16). The M in Eq. (13) was 1 for levels -1 and 0 and 2^l for higher levels. The first run used $N=1024$ in Eq. (13) and the final M was 2^{11}, and the second run used $N=64$ in Eq. (13) and the final M was 2^{15}. Note that N_e is N times the sum of all M_l up to and including the last level. At each level the standard deviation was calculated using Eq. (16) with $N=1024$ for the first pass and $N=64$ for the second pass. Since for large M the functions F_A and G_A are themselves almost integral estimates, this is roughly equivalent to evaluating these integrals 1024 times for $N=1024$ and 64 times for $N=64$ and determining the standard deviations from the resulting variations. For small M, using $N=64$ may result in too few function evaluations to adequately sample the exceptionally large and small regions resulting in standard deviation estimates with rather large standard deviations. As M increases, however, F_A and G_A rapidly become very smooth owing to the fact that such regions are by design becoming very unlikely to be missed. The final estimates incorporated the results of all levels by utilizing Eq. (22) for the expectation values and Eq. (23) for their standard deviations.

Since the expensive part of the calculation is finding $\mathbf{x}(\mathbf{q})$ and evaluating $\psi(\mathbf{x})$, in

TABLE I

Results for H_2^+ (4.19×10^6 Wave Function Evaluations)

N	$\langle V\rangle$ (Ry)	$(\langle T\rangle = \langle\lvert\nabla\psi/\psi\rvert^2\rangle)$ (Ry)	$\langle H\psi/\psi\rangle$ (Ry)	$(\langle H\psi/\psi\rangle - \langle V\rangle - \langle T\rangle)$ (Ry)
1024	-1.975404 ±0.000068	0.969459 ±0.000043	-1.00600545 ±0.00000008	-6.045×10^{-5} $\pm8.045\times10^{-5}$
64	-1.975488 ±0.000048	0.969446 ±0.000022	-1.00600543 ±0.00000004	$+3.66\times10^{-5}$ $\pm5.28\times10^{-5}$

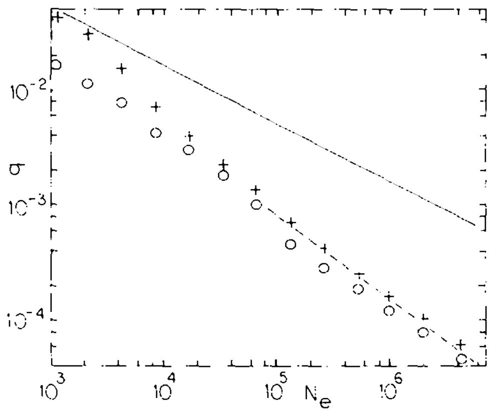

FIG. 2. Standard deviation in $\langle V \rangle$ for H_2^+ versus the number of wave function evaluations (pairing with respect to V). (—) σ_{MC} without pairing. (- -) Line for which $\sigma_A \propto N_e^{-0.72}$. (+) Results using $N = 1024$ in Eq. (16). (○) Results using $N = 64$ in Eq. (16).

addition to $\langle V \rangle$, expectation values of $\langle T \rangle = \langle |\nabla\psi/\psi|^2 \rangle$ and of $\langle H\psi/\psi \rangle$ were estimated in the same runs. However, in the term $F_A - \langle f \rangle G_A$, whose extremal values determine the **q**'s which are to be paired, f was kept equal to V. The final results of the two runs are given in Table I along with the very satisfying fact that to within the error $\langle H\psi/\psi \rangle = \langle T \rangle + \langle V \rangle$. This last is an important test of the integration since an integration by parts has removed the singularity in T which otherwise nearly cancels the singularity in V to give an $H\psi/\psi$ with almost no singularity. Thus $\langle T \rangle$, $\langle V \rangle$, and $\langle H\psi/\psi \rangle$ emphasize somewhat different regions of space and different properties of ψ and will normally not be equal if there are errors either in calculating the derivatives of ψ or in evaluating the integrals. Indeed the ability to make this test is one reason for wanting an accurate $\langle V \rangle$.

Figure 2 shows the standard deviation in $\langle V \rangle$ for H_2^+ as a function of the number of wave function evaluations. The second to the last point plotted required 700 CPU seconds on an IBM 3081, and the last point 1400 CPU seconds. The standard deviation is decreasing as $N_e^{-0.72}$. If the desired quantity were $\langle V \rangle$ with a standard deviation less than 10^{-4} Ry, simple Monte-Carlo would require nearly 100 times as many CPU seconds to reach the desired accuracy.

A more detailed examination of the two runs is shown in Fig. 3, where the ratio of the standard deviation per wave function evaluation at the last level calculated to that of the unpaired Monte-Carlo result is plotted both for $\langle V \rangle$ and also for $\langle H\psi/\psi \rangle$ calculated at the same time using **a** vectors selected to minimize the standard deviation in $\langle V \rangle$. Note that the abscissa in this plot is the level number l, not the number of function evaluations. Even though the standard deviation in the $N = 64$ run at, for example, level 7 is much higher than that for the $N = 1024$ run, the number of function evaluations to find the $N = 64$ result is so much smaller that, as seen in Fig. 2, the $N = 64$ result has the smaller total standard deviation. The standard deviation for $\langle H\psi/\psi \rangle$ and also for $\langle T \rangle$ (not shown because its ratio

581/61/3-10

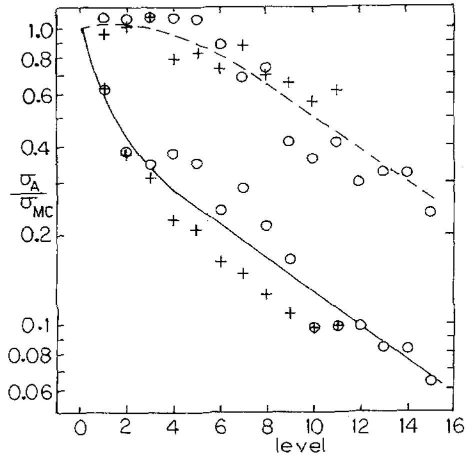

FIG. 3. Ratios σ_A/σ_{MC} for $\langle V \rangle$ and $\langle H\psi/\psi \rangle$ in H_2^+ versus level when pairing is with respect to V. (– – –) σ_A/σ_{MC} for $\langle H\psi/\psi \rangle$. (——) σ_A/σ_{MC} for $\langle V \rangle$. (+) Values for $N=1024$ in Eq. (16). (○) Values for $N=64$ in Eq. (16).

to σ_{MC} is very nearly the same as that for $\langle H\psi/\psi \rangle$) at low levels is constant at the unpaired value, then at about level 5 begins to decrease at nearly the same rate as that in $\langle V \rangle$, indicating that the set of **a**'s being constructed in this case lowers the standard deviations for all three quantities.

If one only wanted to know $\langle H\psi/\psi \rangle$, one would use $H\psi/\psi$ as the f in $F_A - \langle f \rangle G_A$ not V. The results of an $N=400$ run doing this are shown in Fig. 4. The standard deviation in $\langle H\psi/\psi \rangle$, especially for levels less than 5, is less than when the pairing was carried out with respect to V. The standard deviation in $\langle V \rangle$ using these **a**'s, however, is larger and moreover appears to be decreasing for large levels as $N_e^{-0.57}$ rather than $N_e^{-0.72}$. The reason for this is probably the constancy of $H\psi/\psi$ for $\mathbf{x}(\mathbf{q})$'s which cause the potential to vary over wide ranges.

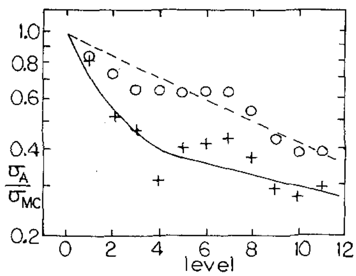

FIG. 4. Ratios σ_A/σ_{MC} for $\langle V \rangle$ and $\langle H\psi/\psi \rangle$ in H_2^+ versus level when pairing is with respect to $H\psi/\psi$. (– – –) σ_A/σ_{MC} for $\langle H\psi/\psi \rangle$. (——) σ_A/σ_{MC} for $\langle V \rangle$. (○) Values for $\langle H\psi/\psi \rangle$, $N=400$ in Eq. (16). (+) Values for $\langle V \rangle$, $N=400$ (same run).

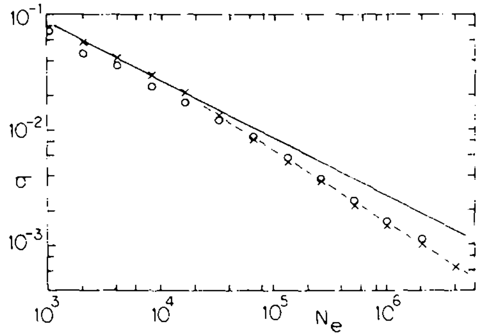

FIG. 5. Standard deviation in $\langle V\rangle$ for spin-aligned H_2 versus the number of wave function evaluations (pairing with respect to V). (——) σ_{MC} without pairing. (– –) Line for which $\sigma_A \propto N_e^{-0.62}$. (×) Results using $N = 1024$ in Eq. (16). (○) Results using $N = 64$ in Eq. (16).

The H_2^+ calculations were made as four-dimensional calculations, but there is an underlying two-dimensional structure. The pairing method in part is simply uncovering this structure and utilizing it to achieve a relatively slow, by two-dimensional standards, convergence rate which is still faster than unpaired Monte-Carlo. The spin-aligned H_2 system which is treated here as an eight-dimensional integral cannot be reduced below five dimensions and is therefore a much more stringent test of the method.

The standard deviation in $\langle V\rangle$ for spin-aligned H_2 as a function of the number of wave function evaluations is shown in Fig. 5. Approximately $2\frac{1}{2}$ hours of IBM 3081 CPU time were required to reach the last x in this figure. The standard deviation σ_A is decreasing as $N_e^{-0.62}$ and is approximately half the size of σ_{MC} at the highest numbers of function evaluations considered. The results of these calculations are shown in Table II. If the goal were to find $\langle V\rangle$ with σ_A less than 10^{-4} Ry approximately 9×10^7 function evaluations would be required assuming

TABLE II

Results for H_2

N	Ne	$\langle V\rangle$ (Ry)	$(\langle T\rangle = \langle\lvert\nabla\psi/\psi\rvert^2\rangle)$ (Ry)	$\langle H\psi/\psi\rangle$ (Ry)	$(\langle H\psi/\psi\rangle - \langle V\rangle - \langle T\rangle)$ (Ry)
1024	4.2×10^6	-4.00002	2.0000873	-2.0000364	$+1.0\times 10^{-4}$
		± 0.00070	± 0.0000078	± 0.0000012	$\pm 7.0\times 10^{-4}$
1024	2.1×10^6	-3.99874	2.000109	-2.0000366	$+14.1\times 10^{-4}$
		± 0.00095	± 0.000012	± 0.0000017	$\pm 9.5\times 10^{-4}$
64	2.1×10^6	-4.00139	2.0000945	-2.0000378	-12.6×10^{-4}
		± 0.00113	± 0.0000091	± 0.0000014	$\pm 11.3\times 10^{-4}$

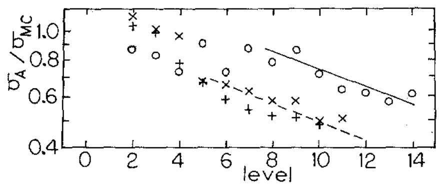

FIG. 6. Ratio σ_A/σ_{MC} for $\langle V \rangle$ for spin-aligned H_2 when pairing is with respect to V. (+, ×) Two independent runs with $N = 1024$ in Eq. (16). (○) $N = 64$ in Eq. (16).

the continuance of this convergence rate. This is less than the number required to reduce σ_{MC} to this size by a factor of 8.5.

In Fig. 6, the ratio of σ_A as given by Eq. (16) to σ_{MC} the unpaired result is shown for two $N = 1024$ runs along with one $N = 64$ run. The $N = 64$ run is not only higher than the $N = 1024$ runs, it is so much higher that the factor of 32 or 5 level advantage in the number of function evaluations is almost exactly canceled, leaving the $N = 1024$ and $N = 64$ σ_A's almost equal as already seen in Fig. 5. An interpretation of this is that the N evaluations must contain rather rare events to properly pair them. Note, however, that the ratio with $N = 64$ drops for the first four levels then fluctuates but remains approximately constant for the next five levels. This implies the existence of gross features which pairing can account for relatively early coupled with finer details which require many more function evaluations.

If one is solely interested in the expectation value of $\langle H\psi/\psi \rangle$, $H\psi/\psi$ should be used as the f in the $F_A - \langle f \rangle G_A$ which determines the locations paired rather than V. An $N = 1024$ run of spin-aligned H_2 carried out to level 9 in which this was done is shown in Fig. 7. The ratio σ_A/σ_{MC} decreases steadily from 1.0 at level 0 to 0.9 at level 9, indicating a slightly improved convergence rate of $N_e^{-0.51}$. In the same run the ratio of σ_A/σ_{MC} for $\langle V \rangle$ using **a** vectors selected to improve σ_A/σ_{MC} for $\langle H\psi/\psi \rangle$ increases rather dramatically to a maximum of 2.76 at level 8. At level 8, 2.6×10^6 wave function evaluations yield the same accuracy in $\langle H\psi/\psi \rangle$ as 3.2×10^6

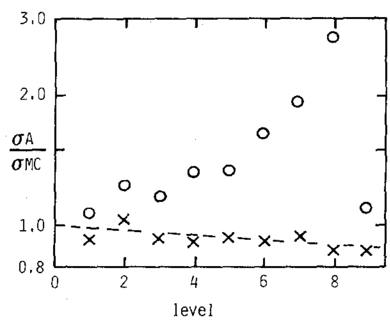

FIG. 7. Ratios σ_A/σ_{MC} for $\langle V \rangle$ and $\langle H\psi/\psi \rangle$ for spin-aligned H_2 when pairing is with respect to $\langle H\psi/\psi \rangle$, $N = 1024$ in Eq. (16). (○) σ_A/σ_{MC} for $\langle V \rangle$. (×) σ_A/σ_{MC} for $\langle H\psi/\psi \rangle$.

unpaired wave function evaluations and also the same accuracy in $\langle V\rangle$ as 3.4×10^5 unpaired wave function evaluations. This comes about because $(H\psi/\psi - E)\,\psi^2$ has both its largest positive and negative values in the same regions where $(V-\langle V\rangle)\,\psi^2$ is all of one sign. Thus the values that cancel to shrink σ_A for $H\psi/\psi$ add to enlarge σ_A for V.

CONCLUSION

The pairing method constructs vectors $\mathbf{a}_j$ containing the coordinate values giving rise to the most extreme values of the F defined in Eq. (18). These vectors contain information about the coordinate values giving rise to F which is usually ignored in Monte-Carlo schemes for estimating integrals. Only a slight modification of Monte-Carlo, i.e., the replacement of $F(x_i)$ by $F_A(x_i)$ as defined in Eq. (13), is required. The overhead is so small as to be almost undetectable while the accuracy gains can be large. In summary, pairing should be considered when truly accurate expectation values are needed and the number of function evaluations is large compared with the expected oscillations in the function being evaluated. Accuracy increases of a factor of 10 or computer time savings of a factor of 100 have been shown in the present work.

APPENDIX A: ERROR ESTIMATION

The statistical error which results from estimating the expectation value, E, using N random values can be determined, as explained in [4], by summing the squares of the differences between the N evaluations of E over $N-1$ values which result from omitting each x_i in turn. The difference in the estimate due to the omission of the ith term is

$$\delta_i = E - \frac{[\langle F\rangle - (1/N)\,F_A(x_i)]}{[\langle G\rangle - (1/N)\,G_A(x_i)]}$$

which noting that $E=\langle F\rangle/\langle G\rangle$ and combining terms becomes

$$\delta_i = \frac{1}{N}\frac{[F_A(x_i) - EG_A(x_i)]}{[(G)-(1/G)\,G_A(x_i)]}$$

where E, F, G, $F_A(x_i)$, and $G_A(x_i)$ are defined in the second section. Since $G_A(x_i)\ll N\{G\}$, we approximate $1-G_A(x_i)/N\{G\}$ by 1 and simplify to obtain

$$\delta_i = \frac{1}{N\langle G\rangle}\,[F_A(x_i) - EG_A(x_i)] \tag{19}$$

so that the variance is given by

$$\sigma^2 = \sum_{i=1}^{N} \sigma_i^2 = \frac{1}{N\langle G \rangle^2} [\langle F^2 \rangle - 2E\langle FG \rangle + E^2 \langle G^2 \rangle]. \tag{20}$$

Note that, as in Eq. (5), M wave function evaluations are required for each of the N values of $F_A(x_i)$ and $G_A(x_i)$ for a total of $N_e = N*M$ wave function evaluations to determine E. The variations of the N values of F_A and G_A which are statistically independent are then used by Eq. (20) to determine σ.

The average over estimates with varying accuracies is done in the standard fashion [5] which minimizes χ^2 with respect to the estimate of E:

$$\frac{\partial(\chi^2)}{\partial E} = 0 = 2 \sum_{i=1}^{L} \left(\frac{E - E_i}{\sigma_i^2} \right). \tag{21}$$

Solving the above expression for E yields

$$E = \sum_{i=1}^{L} \left(\frac{E_i}{\sigma_i^2} \right) \Big/ \sum_{i=1}^{N} \left(\frac{1}{\sigma_i^2} \right) \tag{22}$$

so the weight factor for each estimate is equal to $1/\sigma_i^2$. The σ^2 at each stage is given by

$$\frac{1}{\sigma^2} = \sum_{i=1}^{L} \frac{1}{\sigma_i^2}. \tag{23}$$

A word of caution is required here in case an estimate results with an artificially low variance. The incorrect low estimate could dominate both Eqs. (22) and (23) normally giving rise to both an incorrect E and σ. This case is detectable owing to the incorrect E which causes χ^2 to exceed its normal value. The case where the E_i associated with the incorrect σ_i is not itself in error is much harder to detect. For N sufficiently large, one need not worry about this situation because the error in the variance converges as $1/N$, but one does need to be aware of this possibility in the early levels. We checked our results over the initial stages by performing independent evaluations and comparing the results for consistency at each of the levels.

APPENDIX B: The Biased Selection Method

The biased selection method, as used here, is a specialized form of information sampling involving two centers. Multidimensional points, q_i, are selected with relative probabilities w_i to ensure that all regions of importance are sampled. This method is adapted from the work done by R. L. Coldwell and R. E. Lowther [6].

The H_2^+ ion is positioned in a box of maximum side length of 12.8 Bohr radii. The axis which contains the two nuclei is along the longest diagonal possible between two corners of the box. The arrangement does not make use of the radial symmetry of the H_2^+ ion. We chose to do this to test whether the code could utilize this information without it being explicitly written in the code. We felt that this situation is more in line with what we would expect the code to do for more complex systems, where underlying symmetries are not known.

The electronic coordinates are selected with respect to one of the nuclei. Since the two nuclei are identical, the nucleus used to specify the placement of the electron is a function of a random value, q_1, where nucleus A is used for $q_1 \leqslant 0.5$, and nucleus B for $q_1 \geqslant 0.5$.

The electron distance from the nucleus, r, is selected such that

$$q_2 = \int_0^r h(r')\,dr' \bigg/ \int_0^1 h(r')\,dr' \tag{24}$$

i.e., with a density proportional to $h(r)$. The function $h(r)$ is obtained by doing a linear interpolation between 64 points spaced 0.2 a_B apart of the function

$$g(r) = r^2 \sum_{i=1}^{3} \alpha_i^3 e^{-\alpha_i r} \tag{25}$$

with $\alpha = \{4, 2, 0.5\}$. The first point $h(0)$ is set equal to $h(0.2)$ so that the resulting weight function near the nuclei contains an r^2 that more than concels the $1/r$ in the potential. Values for ϕ and $\mu = \cos\theta$ are specified by $\phi = 2\pi q_3$ and $\mu = 2q_4 - 1$, so that the cartesian coordinates of the electron are

$$\begin{aligned} x(q_1, q_2, q_3, q_4) &= r(q_2)\sin[\phi(q_3)]\sqrt{1-[\mu(q_4)]^2} + X(q_1) \\ y(q_1, q_2, q_3, q_4) &= r(q_2)\cos[\phi(q_3)]\sqrt{1-[\mu(q_4)]^2} + Y(q_1) \\ z(q_1, q_2, q_3, q_4) &= r(q_2)\,\mu(q_4) + Z(q_1). \\ X(q_1) = Y(q_1) &= Z(q_1) = \pm r_{AB}/2\sqrt{3} \end{aligned} \tag{26}$$

where the plus sign is used with one nucleus and the minus sign is used with the other and r_{AB} is the distance between the two nuclei expressed in Bohr radii.

The relative probability of choosing this position in cartesian coordinates from all possible positions is

$$\omega(\mathbf{q}) = \frac{h(r)}{4\pi r^2}. \tag{27}$$

Since the position of the electron may be chosen with respect to either of the two nuclei, the average probability for selecting positions is

$$\bar{\omega}(\mathbf{q}) = \frac{1}{2}\left[\frac{h(r_{wA})}{4\pi r_A^2} + \frac{h(r_B)}{4\pi r_B^2}\right] \tag{28}$$

where the subscripts A and B indicate which of the two nuclei is being used for the placement of the electron.

Using the biased selection method in this manner allows an integral to be written

$$I = \int_0^1 f(\mathbf{x})\, d\mathbf{x} = \int_0^1 \left\{ \frac{f[x(\mathbf{q})]}{\bar{\omega}[x(\mathbf{q})]} \right\} d\mathbf{q}. \tag{29}$$

The MC estimate of this integral is then $\langle f/\omega \rangle$ and its variance is that of $(f/\bar{\omega})$. Note that $f[x(q)]/w[x(q)]$ will tend to be constant, so that when a Monte-Carlo estimate of Eq. (29) is obtained, the variance between individual evaluations of the integrand will be small which will reduce the inherent Monte-Carlo error.

The biased selection procedure for spin-aligned H_2 uses an $h(r)$ in which only $\alpha = 2$ and allows for a 7% probability of placing each electron completely at random within the large box, a 3% probability of placing both electrons with respect to the same nucleus, and a 12% probability of placing the second electron at a random distance between 0 and 1.3 a_B from the first. This last probability removes the singularity in the electron-electron part of the potential. The resulting $w(q)$ is of course averaged over every possible way of reaching the final configuration exactly as in Eq. (28) above.

APPENDIX C: THE TRIAL WAVE FUNCTIONS

The pairing method was tested on two systems consisting of two protons, A and B, 7.8 Bohr radii apart. The first system, H_2^+, has only one electron, while the second, spin-aligned H_2, has two spin-up electrons. This nuclear separation distance is approximately the potential minimum of the $^3\Sigma_g^+$ state of H_2, spin-aligned hydrogen. The one-electron wave function is

$$\psi = \chi(r_{1A} + r_{1B})\, \phi(r_{1A} - r_{1B}) \tag{30}$$

with

$$\chi(x) = e^{-(1/2)x} \sum_{i=0}^{4} a_i x^i \tag{31}$$

and

$$\phi(x) = (e^{(1/2)x} + e^{-(1/2)x}) \sum_{i=0}^{2} a_{(i+5)} x^{2i} + (e^{(1/2)x} - e^{-(1/2)x})\, x \sum_{i=0}^{1} a_{i+8} x^{2i+1} \tag{32}$$

where the parameters given in Table III were found by minimizing

$$\sigma^2 = N \left\{ \sum_{i=1}^{N} \psi_i^4/\omega_i^2 (H\psi_i/\psi_i - \langle E \rangle)^2 \right\} \Big/ \left\{ \sum_{i=1}^{N} \psi_i^2/\omega_i \right\}^2 \tag{33}$$

TABLE III

The Coefficients for $\chi(x)$ and $\phi(x)$ in Eq. (31) and Eq. (32)

i	0	1	2	3	4
a_i	1.0	-9.54565×10^{-3}	-1.75124×10^{-3}	1.06063×10^{-4}	-3.38318×10^{-6}
i	5	6	7	8	9
a_i	1.0				1.07657×10^{-5}

with respect to these parameters for $N = 1000$. The ω_i refers to the bias used in selecting the N configurations as described in Appendix B. The $\sqrt{N\sigma}$ found is 2.8×10^{-4} Ry with $\langle E \rangle = -1.00600545 \pm 8 \times 10^{-8}$ Ry differing from that found by an exact solution, which is possible for H_2^+, by less than the final standard deviation. (The code which found the exact result of -1.0060549 Ry is due to Hyun-Joon Shin [7] using the methods of Barber and Hasse [8, 9].)

The spin-aligned H_2 wave function is

$$\psi = A_{A,B;1,2}\{e^{-r_{1A}-r_{2B}}[\tfrac{1}{2} + v_p(r_{1B}, r_{12})(a_1 + a_2(r_{12} - r_{AB}) + a_3 r_{1A} + a_4 r_{2B} + a_5 r_{1A} r_{2B} + a_6 r_{2B}^2 + a_7(r_{12} - r_{AB})^2)]\} \tag{34}$$

where $A_{A,B;1,2}$ is an operator which antisymmetrizes with respect to nuclei A and B and electrons 1 and 2. The term v_p is given by

$$v_p(r_{1B}, r_{12}) = s(r_{12})/r_{12} - 2s(r_{1B})/r_{1B} + 1/r_{AB} \tag{35}$$

with

$$s(x) = 1 - 4(1 - x)_+^3/3 + 8(0.5 - x)_+^3/3 \tag{36}$$

where

$$\begin{aligned} (x)_+ &= 0, \qquad x < 0 \\ &= x, \qquad x > 0. \end{aligned} \tag{37}$$

This form is suggested by the fact that the Hamiltonian for two separated atoms can be written as

$$\begin{aligned} H = &-\nabla_1^2 - 2/r_{1A} + (1/r_{12} - 2/r_{1B} + 1/r_{AB}) \\ &-\nabla_2^2 - 2/r_{2B} + (1/r_{12} - 2/r_{2A} + 1/r_{AB}) \end{aligned} \tag{38}$$

where the terms in parentheses are small for electron 1 close to nucleus A and electron 2 close to nucleus B. The splines in v_p remove the singularities while leaving v_p identical to the term in parentheses for r_{12} and r_{1B} on the order of 7.8 a_B.

The parameters a_1 through a_7 were found in two stages. First σ^2 was minimized with respect to the a's using $N = 400$. The resulting ψ had

TABLE IV

The Coefficients in $\psi(\mathbf{r}_1, \mathbf{r}_2)$ (Eq. (34))

i	1	2	3	4
a_i	-3.024050	-3.872121×10^{-2}	-4.98549×10^{-2}	-3.965802×10^{-2}
i	5	6	7	
a_i	-1.501606×10^{-2}	5.380148×10^{-3}	3.331075×10^{-3}	

$\langle E \rangle = -2 + (-3.36 \pm 0.17) \times 10^{-5}$ Ry and $\sigma \approx 1.4 \times 10^{-3}$ Ry. This ψ was used to find a set of 256 configurations using the pairing method, plus another 144 "worst" set of configuration picked from 28,000 evaluations of $H\psi/\psi$. These configurations provide a capsulized picture of those areas where ψ is not as good as it could be and allow more minimization of E over a small set of points than would otherwise be possible. Then $\langle E \rangle + 25\sigma$ was minimized with respect to these 400 configurations. The resulting parameters are given in Table IV. For this ψ, $\sqrt{N}\sigma = 2.54 \times 10^{-3}$ Ry and when finally evaluated in the process of this work $\langle E \rangle = -2 + (-3.65 \pm 0.10) \times 10^{-5}$ Ry, which should be compared with the value of -4.09×10^{-5} Ry calculated by Kolos and Wolniewicz [10] using a 60-parameter wave function.

References

1. J. M. Hammersly and D. C. Hanscomb, "Monte Carlo Methods," pp. 60–66, Chapman & Hall, London, 1964.
2. C. B. Haselgrove, *Math. Comp.* **15** (1961), 323.
3. S. A. Burdick, *J. Comput. Chem.* **3** (1982), 117.
4. R. L. Coldwell, *Int. J. Quantum Chem.* **S11** (1977), 215.
5. P. R. Bevington, "Data Reduction and Error Analysis for the Physical Sciences," p. 187, McGraw–Hill, New York, 1969.
6. R. L. Coldwell and R. E. Lowther, *Int. J. Quantum Chem.* **S12** (1978), 329.
7. Hyun-Joon Shin, Private communication.
8. W. G. Barber and H. R. Hasse, *Proc. Cambridge Philos. Soc.* **31** (1935), 564.
9. D. R. Bates, K. Ledsham, and A. L. Stewart, *Philos. Trans. R. Soc. London, Ser. A* **246** (1953), 215.
10. W. Kolos and L. Wolniewicz, *Chem. Phys. Lett.* **24** (1974), 457.