

## A combinatorial approach to the electron correlation problem

Alex J. W. Thom and Ali Alavi<sup>a)</sup>

Chemistry Department, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

(Received 21 July 2005; accepted 15 September 2005; published online 22 November 2005)

Starting from a path-integral formulation of quantum statistical mechanics expressed in a space of Slater determinants, we develop a method for the Monte Carlo evaluation of the energy of a correlated electronic system. The path-integral expression for the partition function is written as a contracted sum over graphs. A graph is a set of distinct connected determinants on which paths can be represented. The weight of a graph is given by the sum over exponentially large numbers of paths which visit the vertices of the graph. We show that these weights are analytically computable using combinatorial techniques, and they turn out to be sufficiently well behaved to allow stable Monte Carlo simulations in which graphs are stochastically sampled according to a Metropolis algorithm. In the present formulation, graphs of up to four vertices have been included. In a Hartree-Fock basis, this allows for paths which include up to sixfold excitations relative to the Hartree-Fock determinant. As an illustration, we have studied the dissociation curve of the  $N_2$  molecule in a VDZ basis, which allows comparison with full configuration-interaction calculations. © 2005 American Institute of Physics. [DOI: 10.1063/1.2114849]

### I. INTRODUCTION

It is well known that the path-integral formulation of quantum statistical mechanics provides, in principle, a method to evaluate the energy of  $N$  interacting electrons.<sup>1</sup> In its “real-space imaginary-time” formulation, the path integral represents a sum over paths in which one propagates the world lines of  $N$  electrons for an imaginary time  $i\hbar\beta$  from a given point  $\mathbf{X}=(\mathbf{x}_1, \dots, \mathbf{x}_N)$  in the  $N$ -electron configuration space back to the same point up to a permutation of the electron labels,  $\hat{P}\mathbf{X}=(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_N})$ . A given path can be described by a continuous function,  $\mathbf{X}(\tau)$ , with  $\tau$  running from 0 to  $\hbar\beta$ . Each such path carries an associated Boltzmann weight determined by the Hamiltonian of the system, multiplied by the sign of the permutation:

$$w[\mathbf{X}(\tau); \mathbf{X}, \hat{P}\mathbf{X}] = \text{sign}(\hat{P}\mathbf{X}) \left( \exp\left[-(1/\hbar) \int_0^{\hbar\beta} (m|\dot{\mathbf{X}}|^2/2 + V[\mathbf{X}(\tau)]) d\tau\right] \right)_{\mathbf{X}(\hbar\beta)=\hat{P}\mathbf{X}(0)}. \quad (1)$$

The partition function is defined as the functional integral of the weight over all such paths, for all permutations of end points and for all initial positions  $\mathbf{X}$ :

$$Q = \frac{1}{N!} \int d\mathbf{X} \sum_{\hat{P}} \mathcal{D}\mathbf{X}(\tau) w[\mathbf{X}(\tau); \mathbf{X}, \hat{P}\mathbf{X}]. \quad (2)$$

The expectation value of the energy at a finite  $\beta=(kT)^{-1}$  can be found through the relation

<sup>a)</sup> Author to whom correspondence should be addressed. Electronic mail: asa10@cam.ac.uk

$$E = -\frac{\partial \ln Q}{\partial \beta} \quad (3)$$

and takes the form

$$E = \frac{1}{Q} \frac{1}{N!} \int d\mathbf{X} \sum_{\hat{P}} \int \mathcal{D}\mathbf{X}(\tau) E[\mathbf{X}(\tau)] w[\mathbf{X}(\tau); \mathbf{X}, \hat{P}\mathbf{X}], \quad (4)$$

where  $E[\mathbf{X}(\tau)]$  is an energy estimator derived from the Hamiltonian of the system.<sup>2</sup> In numerical implementation, the continuous path is discretized into a number  $P$  of time slices:

$$\mathbf{X}(\tau) \rightarrow \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_P, \hat{P}\mathbf{X}_1. \quad (5)$$

Expectation values of the type (4) can be computed by averaging over such paths, with the caveat that the probability to generate a given path is proportional to the absolute value of this weight, leading to

$$E = \frac{\langle \text{sign}(w) E[\mathbf{X}] \rangle_{|w|}}{\langle \text{sign}(w) \rangle_{|w|}}. \quad (6)$$

However, the “minus-sign” problem gives rise to a serious difficulty since an exponential cancellation occurs between the even and odd-permuted paths as either  $\beta$  or  $N$  become large, making the denominator in Eq. (6) very difficult to estimate. An enormous effort has been devoted to solving this problem. The best current methods circumvent this in an approximate way through imposition of constraints in the generation of paths. For example, in the restricted path-integral method of Ceperley,<sup>3</sup> the paths are restricted by the nodes of a trial density matrix. This is the path-integral version of the fixed-node approximation of the zero-temperature diffusion Monte Carlo method.<sup>4</sup> Release-node methods are in principle one way forward and have been applied to small

molecules.<sup>5</sup> However, this method does not seem practical for larger numbers of electrons. The motivation for developing quantum Monte Carlo methods is the prospect of a correlated method with an  $N^3$  scaling with system size,<sup>6</sup> i.e., much more favorable than wave-function-based correlated methods such as coupled cluster [CCSD or CCSD(T)], which typically scale as  $\sim N^6$  or  $N^7$ . The sign problem, however, poses the ultimate difficulty.

In this paper, we formulate the path-integral method in Slater determinant space, which allows the partition function to be cast as a trace over an antisymmetric space. By performing analytic summations, we can contract the partition function to express it as a new sum over “graphs,” rather than “paths.” The advantage of this is that the wildly fluctuating sign problem associated with path integral is replaced with a much more controlled method in which graphs—rather than paths—are stochastically sampled. There is no need for a fixed-node-type approximation. In addition, the method has a favorable scaling with number of electrons and size of basis. In Secs. II and III the general formulation is presented. Sections IV and V deal with the combinatorial techniques required for the analytic summations. Section VI deals with some properties of graphs. Section VII reports some numerical results of the method as applied to the binding-energy curve of the  $H_2$  and  $N_2$  molecules.

## II. SLATER DETERMINANT SPACE

Although the continuous-space representation of path integrals is elegant because it obviates the use of a basis, for the purposes of battling the sign problem, it may be greatly advantageous to work in a finite, discrete, space. Given a set of one-electron spin-orbitals  $u_1(\mathbf{x}), u_2(\mathbf{x}) \cdots u_{2M}(\mathbf{x})$ , it is possible to construct a finite but still exponentially large antisymmetrized  $N$ -electron space using the Slater determinant formalism. Indeed, given  $2M$  spin-orbitals, the Slater determinant space of  $N$  electrons contains  $N_{\text{det}} = \binom{2M}{N}$  determinants. Apart from a certain loss of accuracy which inevitably arises through the use of a finite basis, there is no difficulty in extending the concept of a path to this basis. If we denote  $D_{\mathbf{i}}(\mathbf{X})$  to be a Slater determinant (the label  $\mathbf{i}$  being an ordered  $N$ -tuple of integers):

$$D_{\mathbf{i}} \equiv D_{i_1 i_2 \dots i_N} = \frac{1}{\sqrt{N!}} \det[u_{i_1} \dots u_{i_N}], \quad (7)$$

then a closed path of  $P$  steps in Slater determinant space consists of the sequence

$$D_{\mathbf{i}_1}, D_{\mathbf{i}_2}, \dots, D_{\mathbf{i}_p}, D_{\mathbf{i}_1}, \quad (8)$$

which is analogous to the discretization in of  $\mathbf{X}(\tau)$  in (5). The associated Boltzmann weight for this path in determinant space is

$$w^{(P)}[\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_p, \mathbf{i}_1] = \langle D_{\mathbf{i}_1} | e^{-\beta \hat{H}/P} | D_{\mathbf{i}_2} \rangle \times \langle D_{\mathbf{i}_2} | e^{-\beta \hat{H}/P} | D_{\mathbf{i}_3} \rangle \dots \langle D_{\mathbf{i}_p} | e^{-\beta \hat{H}/P} | D_{\mathbf{i}_1} \rangle, \quad (9)$$

which is analogous to Eq. (1). In Eq. (9),  $\hat{H}$  is the nonrela-

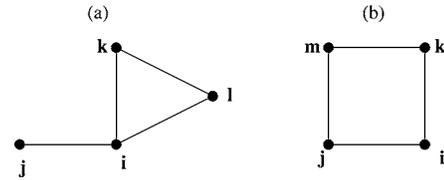


FIG. 1. Two examples of four-vertex graphs. The weight  $w^{(4)}[G_a]$  represents the sum over all paths of length  $P$  which visit the nodes  $\{\mathbf{i}, \mathbf{j}, \mathbf{k}, \mathbf{l}\}$  and  $w^{(4)}[G_b]$  represents the sum over all paths which visit  $\{\mathbf{i}, \mathbf{j}, \mathbf{k}, \mathbf{m}\}$ . Paths which visit only the intersection set  $G_a \cap G_b = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$  are not included in the above weights, but only in  $w^{(3)}[G_a \cap G_b]$ , i.e., a smaller graph.

tivistic Schrödinger Hamiltonian of the system containing all interactions. In our view, three major advantages accrue from this change of representation. The first is that the antisymmetry principle is inserted from the very outset in the basis, and therefore one does not need to perform explicit averaging over permuted (exchanging) paths. Traces over this space are therefore automatically traces over purely antisymmetric parts, as is appropriate for a Fermion system:

$$Q = \text{Tr}[e^{-\beta \hat{H}}] = \sum_{\mathbf{i}} \langle D_{\mathbf{i}} | e^{-\beta \hat{H}} | D_{\mathbf{i}} \rangle, \quad (10)$$

$$= \sum_{\mathbf{i}_1} \sum_{\mathbf{i}_2} \cdots \sum_{\mathbf{i}_p} w^{(P)}[\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_p, \mathbf{i}_1]. \quad (11)$$

On the other hand, one should note that the sign of  $w^{(P)}$  in Eq. (9) is not positive definite, and indeed a very poorly behaved quantity since it depends on the product of the signs of  $P$  factors, implying small variations in the path, can produce wild fluctuations in the sign of the path. Therefore it is not clear if the use of this basis alone will help in solving the sign problem.

This brings us to the second major advantage of working in a finite space, which is the principal topic of this paper. There exists the explicit possibility to sum analytically over exponentially large numbers of paths to give rise to new objects that we shall refer to as graphs. An  $n$ -vertex graph  $G$  consists of a set of  $n$  distinct determinants:

$$G = \underbrace{\{D_{\mathbf{i}}, D_{\mathbf{j}}, D_{\mathbf{k}}, \dots\}}_{n \text{ determinants}} \quad (12)$$

in no particular sequence.<sup>7</sup> The connectivity of the graph is determined by  $\rho_{\mathbf{ij}} = \langle D_{\mathbf{i}} | e^{-\beta \hat{H}/P} | D_{\mathbf{j}} \rangle$ . The size of the graph,  $n$ , can vary from 1 to  $\min(P, N_{\text{det}})$ . A graph  $G$  is an object on which one can represent the paths which visit exclusively all vertices in  $G$ . The weight  $w^{(n)}[G]$  of a given graph is obtained by summing over all such paths of length  $P$ :

$$w^{(n)}[G] = \sum_{\mathbf{i}_1 \in G} \sum_{\mathbf{i}_2 \in G} \sum_{\mathbf{i}_3 \in G} \cdots \sum_{\mathbf{i}_p \in G} ' w^{(P)}[\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_p, \mathbf{i}_1]. \quad (13)$$

The “ $'$ ” indicates that the summation indices  $\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_p$ , etc., are chosen from the set  $G$  in such a way that each vertex in  $G$  is visited at least once. This condition ensures that the weights of two different graphs  $G_a$  and  $G_b$  (i.e., two graphs which differ in at least one vertex) do not include the same paths which visit only the vertices of the set  $G_a \cap G_b$  (see Fig. 1). Therefore, the sum  $w[G_a] + w[G_b]$  will not double

count the contribution of such paths. As a result, the partition function can be represented not as a “sum over paths” but rather a contracted “sum over graphs”:

$$Q = \sum_n \sum_G w^{(n)}[G]. \quad (14)$$

In the above, the sum over  $n$  represents a sum over one-vertex, two-vertex, three-vertex, etc., graphs, and for each  $n$ , the sum over  $G$  represents the sum over all  $n$ -vertex graphs which can be constructed in the system. Since each graph represents a sum over many paths of (typically) alternating sign, the weight of each graph can be expected to be a much better behaved function than that of individual paths [Eq. (9)].

Given the expression Eq. (14) for the partition function, one can now reexpress  $E$  in a manner suitable for stochastic (Monte Carlo) method in which graphs are sampled with an appropriate probability and an energy estimator is averaged. Using Eq. (3), we have

$$E = \frac{1}{Q} \sum_n \sum_G -\frac{1}{w^{(n)}[G]} \frac{\partial w^{(n)}[G]}{\partial \beta} \cdot w^{(n)}[G]. \quad (15)$$

Therefore, if graphs can be sampled with an un-normalized probability given by  $w^{(n)}[G]$  (using a Metropolis Monte Carlo method), then the energy estimator becomes

$$\tilde{E}^{(n)}[G] = -\frac{\partial \ln w^{(n)}[G]}{\partial \beta}, \quad (16)$$

$$E = \langle \tilde{E}^{(n)}[G] \rangle_{w^{(n)}[G]}. \quad (17)$$

In other words, if on step  $t$  of a Monte Carlo simulation consisting of  $K$  steps one is at graph  $G_t$ , then the energy is given by the running average of  $\tilde{E}^{(n)}[G]$ :

$$E = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_t^K \tilde{E}^{(n)}[G_t]. \quad (18)$$

In order to perform a Metropolis Monte Carlo simulation, one needs to ensure that microscopic reversibility is satisfied, and this requires some care. In the present implementation, the graphs were generated stochastically according to an algorithm given in Appendix A. This algorithm generates a fresh graph at each Monte Carlo step. In addition one needs to compute the generation probability of the graph using this algorithm (see Appendix A), in order to unbiased the acceptance ratios for the Metropolis Monte Carlo method. Thus the probability to accept a trial graph  $G'$ , given the current graph  $G$ , is

$$P_{\text{acc}}[G'|G] = \min\left(1, \frac{w^{(n)}[G'] P_{\text{gen}}[G]}{P_{\text{gen}}[G'] w^{(n)}[G]}\right). \quad (19)$$

If the weights  $w^{(n)}[G]$  are positive definite, then the Metropolis Monte Carlo can proceed as normal. No sign problem is encountered and the graphs can be sampled quite straightforwardly. However, this condition cannot be guaranteed for an arbitrary graph at an arbitrary  $\beta$ . Nevertheless, for reasons that will become clear later on, this turns out not to

be catastrophic. Where the weights  $w^{(n)}[G]$  are not positive definite, one has to use the equivalent of Eq. (6):

$$E = \frac{\langle \text{sign}(w^{(n)}[G]) \tilde{E}^{(n)}[G] \rangle_{|w|}}{\langle \text{sign}(w^{(n)}[G]) \rangle_{|w|}}. \quad (20)$$

For this expression to be useful, the denominator  $\langle \text{sign}(w^{(n)}[G]) \rangle_{|w|}$  must be well behaved as  $\beta$  becomes large. In practice this means that the number of sampled “positive” graphs should exceed the the number of “negative” graphs in such a way that does not vanish with increasing  $\beta$  or  $N$ , making this expectation value feasible to compute in practice. One of the purposes of numerical calculations presented in this paper is to investigate this.

### III. SINGLE-REFERENCE FORMULATION

The third major advantage of using a Slater determinant formalism concerns the connection with more traditional quantum chemistry methods and, in particular, the Hartree-Fock method. Thus far the method has been proposed as a finite-temperature technique, with the ground-state energy being sought in the limit of  $\beta \rightarrow \infty$ . Where one is specifically interested in the ground-state energy, it is desirable to accelerate convergence with respect to  $\beta$ . As is shown below this can be achieved by working with an appropriate reference. In this case  $\beta$  loses its strict thermodynamic meaning and becomes a convenient calculation parameter. In fact, one can directly obtain the correlation energy of the ground state.

Suppose the one-particle orbitals  $\{u_1(\mathbf{x}), \dots, u_{2M}(\mathbf{x})\}$  are the Hartree-Fock orbitals (occupied and virtual), and let  $D_i$  be the Hartree-Fock determinant whose energy is  $E_{\text{HF}} = \langle D_i | \hat{H} | D_i \rangle$ . In this basis, the exact (full configuration-interaction) ground-state wave function  $\Psi_0$  of  $\hat{H}$  can be expressed as  $\Psi_0 = D_i + \sum_{j \neq i} C_j D_j$ , where the coefficients  $C_j$  are obtained by solving the Schrödinger equation  $(\hat{H} - E_0)\Psi_0 = 0$ . The correlation energy is  $E_0 - E_{\text{HF}}$ . We seek a procedure which will deliver the correlation energy without any attempt to calculate  $\Psi_0$ .

Consider the matrix element of the Boltzmann operator

$$w_i = \langle D_i | e^{-\beta \hat{H}} | D_i \rangle \quad (21)$$

and the function  $\tilde{E}_i(\beta)$  defined as

$$\tilde{E}_i(\beta) = -\frac{\partial \ln w_i}{\partial \beta}. \quad (22)$$

As a function of  $\beta$ ,  $\tilde{E}_i$  satisfies two interesting limits:

$$\lim_{\beta \rightarrow 0} \tilde{E}_i = \langle D_i | \hat{H} | D_i \rangle = E_{\text{HF}}, \quad (23)$$

$$\lim_{\beta \rightarrow \infty} \tilde{E}_i = E_0, \quad (24)$$

where the latter is true as long as  $\langle D_i | \Psi_0 \rangle \neq 0$ . By transforming to the exact eigenvalue basis of the Hamiltonian it is also easy to see that

$$\tilde{E}_i \geq E_0. \quad (25)$$

In other words, the range of variation of  $\tilde{E}_i$  as a function of  $\beta$  is bracketed from above by the Hartree-Fock energy and from below by the exact ground-state energy (the full configuration-interaction result). Therefore the range of variation of  $\tilde{E}_i$  is precisely the correlation energy.

The charm of Eq. (22) is that it provides a way to calculate the correlation energy without reference to a wave function. The expression for  $w_i$  in Eq. (21) can be cast as a graph sum, analogous to Eq. (14):

$$w_i = \sum_n \sum_G w_i^{(n)}[G], \quad (26)$$

where  $w_i^{(n)}[G]$  is defined analogously to Eq. (13), except for the fact that  $D_i$  is considered the beginning and end points of paths:

$$w_i^{(n)}[G] = \sum_{i_2 \in G} \sum_{i_3 \in G} \cdots \sum_{i_p \in G} w^{(p)}[\mathbf{i}, \mathbf{i}_2, \dots, \mathbf{i}_p, \mathbf{i}]. \quad (27)$$

In Eq. (27) it is required that  $G$  contains  $D_i$  among its members. Using Eqs. (22) and (26) leads to an expression for  $\tilde{E}_i$  suitable for a stochastic sampling of graphs, analogous to Eq. (16):

$$\tilde{E}_i = \langle \tilde{E}_i^{(n)}[G] \rangle_{w_i^{(n)}[G]} \quad (28)$$

or in the case where  $w_i^{(n)}[G]$  is not positive definite:

$$\tilde{E}_i = \frac{\langle \text{sign}(w_i^{(n)}[G]) \tilde{E}_i^{(n)}[G] \rangle_{|w|}}{\langle \text{sign}(w_i^{(n)}[G]) \rangle_{|w|}}. \quad (29)$$

One might expect the convergence of Eq. (26) with respect to the graph size  $n$  to be reasonably rapid. Starting from the Hartree-Fock determinant, two-vertex graphs sum over paths involving double-excitations, three-vertex graphs include those with triple and quadruple excitations, and in general the  $n$ -vertex graph include paths with up to  $2(n-1)$ -fold excitations. From a physical view point, it is reasonable to expect that a lot of correlations can be captured with fairly small graphs. Here the inverse-temperature-like parameter  $\beta$  plays a useful role. As we shall see in Sec. IV, for small  $\beta$ , the convergence with respect to  $n$  is rapid, since at small  $\beta$ , the paths cannot move far from  $i$ . At large  $\beta$  the convergence at a given graph size will be poor, and unphysical results will ensue. However, one can hope that meaningful results could be extracted at an intermediate  $\beta$  before convergence breakdown. One should therefore increase  $\beta$  starting from a small value and monitor the energy. We return to this later in actual applications of the formalism.

Evidently the first major challenge here is to perform summations indicated in Eq. (13) or Eq. (27), in order to obtain explicit analytic forms for  $w^{(n)}[G]$  and its single-reference variant  $w_i^{(n)}[G]$ . Fortunately, this turns out to be possible, thanks to a powerful combinatorial identity, which forms the topic of the next section.

#### IV. COMBINATORIAL ANALYSIS

To simplify the notation, let us denote high-temperature density matrix  $e^{-\beta\hat{H}/P}$  by  $\rho$ , and an element of it by  $\rho_{ij} \equiv \langle D_i | e^{-\beta\hat{H}/P} | D_j \rangle$ . For a sufficiently large  $P$ , the elements of  $\rho$  can be computed to high accuracy. If we set  $\hat{H} = \hat{F} + \hat{V}$ , where  $\hat{F}$  is the Fock operator (and therefore diagonal in the Hartree-Fock basis) and  $\hat{V} = \hat{H} - \hat{F}$  is the residual interaction, then to order  $(\beta/P)$ :

$$\rho_{ij} = \langle D_i | \rho | D_j \rangle = e^{-\beta(E_i + E_j)/2P} [\delta_{ij} - (\beta/P) \langle D_i | \hat{V} | D_j \rangle] + \mathcal{O}[(\beta/P)^2], \quad (30)$$

where  $E_j = \langle D_j | \hat{F} | D_j \rangle$  are the matrix elements of the Fock operator in the Hartree-Fock basis. To this order, therefore,  $\rho_{ij}$  has the same connectivity as the Hamiltonian, i.e., since  $\hat{H}$  contains at most two-body interactions,  $\rho_{ij}$  is nonzero only between determinants which differ by two or fewer orbitals. Equation (30) also shows that the  $\rho$  matrix is diagonally dominant.

In terms of the elements of  $\rho$ , we can write

$$w_i = \langle D_i | \rho^P | D_i \rangle = \sum_j \sum_k \cdots \sum_l \underbrace{\rho_{ij} \rho_{jk} \cdots \rho_{li}}_{P \text{ steps}}. \quad (31)$$

A typical path in Eq. (31) consists of  $P$  steps, which will involve a number of ‘‘hops’’ in which one moves from one determinant to another connected one, e.g.,  $D_i \rightarrow D_j$ , incurring a factor  $\rho_{ij}$ , and a number of ‘‘stay puts,’’ which incur diagonal factors such as  $\rho_{ii}$  and  $\rho_{jj}$ . Since the  $\rho$  matrix is diagonally dominant, paths which involve many hops will have a smaller weight than those with relatively few hops. Consider the following rearrangement, which amounts to re-writing the sum, Eq. (31), as sum over paths with a certain numbers of hops between connected determinants:

$$w_i = \rho_{ii}^P + \sum_{j \neq i} \sum_{n=0}^{P-2} \sum_{m=0}^{P-2-n} \rho_{ii}^n \rho_{ij} \rho_{jj}^m \rho_{ji} \rho_{ii}^{P-2-m-n} + \sum_{j \neq k \neq i} \sum_{n=0}^{P-3} \sum_{m=0}^{P-3-n} \sum_{l=0}^{P-3-n-m} \rho_{ii}^n \rho_{ij} \rho_{jj}^m \rho_{jk} \rho_{kk}^l \rho_{ki} \rho_{ii}^{P-3-n-m-l} + \cdots \quad (32)$$

Noting that we are getting nested sums, it proves useful to define a function  $Z_h^{(P)}(x_1, x_2, \dots, x_h)$  as follows:

$$Z_h^{(P)}(x_1, x_2, \dots, x_h) = \sum_{n_1=0}^{P-h} \sum_{n_2=0}^{P-h-n_1} \cdots \sum_{n_h=0}^{P-h-\sum_{i=1}^{h-1} n_i} x_1^{n_1} x_2^{n_2} \cdots x_h^{n_h}. \quad (33)$$

We will shortly discuss the properties of such nested sums. Let us for the moment note that we can cast Eq. (32) as follows:

$$\begin{aligned}
w_i = & \rho_{ii}^P \left( 1 + \sum_j \frac{\rho_{ij}\rho_{ji}}{\rho_{ii}^2} Z_2^{(P)} \left( 1, \frac{\rho_{ij}}{\rho_{ii}} \right) + \sum_j \sum_k \frac{\rho_{ij}\rho_{jk}\rho_{ki}}{\rho_{ii}^3} \right. \\
& \times Z_3^{(P)} \left( 1, \frac{\rho_{ij}}{\rho_{ii}}, \frac{\rho_{kk}}{\rho_{ii}} \right) + \sum_j \sum_k \sum_l \frac{\rho_{ij}\rho_{jk}\rho_{kl}\rho_{li}}{\rho_{ii}^4} \\
& \left. \times Z_4^{(P)} \left( 1, \frac{\rho_{ij}}{\rho_{ii}}, \frac{\rho_{kk}}{\rho_{ii}}, \frac{\rho_{ll}}{\rho_{ii}} \right) + \dots \right). \quad (34)
\end{aligned}$$

In the sums above, each sum is prime with respect to (i.e., excludes) the index of the previous sum, except for the innermost sum, which is also prime with respect to  $\mathbf{i}$ . For example, in the four-hop term,  $\mathbf{k}$  can equal  $\mathbf{i}$ , but not  $\mathbf{j}$ . Similarly,  $\mathbf{l}$  can equal  $\mathbf{j}$ , but not  $\mathbf{k}$  or  $\mathbf{i}$ .

Physically,  $Z_h^{(P)}$  is a combinatorial factor which accounts for the number of ways the  $P-h$  stay-put terms can be arranged for a given number of hops  $h$ . The rationale for this type of expansion is due to the diagonal dominance of the  $\rho_{ij}$  matrix. Every time we hop, we incur a penalty; therefore paths which involve many hops each carry an exponentially decreasing weight. On the other hand, the number of such paths increases exponentially. Therefore, the rate of convergence of the expansion above is difficult to predict. At high temperatures (or weak coupling) it will be fast (the off-diagonal  $\rho_{ij}$  are small) and otherwise slow. Our aim will be further summations in Eq. (34) to produce a rapidly convergent series.

Using induction,<sup>8</sup> one can prove the following identity:

$$Z_h^{(P)}(x_1, \dots, x_h) = \frac{1}{2\pi i} \oint_C \frac{z^P - 1}{(z-1)\prod_{\alpha}(z-x_{\alpha})} dz, \quad (35)$$

where  $C$  is a contour which encloses the poles  $(x_1, \dots, x_h)$ . This contour integral has many useful properties which we shall shortly enjoy.<sup>9</sup> For example, if the poles in the denominator are simple,<sup>10</sup> the contour integral can be immediately done using the residue theorem, to obtain<sup>11</sup>

$$Z_h^{(P)}(x_1, \dots, x_h) = \sum_{i=1}^h \frac{x_i^P - 1}{(x_i - 1)\prod_{j \neq i}(x_i - x_j)}. \quad (36)$$

However, it proves expedient *not* to evaluate the  $Z$  sums explicitly yet, but to leave them in contour-integral form. We can now sum over paths with different numbers of hops. For example, consider three cycles  $\mathbf{i} \rightarrow \mathbf{j} \rightarrow \mathbf{k} \rightarrow \mathbf{i}$  involving the transition matrix elements  $\rho_{ij}\rho_{jk}\rho_{ki}$ . These three-cycles (whose weight can be negative) will be involved in the following three-hop, six-hop, etc., paths which form a typical alternating series:

$$S = \frac{\rho_{ij}\rho_{jk}\rho_{ki}}{\rho_{ii}^3} Z_3 + \left( \frac{\rho_{ij}\rho_{jk}\rho_{ki}}{\rho_{ii}^3} \right)^2 Z_6 + \dots$$

Defining

$$A_{ijk}(z) = \frac{\rho_{ij}\rho_{jk}\rho_{ki}}{(z\rho_{ii} - \rho_{ii})(z\rho_{ii} - \rho_{jj})(z\rho_{ii} - \rho_{kk})}, \quad (37)$$

we obtain

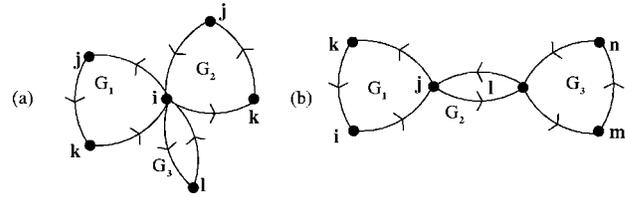


FIG. 2. (a) An example of a star graph, in which three distinct cyclic paths denoted by  $G_1, G_2$ , and  $G_3$  have been linked by the pivot  $\mathbf{i}$ . In this example,  $G_1$  and  $G_2$  contain the same vertices  $\mathbf{i}, \mathbf{j}$  and  $\mathbf{k}$ , but correspond to different directions of going around three cycles. In star graphs composed out of distinct cyclic paths, the execution of any given cycle can be done independently of the others. As a result, if each cycle  $G_i$  is to be executed a certain of times  $n_i$ , then the total number of the ways this can be achieved is given by the multinomial coefficient in Eq. (40). (b) An example of a chain graph composed out of three cyclic paths  $G_1, G_2$ , and  $G_3$ . In this case cycle  $G_2$  can be accessed only through  $G_1$ , and similarly  $G_3$  from  $G_2$ . The total number of ways in which each cycle  $G_i$  can be executed  $n_i$  times is given by Eq. (43).

$$S = \sum_n \frac{1}{2\pi i} \oint_C \frac{z^P - 1}{z - 1} [A_{ijk}(z)]^n. \quad (38)$$

Switching the order of summation and integration (the contour can always be deformed to keep away from singularities), we obtain

$$S = \frac{1}{2\pi i} \oint_C \frac{z^P - 1}{z - 1} \frac{1}{1 - A_{ijk}(z)}. \quad (39)$$

Since  $1 - A_{ijk}(z) = 0$  is a cubic equation in  $z$ , it has three roots, which can be used to evaluate Eq. (39) in terms of three residues, analogous to Eq. (36).

The above is a simple example of the analytic summation over an alternating series. One can similarly treat more complex paths, which consist of sums of *combinations of cyclic paths*. In these cases, combinatorial reasoning is called for. Consider, first, the *star graphs*, which consist of distinct cyclic paths,  $G_1, G_2, \dots, G_g$ , all of which are pivoted about the vertex  $\mathbf{i}$  [Fig. 2(a)]. Assuming that we execute cycle  $G_1$  a number ( $n_1$ ) times, cycle  $G_2$  a number ( $n_2$ ) times, etc., the complete contribution is found to be

$$S_i^{\text{star}} = \sum_{n_1, \dots, n_g} \frac{(n_1 + n_2 + \dots)!}{n_1! n_2! \dots} \frac{1}{2\pi i} \oint_C \frac{z^P - 1}{z - 1} A_{G_1}^{n_1} A_{G_2}^{n_2} \dots \quad (40)$$

$$= \frac{1}{2\pi i} \oint_C \frac{z^P - 1}{z - 1} \frac{1}{(1 - A_{G_1} - A_{G_2} - \dots - A_{G_g})}, \quad (41)$$

where the multinomial coefficient accounts for the number of ways of selecting  $n_1, n_2, \dots, n_g$  out of  $n_1 + n_2 + \dots + n_g$ .

A special case of a star graph is the simple two-vertex graph  $\{\mathbf{i}, \mathbf{j}\}$ , in which the paths correspond to hopping between the two determinants  $D_i$  and  $D_j$ :

$$S_i^{(2)}[\{\mathbf{i}, \mathbf{j}\}] = \frac{1}{2\pi i} \oint_C \frac{z^P - 1}{z - 1} \frac{1}{1 - A_{ij}(z)}, \quad (42)$$

where  $A_{ij}(z) = \rho_{ij}\rho_{ji} / (z\rho_{ii} - \rho_{ii})(z\rho_{ii} - \rho_{jj})$  is defined analogously to Eq. (37).

Next, consider *chain graphs*, which consist of a sequential set of cyclic paths,  $G_1, G_2, \dots, G_g$ , starting from the vertex  $i$  [Fig. 2(b)]. By sequential we mean that the cycle  $G_i$  is accessible only through  $G_{i-1}$ , through a specified common vertex. The combinatorial factors must take into account this constraint. In general, for a chain of  $g$  sequential cyclic paths, the total number of ways that cycle  $G_1$  can be executed  $n_1$  times, cycle  $G_2$  can be executed  $n_2$  times, etc., is

$$\binom{n_1 + n_2 - 1}{n_2} \times \binom{n_2 + n_3 - 1}{n_3} \times \dots \quad (43)$$

and therefore

$$\begin{aligned} S_i^{\text{chain}} &= \sum_{n_1, \dots, n_g} \binom{n_1 + n_2 - 1}{n_2} \times \binom{n_2 + n_3 - 1}{n_3} \\ &\times \dots \frac{1}{2\pi i} \oint_C \frac{z^P - 1}{z - 1} A_{G_1}^{n_1} A_{G_2}^{n_2} \dots A_{G_g}^{n_g} \\ &= \frac{1}{2\pi i} \oint_C \frac{z^P - 1}{z - 1} \frac{1}{1 - A_{G_1} [1 - A_{G_2} (1 - \dots)]}, \quad (44) \end{aligned}$$

i.e., the kernel inside the contour integral is a continued fraction.

Using these results, one can derive the kernels for processes that combine both star-type and chain-type cyclic paths. Of particular importance are paths which can be constructed on the complete three-vertex graph, in which the vertices  $\{i, j, k\}$  are all connected [Fig. 3(a)]. To evaluate the sum for this graph, consider the possible distinct circuits on such a graph. There are four circuits, two of which can be

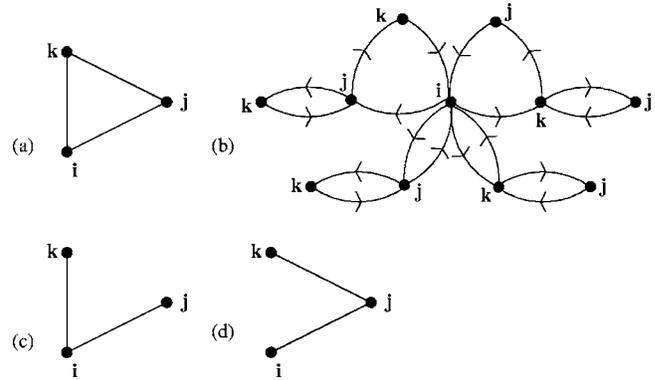


FIG. 3. (a) The complete three-vertex graph and (b) its unfolded representation about  $i$ . In this form, each of the four distinct circuits on the three-vertex graph is shown, forming a “star” around  $i$ . In general, the representation of a graph such as in (a) depicts the connectivity of the graph, whereas the unfolded representation shows the distinct circuits that can be constructed on the graph. (c) and (d) are the other two types of three-vertex graphs, respectively, a star and a “chain” relative to  $i$ . Both (c) and (d) are trees. In terms of the Hamiltonian, in (a) all three determinants  $\{i, j, k\}$  must be single or double excitations of each other. In (c) both  $j$  and  $k$  must be single or double excitations of  $i$ , but not necessarily of each other. In (d)  $k$  can be a triple or quadruple excitation of  $i$ .

considered as paths generated by chaining the two two-cycles, i.e.,  $(i, j)(j, k)$  and  $(i, k)(k, j)$ . The remaining two circuits involve three cycles and can be considered chains of the type  $(i, j, k)(j, k)$  where the common vertex can be chosen to be either  $j$  or  $k$ . Setting these chains onto a star, we obtain Fig. 3(b), which can be thought of as the “unfolded” representation of Fig. 3(a). The corresponding kernel in the contour integral can be written down by inspection:

$$\begin{aligned} S_i^{(3)}[\{i, j, k\}] &= \frac{1}{2\pi i} \oint_C \frac{z^P - 1}{z - 1} \frac{1}{1 - A_{ij}/(1 - A_{jk}) - A_{ik}/(1 - A_{jk}) - 2A_{ijk}/(1 - A_{jk})} \\ &= \frac{1}{2\pi i} \oint_C \frac{z^P - 1}{z - 1} \frac{1 - A_{jk}}{1 - A_{ij} - A_{ik} - A_{jk} - 2A_{ijk}}. \quad (45) \end{aligned}$$

The denominator of the three-vertex term,  $1 - A_{ij} - A_{ik} - A_{jk} - 2A_{ijk} = 0$ , reduces to a cubic polynomial in  $z$ , which can be solved to yield three roots, which are in turn used to evaluate the contour integral using the residue theorem.

There are two other types of three-vertex graphs which correspond to deleting an edge from the complete graph [Figs. 3(c) and 3(d)]. The corresponding kernel is obtained by setting, respectively, either  $A_{ij} = 0$  or  $A_{jk} = 0$ , together with the three-cycle  $A_{ijk} = 0$  in Eq. (45).

By a similar technique of enumerating the distinct circuits on a complete four-vertex graph, one can construct the unfolded representation and hence evaluate  $S_i^{(4)}[\{i, j, k, l\}]$ . This procedure is carried out in Appendix B. Although the resulting expression is quite unwieldy, the final form simpli-

fies remarkably to a quartic polynomial on the denominator of the contour integral, implying that at most four residues are required for its evaluation.

This approach should be extendable to graphs with a larger number of vertices. However, it is a difficult work. Given the substantial simplifications which may occur, it may be possible that an alternative procedure can be found in which the sums for larger graphs can be more easily evaluated. This is the topic of current work. For the present, we will show the results of calculations up to the four-vertex level, which are nevertheless very encouraging.

## V. DOUBLE COUNTING OF PATHS

Closer examination of the terms included in the sum for  $S_i^{(3)}$ , Eq. (45), shows that this includes all paths involving

$\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ , including paths which visit only  $\{\mathbf{i}, \mathbf{j}\}$  and  $\{\mathbf{i}, \mathbf{k}\}$  i.e., the two possible two-vertex terms. However, as discussed in Sec. II, what we need are sums over paths in which all vertices in a graph have been visited, i.e., the “pure”  $n$ -vertex terms,  $w_i^{(n)}$ . Fortunately this turns to be easily computable in an iterative fashion. First, note that  $S_i^{(2)}$  is already pure, i.e.,  $w_i^{(2)}[\{\mathbf{i}, \mathbf{j}\}] = S_i^{(2)}[\{\mathbf{i}, \mathbf{j}\}]$ . This is because in the calculation of

$$S_i^{(2)} = \sum_{n=0} (\rho_{ij}\rho_{ji})^n \frac{1}{2\pi i} \oint_C \frac{z^P - 1}{z - 1} \frac{1}{(z\rho_{ii} - \rho_{ii})^n (z\rho_{ii} - \rho_{jj})^n} dz \quad (46)$$

one can equally set the lower limit of the sum to  $n=1$  without changing the result: the  $n=0$  term leaves  $(z^P - 1)/(z - 1)$  which is analytic and therefore yields zero upon contour integration. As a result  $S_i^{(2)}$  sums over paths which execute at least one  $(\mathbf{i}, \mathbf{j})$  cycle.

With this in hand, to obtain  $w_i^{(3)}[\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}]$  from  $S_i^{(3)}[\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}]$ , we need to subtract out  $S_i^{(2)}[\{\mathbf{i}, \mathbf{j}\}]$  and  $S_i^{(2)}[\{\mathbf{i}, \mathbf{k}\}]$ :

$$w_i^{(3)}[\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}] = S_i^{(3)}[\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}] - S_i^{(2)}[\{\mathbf{i}, \mathbf{j}\}] - S_i^{(2)}[\{\mathbf{i}, \mathbf{k}\}]. \quad (47)$$

For higher-order graphs, a similar purification procedure can be done, although the effort is greater. For example, for the four-vertex graphs,  $S_i^{(4)}[\{\mathbf{i}, \mathbf{j}, \mathbf{k}, \mathbf{l}\}]$ , we need to subtract out the contribution of the three-vertex terms taking care to add back in terms from the two-vertex terms which have been removed twice:

$$w_i^{(4)}[\{\mathbf{i}, \mathbf{j}, \mathbf{k}, \mathbf{l}\}] = S_i^{(4)}[\{\mathbf{i}, \mathbf{j}, \mathbf{k}, \mathbf{l}\}] - S_i^{(3)}[\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}] - S_i^{(3)}[\{\mathbf{i}, \mathbf{j}, \mathbf{l}\}] - S_i^{(3)}[\{\mathbf{i}, \mathbf{k}, \mathbf{l}\}] + S_i^{(2)}[\{\mathbf{i}, \mathbf{j}\}] + S_i^{(2)}[\{\mathbf{i}, \mathbf{k}\}] + S_i^{(2)}[\{\mathbf{i}, \mathbf{l}\}]. \quad (48)$$

Equations (42), (47), and (48) form the working expressions from which the weights  $w_i^{(n)}[G]$  up to  $n=4$  can be calculated. [The one-vertex weight is, of course,  $w_i^{(1)} = \rho_{ii}^P$ .]

To obtain  $w^{(n)}[G]$  from  $w_i^{(n)}[G]$ , this can be achieved according to Eqs. (13) and (27) by summing over all  $\mathbf{i} \in G$ :

$$w^{(n)}[G] = \sum_{\mathbf{i} \in G} w_i^{(n)}[G]. \quad (49)$$

For small graphs containing only a few vertices, this does not pose much extra work. For proper finite-temperature work,  $w^{(n)}[G]$  should be used rather than the single-reference form.

## VI. SIGN OF GRAPHS

Although the sign of  $w^{(n)}[G]$  (or its single-reference variant  $w_i^{(n)}[G]$ ) of an arbitrary  $n$ -vertex graph  $G$  can be either positive or negative for a given  $\beta$ , certain graphs are positive definite for all  $\beta$ . The simplest case are all two-vertex graphs, which are positive definite since the combination  $\rho_{ij}$  always occurs in combination with  $\rho_{ji} = \rho_{ij}^\dagger$  for paths which return to  $\mathbf{i}$ . This argument can be generalized to graphs which do not have cycles (the so-called *trees* in graph-theoretical language). Of the three types of three-vertex graphs (Fig. 3), two are trees. The distribution of graphs among the types encountered in a physical system is not uniform, but depends on the connectivity of the  $\rho_{ij}$  matrix,

which, to order  $(\beta/P)$ , is determined directly by the elements of the Hamiltonian matrix  $H_{ij}$ . Therefore the three-vertex cyclic graph requires that all three determinants to be double excitations of each other, whereas for trees this restriction does not exist. As a result, the three-vertex trees greatly outnumber the three-vertex cyclic graphs. We observe a similar pattern with four-vertex graphs, where four-vertex trees are much more prevalent than four-vertex “cyclic” graphs (i.e., graphs containing cycles). For this to translate into a benefit for the Monte Carlo sampling, one needs to assess what fraction  $f$  of the sampled graphs  $G$  are trees ( $T$ ), positive cyclic graphs ( $C^+$ ), and negative cyclic graphs ( $C^-$ ). Useful diagnostics which measure these can be symbolically represented as

$$f_T = \langle \delta[G - T] \rangle_{|w|}, \quad (50)$$

$$f_{C^+} = \langle \delta[G - C^+] \rangle_{|w|}, \quad f_{C^-} = \langle \delta[G - C^-] \rangle_{|w|}.$$

Since  $\langle \text{sign}(w^{(n)}[G]) \rangle_{|w|} = f_T + f_{C^+} - f_{C^-}$  what is needed, therefore, is that

$$f_T + f_{C^+} \geq f_{C^-}. \quad (51)$$

If this condition is satisfied, then one can be confident that the Monte Carlo evaluation of  $\langle \text{sign}(w^{(n)}[G]) \rangle_{|w|}$  will be reliable, and that the energy can be evaluated according to Eq. (20) or Eq. (29).

## VII. APPLICATIONS

In this section we report on the results of evaluating with energy for two molecular systems,  $H_2$  and  $N_2$ , at their equilibrium geometries and at stretched geometries. The  $H_2$  molecule is presented as a simple illustration for which many exact quantities can be computed. Most of the emphasis will be on the  $N_2$  molecule, which, due to its multiple-bond character, represents a critical system: the dissociation of multiple bonds is known to pose great challenges to correlated methods. Being typical “chemical” systems where the electronic temperature is not itself of primary concern, we have used the single-reference version of the method, i.e., specifically to evaluate  $\tilde{E}_i(\beta)$  for the Hartree-Fock determinant at large  $\beta$ . Both have been done in basis sets for which the full configuration-interaction ground state have been computed. In addition, for  $H_2$ , the full configuration interaction (FCI) excited states were calculated, whence the exact  $\tilde{E}_i$ .

There are essentially two separate issues to address. The first regards the accuracy of a truncation of the graph-sum Eq. (26) at a particular vertex level. This issue must be addressed separately from a potential sign problem which could arise in a stochastic sampling of the graphs. In order to address these two issues, two types of calculations were performed. In the first, we compute  $w_i$  and  $\tilde{E}_i$  by truncating Eq. (26) at a particular vertex level such as  $n=2, 3$ , or  $4$ , but summing all possible graphs at this vertex level (the complete  $n$ -vertex sum). This delivers a correlation energy free of any statistical error, albeit with a systematic error that arises from the truncation. Such calculations display the convergence behavior of  $\tilde{E}_i$  as a function of  $\beta$  and the expected

accuracy given a level of truncation. The cost of summing over all  $n$ -vertex graphs grows as  $N^{2(n-1)}M^{2(n-1)}$ , where  $N$  is the number of electrons and  $M$  is the number of spatial orbitals. Therefore this method is limited to small systems. For example, for the  $N_2$  molecule treated with a VDZ basis, there are roughly  $3 \times 10^9$  four-vertex graphs containing the Hartree-Fock determinant, putting such a calculation at the limit of practicality. On a Pentium 4 processor a complete four-vertex sum for this system takes about one week of calculation.

The second set of calculations was to perform Monte Carlo sampling of graphs, up to the four-vertex level. The aim here is to establish the stability of this procedure with regards the sign problem. Comparison with the complete  $n$ -vertex sum allows us to judge its accuracy. The costliest part of the Monte Carlo is the evaluation of the matrix elements of  $\rho$  required to set up the polynomials which enter the contour-integral formulas; the subsequent evaluation of the contour integral requires numerical root-finding of small polynomials, which is very fast. Therefore the scaling behavior of the Monte Carlo is dominated by the calculation of the Hamiltonian elements. If the two-electron integrals have been computed over the Hartree-Fock orbitals in an initialization step (an  $\sim M^4$  process), then according to the Slater-Condon rules the calculation of the Hamiltonian matrix elements has only  $N^2$  scaling. The graphs were generated stochastically according to the algorithm given in Appendix A. This graph-generation strategy is effective for the small graphs (of up to four vertices) used in the present study. The acceptance probability of a new graph over the current one is given by Eq. (19). A typical Monte Carlo run of  $10^7$  cycles takes, in the case of VDZ  $N_2$ , a few hours on a Pentium 4 processor.

The restricted Hartree-Fock orbitals were computed on a Gaussian basis using MOLPRO.<sup>12,13</sup> The two-electron integrals over the Hartree-Fock orbitals were also computed at this stage. These, together with the one-electron integrals, were read into our program. The elements of the  $\rho$  matrix were then constructed “on the fly” using Eq. (30). In the calculation of  $\rho$ ,  $(\beta/P)$  was set to  $10^{-3}$ , and for a given  $\beta$ ,  $P$  was obtained through this ratio. Typically  $P$  varied from 100 to 50 000. We also set to zero any  $\rho_{ij}$  element whose magnitude was less than  $10^{-6}$ . Given the Trotter decomposition and the value of  $\beta/P$  used, this is a justifiable cutoff which negligibly affects the accuracy—something we tested—but which is useful since it speeds up the calculations significantly.

The  $H_2$  molecule was computed in a VTZ basis, with 56 spin-orbitals. This system is sufficiently small to be exactly diagonalized (including excited states), allowing for the exact calculation of  $\tilde{E}_1$  via

$$\tilde{E}_1^{(\text{exact})}(\beta) = \frac{\sum_a E_a |\langle D_i | \Psi_a \rangle|^2 e^{-\beta E_a}}{\sum_a |\langle D_i | \Psi_a \rangle|^2 e^{-\beta E_a}}, \quad (52)$$

where  $\{E_a, \Psi_a\}$  are the exact energies and eigenstates within the basis.

The results of the complete vertex sums at two-, three-,

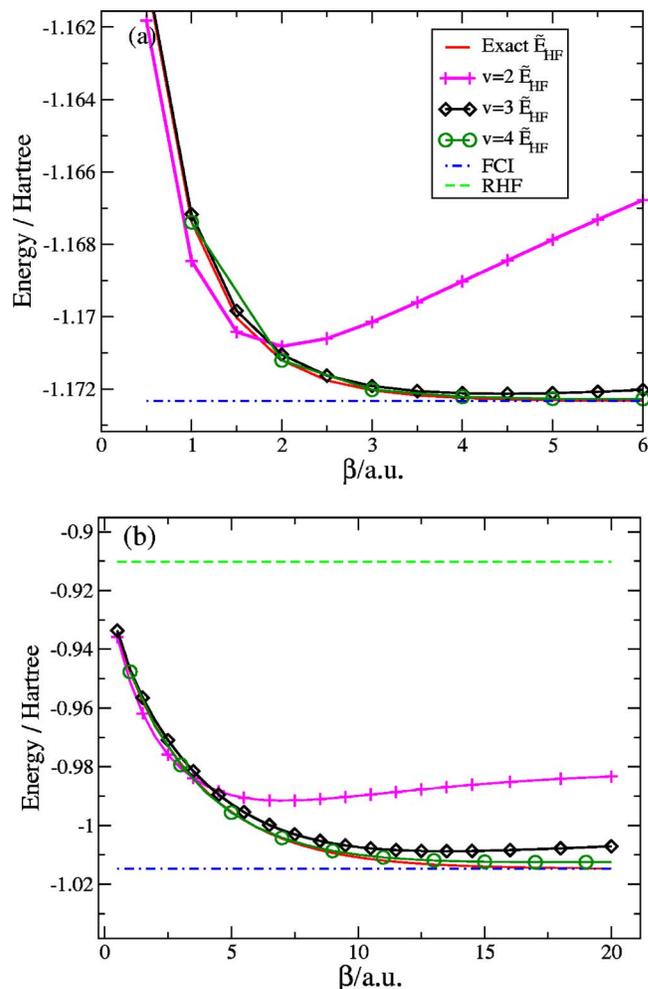


FIG. 4.  $\tilde{E}_{\text{HF}}(\beta)$  for  $H_2$  at (a)  $r=1.4a_0$  (b)  $r=4a_0$  for various methods. The exact value  $\tilde{E}_{\text{HF}}(\beta)$  has been obtained using Eq. (52). The complete  $n$ -vertex sums up to four-vertex graphs are shown. The restricted HF (RHF) energy in (a) is  $-1.131\,662\,82$  a.u.

and four-vertex levels, together with  $\tilde{E}_1^{(\text{exact})}$ , are shown in Fig. 4 at  $r=1.4$  and  $4a_0$ . Convergence towards the FCI curve is smooth and requires  $\beta=5$  at the equilibrium geometries and about 15 at the stretched one. The truncated sums all exhibit minima as a function of  $\beta$ , implying a breakdown of the series at sufficiently large  $\beta$ . This is not surprising. As  $\beta$  grows, one can expect ever larger graphs to contribute. In practice, this means that there is an optimal  $\beta$  for a given truncation. Nevertheless, the description of the entire binding curve, at the three-vertex level (Fig. 5), is satisfactory.

The  $N_2$  molecule was studied with a VDZ basis, for which FCI calculations have been previously reported.<sup>14,15</sup> The convergence of  $\tilde{E}_1$  as a function of  $\beta$  (Fig. 6) for the complete two- and three-vertex sums shows a minimum which approaches the FCI energy for the higher vertex result. At the three-vertex level, one can capture some 90% of the correlation energy at the equilibrium geometry, using  $\beta=2$ . The binding-energy curve for the complete three-vertex sum is shown in Fig. 7, where fair agreement can be seen by comparison to the FCI curve, particularly near the equilibrium geometry. The complete three-vertex results lie between the Hartree-Fock and FCI, but the agreement with the FCI

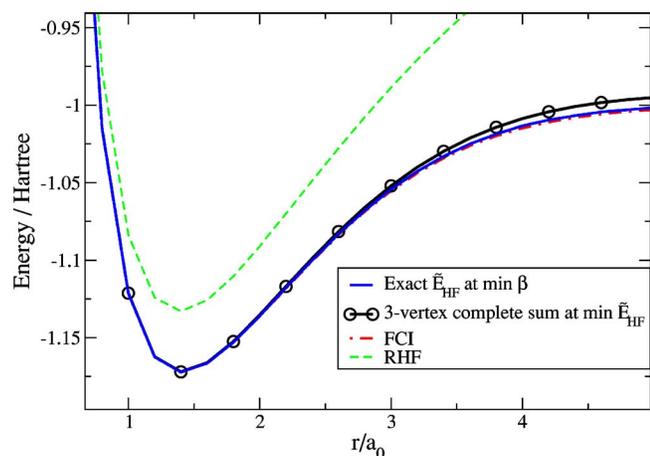


FIG. 5.  $H_2$  dissociation curves showing the result from the three-vertex  $\tilde{E}_{HF}$  and the exact  $\tilde{E}_{HF}$  evaluated at the  $\beta$  at which the former was minimized. The FCI and the RHF curves are also shown.

worsens as the bond is stretched beyond about  $r=3a_0$ . This is an indication of a slowing down in convergence of the vertex series in this limit. The shape of the curve, however, is physical, and there is no evidence of an artificial “hump.”

Turning to the Monte Carlo simulations, we first investigate the stability with respect to increasing  $\beta$ . Shown in Figs. 8–10 are traces from the simulation as  $\beta$  is increased from 0.5 to 16 in powers of 2. Each point on these graphs is a block average over  $10^6$  steps, and the simulations run over  $10^7$  steps. As  $\beta$  is increased, the energy generally decreases until about  $\beta=4$ , after which it begins to increase, indicating that  $\beta \approx 4$  is optimal for the present system. In general we observe that the fluctuations in the energy are rather stable. Equally importantly, the expectation value of the sign of the graphs is also well behaved. Both at small and large  $\beta$ ,  $\text{sign}(w)$  shows small fluctuations about a reasonable finite value. At small  $\beta$ , the overwhelming majority of the sampled graphs are trees. The most problematic  $\beta$  are in the intermediate regime, where the fluctuations in the sign are observed to increase, though not catastrophically. These coincide with the increase in importance of negative cyclic graphs. At large  $\beta$ , we observe a good cancellation between the positive cy-

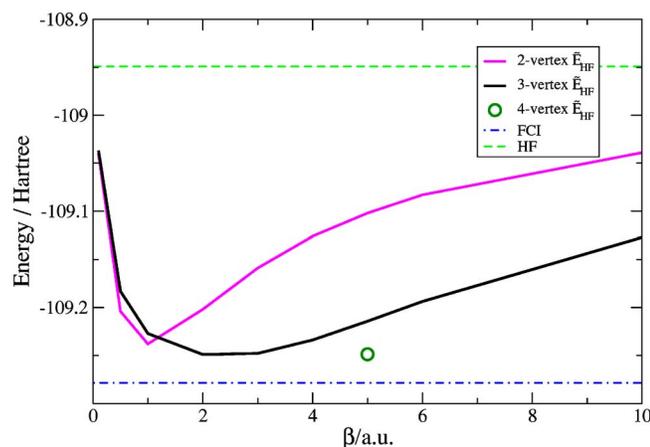


FIG. 6.  $\tilde{E}_{HF}(\beta)$  for the  $N_2$  molecule at  $r=2.118a_0$ , for the complete two-, three-, and four-vertex sums. The HF and FCI energies are also shown.

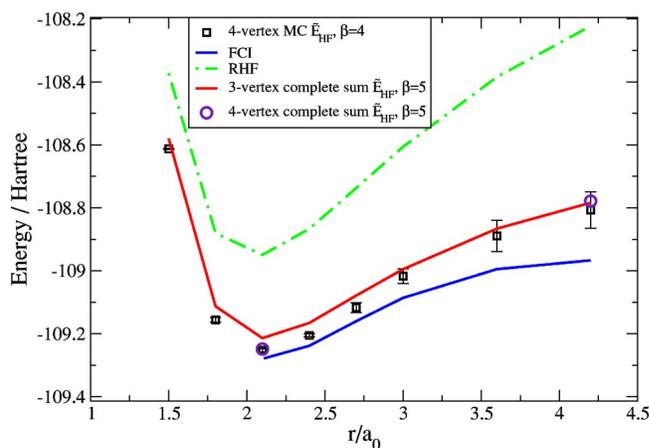


FIG. 7.  $N_2$  binding curves. Shown are the results from four-vertex MC simulations and the complete three-vertex sums. At two geometries ( $r=2.118, 4.2a_0$ ), we performed the complete four-vertex sums (shown in circles). For reference, the FCI and RHF curves are shown.

clic graphs and the negative ones. Even in this limit the trees account for some 75% of the sampled graphs.

We computed the binding-energy curve (Fig. 7) using the four-vertex Monte Carlo (MC) taking  $\beta=4$ . For each geometry, we ran simulations of length up to  $2^{26}=67 \times 10^6$  cycles. Since neighboring points along a MC simulation are correlated (in the present case due to strings of rejections of graphs), we used the “blocking method” as described by Flyvbjerg and Petersen<sup>16</sup> to estimate a correlation time. This analysis indicated that runs of length of  $5 \times 10^6$  cycles could be regarded as independent samples. We thus divided the complete run into segments of this length, and for each computed the required expectation values. For the systems with bond lengths smaller than about  $3a_0$ , the expectation values of the  $\tilde{E}_{HF}$ , Eq. (29), from each segment had a small spread, from which we could obtain the error in the estimated mean, indicated in Fig. 7. For the two more highly stretched cases ( $3.6$  and  $4.2a_0$ ) we found that the spread in the expectation value of  $\tilde{E}_{HF}$  to be  $\pm 0.2\text{Ha}$ , which is perhaps undesirably large. This increase is due to two (possibly related) reasons. First, we observe a modest decrease in the expectation value of the sign of the graphs (to  $\approx 0.3$ ) and, in addition, an increase in the importance of four-vertex graphs, the sampling of which may not be optimally done using our current graph-generation algorithm. Both factors lead to an increase in the spread of  $\tilde{E}_{HF}$ . A improved graph-generation method which more effectively generates the significant four-vertex graphs should help the situation. Otherwise, longer overall runs are required to reduce the error at these geometries, although this strategy is not efficient since the errors decrease only as the square root of the length of the run.

It should be noted, however, that the statistical error bars are nevertheless smaller than the systematic error (which is due to the truncation of the vertex sums). It is likely that the increasingly highly multiconfigurational character of the ground state of the  $N_2$  molecule as the bond is stretched implies that we will need to go to somewhat larger graphs in order to describe the full dissociation curve properly. Work

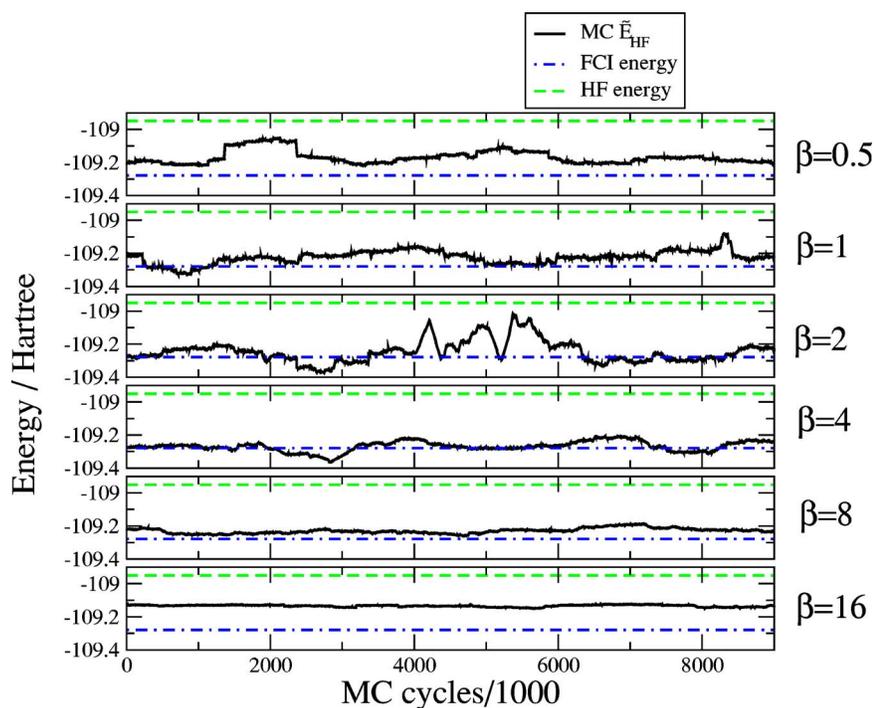


FIG. 8. Block average of  $\bar{E}_{\text{HF}}^{(n)}[G]$  during the Monte Carlo runs, for several different  $\beta$ . The FCI energy and HF energy are shown for reference. Energies are in a.u.

in this direction is in progress. Nevertheless we are very encouraged by the success of the present MC simulations. Put in perspective, the complete four-vertex sums take approximately one week to calculate, but the Monte Carlo simulations take only a few hours; this is an indication of the potential usefulness of the Monte Carlo sampling.

A final issue concerns size consistency, namely, whether the energy of a system consisting of two distant fragments (i.e., a “supermolecule”) is equal to sum of the isolated constituents. The question that arises is whether our truncation of the vertex series is size consistent in the zero-temperature limit.<sup>17</sup> We have made a preliminary analysis of this difficult issue. Graphs which contain determinants consisting of simultaneous excitations on both fragments give rise to size inconsistency, except in special cases. For example, two-

vertex graphs containing the Hartree-Fock (HF) determinant are size consistent (this arises in a manner analogous to that in MP2). A general truncation of the vertex series, however, is not size consistent. It may well be that we can generate size-consistent results by suitably neglecting size-inconsistent graphs. That said, we are not certain if this procedure is physically valid, given the inherent non-size-consistent nature of finite-temperature systems.

We end this section on a note of optimism. The  $\text{N}_2$  molecule is a particularly difficult system, and one would hope that present truncation at four-vertex graphs would generally suffice for many systems, bearing in mind that such graphs allow correlations up to six-fold excitations to be captured. For example, the dispersion interaction, which is essentially electron correlation coming at the double-excitation level,

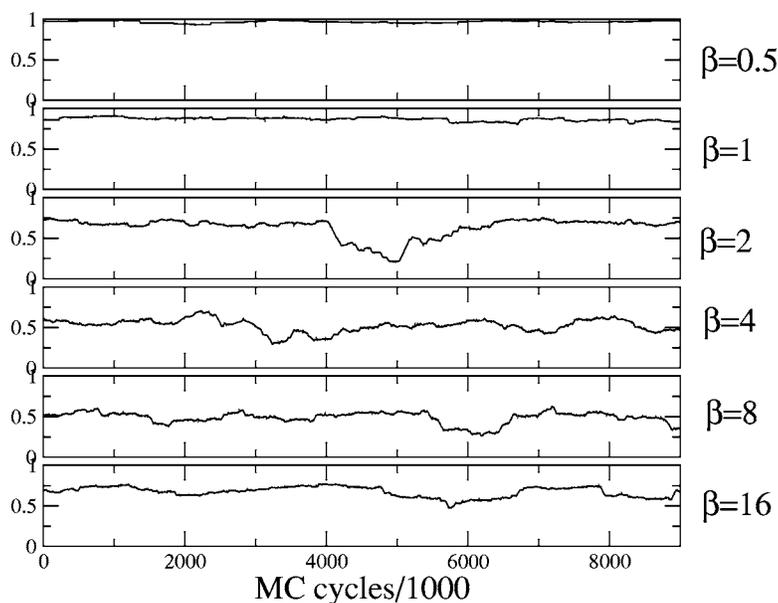


FIG. 9. Block averages of  $\langle \text{sign}(w^{(n)}[G]) \rangle_{|w|}$  during the Monte Carlo runs, for several different  $\beta$ .

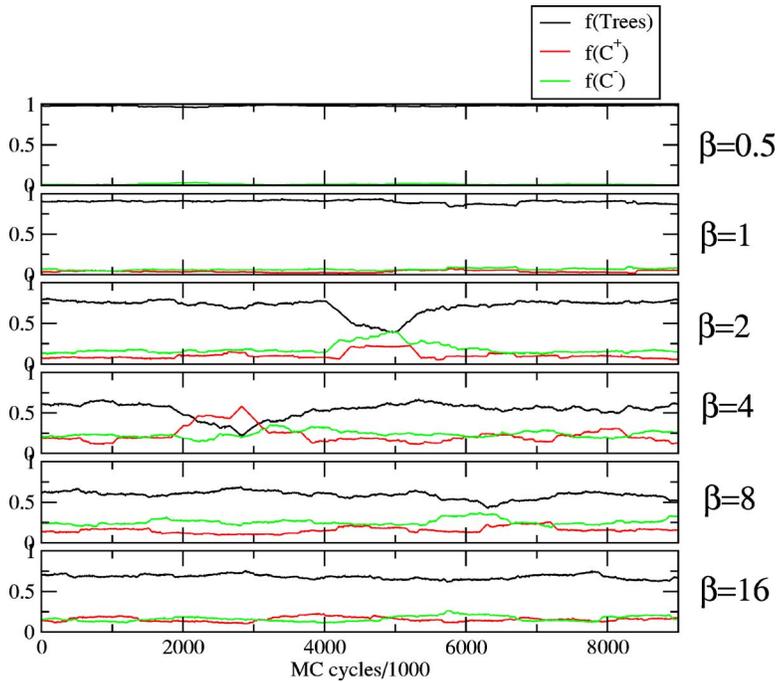


FIG. 10. Block averages of  $f_T, f_{C^+}$ , and  $f_{C^-}$  during the Monte Carlo runs, for several different  $\beta$ .

should be well described by the four-vertex approximation. This remains to be tested in future work. Perhaps most important of all, however, is the very ability to sample graphs without encountering catastrophic sign problems. The present paper is intended as a first account of these new ideas and does not address all issues which arise from it. These we hope to come to in future work.

## ACKNOWLEDGMENTS

We are grateful for the support of the EPSRC through a Portfolio Award, and a studentship to one of the authors (A.J.W.T.).

## APPENDIX A: AN ALGORITHM TO GENERATE GRAPHS WITH A COMPUTABLE GENERATION PROBABILITY

In order to sample graphs, it is necessary to generate them stochastically. We adopted a Markov chain algorithm to do this, in which successive determinants are added to a list until the desired size of graph is reached. Since the connectivity of the determinants is not uniform, such an algorithm will, in general, result in a generation probability which is not uniform. As a result, in order to remove any bias from the Monte Carlo simulation, it is necessary to be able to compute the generation probability. The following algorithm will generate a connected  $n$ -vertex graph with a computable normalized generation probability.

Starting at  $D_i$ , select a connected determinant  $D_j$ , with probability  $p_{ij}$ . This results in a two-vertex graph  $G = \{\mathbf{i}, \mathbf{j}\}$ . Next, select a determinant,  $D_k$ , connected to  $D_j$ , with probability  $p_{jk}$ . If  $D_k$  is distinct, then  $D_k$  is added to the list:  $G = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ . Otherwise a new determinant is generated from the current position, i.e., the last visited determinant, via a

suitable, randomly chosen single or double excitation. This process is continued until  $n$  distinct determinants have been generated.

The probability to generate a given graph  $G$  can be computed by examining all possible ways of generating  $G$  according to this algorithm. For example, consider the three-vertex graph  $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ . According to the above algorithm, the generation probability is

$$P_{\text{gen}}[G] = \sum_{n=0}^{\infty} (p_{ij}p_{ji})^n \times p_{ij}p_{jk} + \sum_{n=1}^{\infty} (p_{ij}p_{ji})^n \times p_{ik} + \sum_{n=0}^{\infty} (p_{ik}p_{kj})^n \times p_{ik}p_{kj} + \sum_{n=1}^{\infty} (p_{ik}p_{ki})^n \times p_{ij} \quad (\text{A1})$$

$$= \frac{p_{ij}(p_{jk} + p_{ji}p_{ik})}{1 - p_{ij}p_{ji}} + \frac{p_{ik}(p_{kj} + p_{ki}p_{ij})}{1 - p_{ik}p_{ki}}. \quad (\text{A2})$$

For an  $n$ -vertex graph, for  $n > 3$ , the generation probability is similarly computable. It is most compactly expressed in matrix notation. In order to simplify notation, let us call our  $n$  vertices  $G = \{\mathbf{i}_1, \dots, \mathbf{i}_n\}$ , with  $\mathbf{i}_1 \equiv \mathbf{i}$ , and no two vertices in this list are the same. Consider the generation probability of  $G$  in the given order  $(\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_n)$ . According to the algorithm, this means that we first visit  $\mathbf{i}_2$  from  $\mathbf{i}_1$ , then we visit  $\mathbf{i}_3$  for the first time from either  $\mathbf{i}_1$  or  $\mathbf{i}_2$ , etc. In general the  $(k+1)$ th vertex can be visited for the first time from any of the previous  $k$  vertices. The algorithm terminates when we first visit the  $n$ th vertex.

Let us construct a sequence of transition matrices  $P^{(k)}[\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_k]$  such that, in  $P^{(k)}$ , the vertex  $\mathbf{i}_k$  is an absorbing state. Such a matrix is

$$P^{(k)}[\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_k] = \begin{pmatrix} 0 & p_{\mathbf{i}_1\mathbf{i}_2} & p_{\mathbf{i}_1\mathbf{i}_3} & \dots & p_{\mathbf{i}_1\mathbf{i}_k} \\ p_{\mathbf{i}_2\mathbf{i}_1} & 0 & p_{\mathbf{i}_2\mathbf{i}_3} & \dots & p_{\mathbf{i}_2\mathbf{i}_k} \\ \dots & \dots & \dots & \dots & \dots \\ p_{\mathbf{i}_{k-1}\mathbf{i}_1} & 0 & p_{\mathbf{i}_{k-1}\mathbf{i}_2} & \dots & 0 \\ 0 & 0 & \dots & \dots & 1 \end{pmatrix}. \quad (\text{A3})$$

The matrix (A3) is a  $k \times k$  matrix which, in the terminology of the Markov chain theory<sup>18</sup> is a substochastic matrix (i.e., the sum across a row is less than or equal to 1).

The advantage of using such a transition matrix is that it guarantees that the Markov chain terminates once the vertex  $\mathbf{i}_k$  is arrived at. Noting that  $[(P^{(k)})^n]_{\mathbf{i}_{k-1}\mathbf{i}_k}$  gives the probability of arriving at  $\mathbf{i}_k$  in exactly  $n$  steps given that we started at  $\mathbf{i}_{k-1}$  (passing through some or all of the vertices  $\{\mathbf{i}_1, \dots, \mathbf{i}_{k-1}\}$ ), we see that the total probability of arriving at  $\mathbf{i}_k$  is simply the geometric series:

$$\sum_{l=0}^{\infty} [(P^{(k)}[\mathbf{i}_1, \dots, \mathbf{i}_k])^l]_{\mathbf{i}_{k-1}\mathbf{i}_k} = \left( \frac{1}{I - P^{(k)}[\mathbf{i}_1, \dots, \mathbf{i}_k]} \right)_{\mathbf{i}_{k-1}\mathbf{i}_k}, \quad (\text{A4})$$

where the right-hand side of Eq. (A4) requires the calculation of the inverse of the  $k \times k$  matrix  $(I - P^{(k)}[\mathbf{i}_1, \dots, \mathbf{i}_k])$ . Therefore the probability of generating the sequence  $\mathbf{i}_1 \rightarrow \mathbf{i}_2 \rightarrow \dots \rightarrow \mathbf{i}_n$  is

$$P_{\text{gen}}[\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_n] = \left( \frac{1}{I - P^{(2)}[\mathbf{i}_1, \mathbf{i}_2]} \right)_{12} \times \left( \frac{1}{I - P^{(3)}[\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3]} \right)_{23} \times \dots \times \left( \frac{1}{I - P^{(n)}[\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_n]} \right)_{n-1,n} = \prod_{k=2}^n \left( \frac{1}{I - P^{(k)}[\mathbf{i}_1, \dots, \mathbf{i}_k]} \right)_{k-1,k}. \quad (\text{A5})$$

The total generation probability of this graph irrespective of the sequence, in which  $\mathbf{i}_2, \mathbf{i}_3, \dots, \mathbf{i}_n$  is chosen, is the sum over the  $(n-1)!$  permutations in which this set of  $n-1$  vertices

can be generated. The total generation probability is therefore

$$P_{\text{gen}}[G] = P_{\text{gen}}[\{\mathbf{i}_1, \dots, \mathbf{i}_n\}] = \sum_{\hat{P}} P_{\text{gen}}[\mathbf{i}_1, \hat{P}(\mathbf{i}_2, \dots, \mathbf{i}_n)], \quad (\text{A6})$$

where  $\hat{P}$  is permutation operator over the set of vertices  $\mathbf{i}_2 \dots \mathbf{i}_n$ .

In practice, we deal with graphs consisting of a rather modest number of vertices, say three or four, which means we need invert matrices no larger than  $4 \times 4$ , a modest computational task.

There is considerable freedom in choosing the probabilities  $p_{\mathbf{ij}}$ . The simplest choice (and the one we used) is to take

$$p_{\mathbf{ij}} = \frac{1}{N_{\mathbf{i}}}, \quad (\text{A7})$$

where  $N_{\mathbf{i}}$  is the number of determinants connected to  $\mathbf{i}$ . In other words the selection of determinant  $D_{\mathbf{j}}$  is done uniformly, save the fact that it must be connected to  $D_{\mathbf{i}}$  (i.e.,  $p_{\mathbf{ij}} \neq 0$ ). Certainly we would expect to be able to improve on this selection probability by using a nonuniform distribution. This is a topic of on-going research in our group. A second variant on the above algorithm is to select the new determinant at any stage from any of the previously visited determinant, rather than from the current position. The generation probability is computable along similar lines to the above. We found this to be an effective alternative in generating graphs in which the Hartree-Fock determinant remains central.

### APPENDIX B: THE COMPLETE FOUR-VERTEX GRAPH

In the complete four-vertex graph, all vertices are connected to each other, i.e., there are  $\binom{4}{2}$  edges. To obtain the unfolded representation, one needs to enumerate all possible distinct circuits on a four-vertex graph. Carrying this procedure out,<sup>19</sup> setting the result on a star graph pivoted at  $\mathbf{i}$ , one obtains the following expression for the total weight for paths of length  $P$  which start and finish at  $\mathbf{i}$ :

$$S_{\mathbf{i}}^{(4)}[\{\mathbf{i}, \mathbf{j}, \mathbf{k}, \mathbf{l}\}] = \frac{1}{2\pi i} \oint_C \frac{z^P - 1}{z - 1} \times 1 \left/ \left[ 1 - \frac{A_{\mathbf{ij}}}{1 - \frac{A_{\mathbf{jk}}}{1 - A_{\mathbf{kl}}} - \frac{A_{\mathbf{jl}}}{1 - A_{\mathbf{kl}}} - \frac{2A_{\mathbf{jkl}}}{1 - A_{\mathbf{kl}}}} - \frac{A_{\mathbf{ik}} + A_{\mathbf{ijl}} + A_{\mathbf{ijkl}} + A_{\mathbf{ijlk}}}{\left(1 - \frac{A_{\mathbf{ij}}}{1 - A_{\mathbf{kl}}} - \frac{A_{\mathbf{jl}}}{1 - A_{\mathbf{kl}}} - \frac{2A_{\mathbf{jkl}}}{1 - A_{\mathbf{kl}}}\right)(1 - A_{\mathbf{kl}})} - \frac{A_{\mathbf{ik}}}{1 - \frac{A_{\mathbf{jk}}}{1 - A_{\mathbf{jl}}} - \frac{A_{\mathbf{kl}}}{1 - A_{\mathbf{jl}}} - \frac{2A_{\mathbf{jkl}}}{1 - A_{\mathbf{jl}}}} - \frac{A_{\mathbf{ijk}} + A_{\mathbf{ikl}} + A_{\mathbf{ikjl}} + A_{\mathbf{iklj}}}{\left(1 - \frac{A_{\mathbf{ij}}}{1 - A_{\mathbf{jl}}} - \frac{A_{\mathbf{kl}}}{1 - A_{\mathbf{jl}}} - \frac{2A_{\mathbf{jkl}}}{1 - A_{\mathbf{jl}}}\right)(1 - A_{\mathbf{jl}})} - \frac{A_{\mathbf{il}}}{1 - \frac{A_{\mathbf{jl}}}{1 - A_{\mathbf{jk}}} - \frac{A_{\mathbf{kl}}}{1 - A_{\mathbf{jk}}} - \frac{2A_{\mathbf{jkl}}}{1 - A_{\mathbf{jk}}}} - \frac{A_{\mathbf{ijl}} + A_{\mathbf{ikl}} + A_{\mathbf{iljk}} + A_{\mathbf{ilkj}}}{\left(1 - \frac{A_{\mathbf{ij}}}{1 - A_{\mathbf{jk}}} - \frac{A_{\mathbf{kl}}}{1 - A_{\mathbf{jk}}} - \frac{2A_{\mathbf{jkl}}}{1 - A_{\mathbf{jk}}}\right)(1 - A_{\mathbf{jk}})} \right]. \quad (\text{B1})$$

This remarkably simplifies to

$$S_i^{(4)}[\{\mathbf{i}, \mathbf{j}, \mathbf{k}, \mathbf{l}\}] = \frac{1}{2\pi i} \oint_C \frac{z^P - 1}{z - 1} \frac{1 - A_{jk} - A_{jl} - A_{kl} - 2A_{jkl}}{[1 - A_{ij}(1 - A_{kl}) - A_{ik}(1 - A_{jl}) - A_{il}(1 - A_{jk}) - A_{jk} - A_{jl} - A_{kl} - 2A_{ijk} - 2A_{ijl} - 2A_{ikl} - 2A_{jkl} - 2A_{ijkl} - 2A_{ijlk} - 2A_{ikjl}]} dz \quad (\text{B2})$$

The denominator reduces to a quartic polynomial in  $z$ , whose roots lead to four residues. The quartic polynomial can be solved iteratively using standard numerical techniques such as the false-position method.

<sup>1</sup>R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

<sup>2</sup>M. E. Tuckerman and A. Hughes, in *Classical and Quantum Dynamics in Condensed Systems*, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, Singapore, 1998).

<sup>3</sup>D. M. Ceperley, Phys. Rev. Lett. **69**, 331 (1992).

<sup>4</sup>J. B. Anderson, J. Chem. Phys. **63**, 1499 (1975); **65**, 4121 (1976).

<sup>5</sup>A. Luchow and J. B. Anderson, J. Chem. Phys. **105**, 4636 (1996).

<sup>6</sup>W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).

<sup>7</sup>In this paper, we will use curly brackets to refer to sets in which the order of the elements is not important.

<sup>8</sup>D. E. Knuth, *The Art of Computer Programming, Fundamental Algorithms*, Vol. 1 (Addison-Wesley, Reading, MA, 1973).

<sup>9</sup>Another useful property is that  $Z_h^{(P)}(x_1, \dots, x_h) = 0$  when  $h > P$  for any  $(x_1, \dots, x_h)$ . This allows one to replace the upper limits in summations over "hops"  $h$  to infinity, rather than  $P$ , which considerably simplifies the algebra in handling  $Z$  sums.

<sup>10</sup>In cases where the poles are multiple rather than simple, the residue theorem still holds but one has to use its generalized version, i.e.,

$$\frac{1}{2\pi i} \oint_C f(z) dz = (\text{sum of enclosed residues}),$$

where a residue for a pole of order  $m$  at  $z_0$  is given by

$$\frac{1}{(m-1)!} \frac{d^{m-1}}{dz^{m-1}} [(z - z_0)^m f(z)].$$

<sup>11</sup>The identity, Eq. (36), has also been derived by R. W. Hall, J. Chem. Phys. **116**, 1, (2002), by an independent method.

<sup>12</sup>H.-J. Werner, P. J. Knowles, R. Lindh *et al.*, MOLPRO, version 2000, a package of *ab initio* programs

<sup>13</sup>T. H. Dunning, Jr., J. Chem. Phys. **53**, 2823 (1989).

<sup>14</sup>J. W. Krogh and J. Olsen, Chem. Phys. Lett. **344**, 578 (2001).

<sup>15</sup>G. K-C. Chan, M. Kállay, and J. Gauss, J. Chem. Phys. **121**, 6110 (2004).

<sup>16</sup>H. Flygberg and H. G. Petersen, J. Chem. Phys. **91**, 461 (1989).

<sup>17</sup>Let us note that size consistency is, strictly speaking, a property only of the ground states of the supermolecule and isolated fragments. At finite temperature, contributions from charge-transfer excited states arise which affect the energy of the supermolecule and are not present in the isolated fragments. For example, in a dissociating  $N_2$  molecule, the charge-transfer state in which an electron is removed from one atom and given to the other is an excited state which cannot be described by a single N atom. An exact finite-temperature method should become increasingly size consistent as the temperature is lowered, where the contribution from such charge-transfer excitations will become exponentially small.

<sup>18</sup>W. Feller, *Introduction to Probability Theory and Its Applications* (Wiley, New York, 1950), Chap. 15, Vol. I.

<sup>19</sup>A. J. W. Thom, Certificate for Postgraduate Study, University of Cambridge (2004).