Stochastic Perturbation Theory: A Low-Scaling Approach to Correlated Electronic Energies

Alex J. W. Thom* and Ali Alavi

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

(Received 20 April 2007; published 4 October 2007)

We discuss a stochastic implementation of Møller-Plesset (MP) theory based upon the concept of a "graph," a set of connected Slater determinants. We show how contributions from an arbitrary level, MPn, of perturbation theory can be expressed diagrammatically in terms of graphs, and that these may be stochastically sampled to give a good estimate of the energy. We show this to be the case for Ne, Ar, N_2, and H_2O molecules. N-molecule chains of He atoms and H_2 molecules at equilibrium and stretched geometries show an effective scaling of $\mathcal{O}[N^{2.6}]$ and $\mathcal{O}[N^{5.6}]$ for MP2 and MP3 theories.

DOI: 10.1103/PhysRevLett.99.143001 PACS numbers: 31.15.Md, 02.70.Ss

Wave-function-based methods such as Møller-Plesset (MP) perturbation theory [1,2], can provide high accuracy for calculating the energy of electronic systems and are capable, for example, of describing long-range dispersion interactions between closed-shell molecules. However, they are limited in their application to large systems by prohibitive scaling with system size: MP2 and MP3 scale, respectively, as $\mathcal{O}[N^4]$ and $\mathcal{O}[N^6]$, where N is the number of electrons [3]. Depending on the system, a reduction in scaling can be achieved by exploiting spatial locality in computing the matrix elements [1,2]. Alternatively, stochastic approaches such as quantum Monte Carlo provide a more general approach at a reasonable scaling [4]. Here we use insight gained from such stochastic methodology in a quantum chemical context. Specifically, we develop a stochastic implementation of MP theory in terms of "graphs," the notion of which we have used in a previous path-integral-based Monte Carlo method [5]. A "graph" is a set of connected Slater determinants in terms of which we can express the contributions from an arbitrary level of perturbation theory. The machinery we have previously developed to generate graphs according to a weighted, yet normalized, probability distribution allows us to evaluate the MPn energies stochastically.

Much of the complexity involved in diagrammatic perturbation theory [6] stems from the nature of the interaction operator, $\hat{u}(\mathbf{r}_1, \mathbf{r}_2)$. As a two-particle operator, diagrams and corresponding expressions in this formulation require interactions between two holes and two particles, $\langle ij|\hat{u}|ab\rangle$. We have found considerable simplification is possible by working, not in this space of orbitals as is conventional, but in the space of Slater determinants, each of which is a distinct N-electron configuration. The perturbation theory may be expanded in this basis and the interaction instead becomes the matrix element between determinants, $\langle D_\mathbf{i}|\hat{V}|D_\mathbf{j}\rangle$. This allows us to use a "one-particle-like" perturbation theory. We define $\hat{H}_1 = \hat{H} - \hat{H}_0$ as the difference between the full N-electron Hamiltonian and the reference Hamiltonian, which is the many electron Fock operator, $\hat{H}_0 = \sum_i \varepsilon_i c_i^\dagger c_i$, where ε_i, $c_i^\dagger$, and c_i are the energy, creation, and annihilation operators for an electron in the ith Hartree-Fock orbital.

Following Brueckner's bracketing formulation [7] of Rayleigh-Schrödinger perturbation theory, we may form expressions for the nth order energy by considering how many ways we can place a pair of brackets, $\langle$ and $\rangle$, each enclosing a $\hat{V}$, into the bracketed string $\langle \hat{V}\hat{R}\hat{V}\hat{R}\ldots\hat{R}\hat{V}\rangle$. In such a string there are $n\hat{V}$'s and $(n-1)$ $\hat{R}$'s. Here $\hat{R}$ represents the resolvent operator,

$$\hat{R} = \sum_{\mathbf{i}\neq 0} \frac{|D_\mathbf{i}\rangle\langle D_\mathbf{i}|}{E_0 - \hat{H}_0}, \tag{1}$$

in which the sum over $\mathbf{i}$ spans the complete space of determinants excluding the reference determinant D_0. We initially will take $\hat{V} = \hat{H}_1$. As an example, the MP3 case is instructive. There are two possible ways to insert brackets: $\langle \hat{V}\hat{R}\hat{V}\hat{R}\hat{V}\rangle$, $\langle \hat{V}\hat{R}\langle\hat{V}\rangle\hat{R}\hat{V}\rangle$. To interpret each expression, we substitute $\langle D_0|$ for $\langle$ and $|D_0\rangle$ for $\rangle$, and negate if there are an even number of $\langle\,\rangle$ pairs. For the former,

$$\begin{aligned}\langle D_0|\hat{V}\sum_{\mathbf{i}\neq 0}\frac{|D_\mathbf{i}\rangle\langle D_\mathbf{i}|}{E_0-\hat{H}_0}\hat{V}\sum_{\mathbf{j}\neq 0}\frac{|D_\mathbf{j}\rangle\langle D_\mathbf{j}|}{E_0-\hat{H}_0}\hat{V}|D_0\rangle \\ = \sum_{\mathbf{i},\mathbf{j}\neq 0}\frac{\langle D_0|\hat{V}|D_\mathbf{i}\rangle\langle D_\mathbf{i}|\hat{V}|D_\mathbf{j}\rangle\langle D_\mathbf{j}|\hat{V}|D_0\rangle}{(E_0-E_\mathbf{i})(E_0-E_\mathbf{j})},\end{aligned} \tag{2}$$

and the latter,

$$\begin{aligned}-\langle \hat{V}\hat{R}\langle\hat{V}\rangle\hat{R}\hat{V}\rangle &= -\langle\hat{V}\rangle\langle\hat{V}\hat{R}\hat{R}\hat{V}\rangle \\ &= -\langle D_0|\hat{V}|D_0\rangle \\ &\quad\times\langle D_0|\hat{V}\sum_{\mathbf{i}\neq 0}\frac{|D_\mathbf{i}\rangle\langle D_\mathbf{i}|}{E_0-\hat{H}_0}\sum_{\mathbf{j}\neq 0}\frac{|D_\mathbf{j}\rangle\langle D_\mathbf{j}|}{E_0-\hat{H}_0}\hat{V}|D_0\rangle \\ &= -\langle D_0|\hat{V}|D_0\rangle\sum_{\mathbf{i}\neq 0}\frac{\langle D_0|\hat{V}|D_\mathbf{i}\rangle\langle D_\mathbf{i}|\hat{V}|D_0\rangle}{(E_0-E_\mathbf{i})^2}.\end{aligned} \tag{3}$$

Brackets can, therefore, only be placed surrounding an expression with a $\hat{V}$ at both ends, as placing a bracket next

to an $\hat{R}$ would result in an expression containing $\sum_{\mathbf{i}\neq 0}\langle D_0|D_\mathbf{i}\rangle = 0$. This bracketing can prove somewhat cumbersome, so it is convenient to perform it diagrammatically. We may describe a given term with a diagram where the resolvent $\hat{R}$ features as a vertex, and the interaction $\hat{V}$ as an edge. Figure 1 shows this process in the MP3 case. We have denoted the vertices by squares to indicate that they include the denominator of the resolvent, with a filled square indicating the reference determinant (which has no denominator). Summations over all possible determinants at each unfilled vertex must be performed to recover the full energy contribution.

These diagrams differ considerably from Brandow [8], Goldstone [9], and Hugenholtz [10] diagrams, which are essentially Feynman diagrams, where lines indicate hole and particle creation operators, and vertices indicate interaction operators. Our diagrams are essentially duals of the one-particle interaction operator diagrams.

The process of bracketing involves taking a linked subset of the unfilled vertices in the diagram, along with the edges that bind them to the diagram, and disconnecting it. The dangling edges in this disconnected diagram are then bound to a filled vertex, the reference determinant. The terminating vertices in the main diagram are fused together to reform a closed diagram. This is illustrated in Figs. 1(b) and 1(c). This process also allows the removal of a single edge as shown in the figure. The component subdiagrams formed must have their contributions multiplied to evaluate contribution of the diagram.

The MPn energy is the sum of diagrams with n vertices. Such a diagram will only have a value if two neighboring vertices are connected (i.e., have a nonzero $\langle D_\mathbf{i}|\hat{V}|D_\mathbf{j}\rangle$ element). Furthermore, the unrestricted nature of the sums over determinant space allows a diagram to contain two (or more) determinants which are identical. Such diagrams can be depicted as in Fig. 1(d). When summing over diagram space, it is convenient to put diagrams into groups dependent upon the number of *distinct* determinants contained in them, and to sum them together. We should be careful, however, only to sum up such diagrams as required for a given level of MPn theory. The linked diagram theorem [11] (where "linked" applies to the Feynman-type diagram representation) shows that in such full summations up to a given MPn level, the contributions from non-size-consistent unlinked diagrams will cancel, so we need not currently be concerned with ensuring we only have linked diagrams.

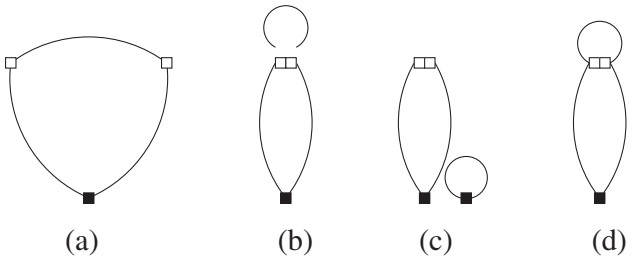

FIG. 1. (a) A typical MP3 diagram corresponding to $\langle \hat{V}\,\hat{R}\,\hat{V}\,\hat{R}\,\hat{V}\rangle$. The filled square corresponds to the reference determinant, and the unfilled squares to resolvent operators. The lines connecting determinants correspond to the interaction operator $\hat{V}$; (b) the bracketing process in $\langle \hat{V}\,\hat{R}\langle\hat{V}\rangle\hat{R}\,\hat{V}\rangle$. The topmost line has been disconnected, and the two resolvents fused; (c) the dangling edges are attached to a filled vertex giving $-\langle \hat{V}\,\hat{R}\,\hat{R}\,\hat{V}\rangle\langle\hat{V}\rangle$, which is negative owing to an even number of cycles; (d) an MP3 term with only two distinct determinants is achieved when $\mathbf{i} = \mathbf{j}$ in (a) [Eq. (2)]. Its contribution is accounted for in the two-vertex graph (0$\mathbf{i}$).

We may now note that the two possible diagrams for MP3, Figs. 1(a) and 1(c), can be combined into a single term if instead we define $\hat{V} = \hat{H}_1 - E_1$, where E_1 is the first-order energy, $\langle D_0|\hat{H}_1|D_0\rangle$. In Eq. (2) the E_1 terms vanish in the first and last brackets, and the E_1 of the middle bracket produces precisely the value of Eq. (3). With this new $\hat{V}$ operator, $\langle\hat{V}\rangle$ also vanishes, so contributions with a component which has a single line such as Fig. 1(c) are zero, considerably simplifying the required expressions for higher n.

We may now relate these diagrams to graphs. Although we have explored the concept of graphs in Slater determinant space in a previous paper [5], it is convenient to review the definition. A graph is a set of vertices, which represent distinct Slater determinants, and a set of edges which connect them. The vertices are labeled according an ordered N vector which contains the occupied orbitals in the Slater determinant, which we designate $\mathbf{i}, \mathbf{j}, \ldots$. The edges correspond to a matrix element between different determinants (here the edge between $\mathbf{i}$ and $\mathbf{j}$ corresponds to the matrix element $\langle D_\mathbf{i}|\hat{V}|D_\mathbf{j}\rangle$), or of a determinant and itself, e.g., $\langle D_\mathbf{i}|\hat{V}|D_\mathbf{j}\rangle$. Using this definition we may use the graph to represent a value which can be calculated from these matrix elements. Our previous use of graphs had high-temperature density matrix elements as the edges, and had the value of a graph being the sum of all possible paths using the vertices of the graph. Here we shall take the value of a graph to be the sum of the energy contributions of all levels up to MPn theory using exclusively all determinants in the graph. This represents a very minor efficiency saving in the case of full MP calculations as the main computational cost is that of enumerating the appropriate graphs at the n-vertex level. However, we may also use the algorithms previously derived [5] to generate graphs randomly, and evaluate MP energies with Monte Carlo (MC) calculations.

The full MPn sum may be written as a sum, first over graph size v and then over graphs with v vertices, G_v,

$$E = \sum_{v=1}^{n}\sum_{G_v} E_{v\to n}[G_v]. \tag{4}$$

Here $E_{v\to n}$ includes energy contributions from MPv through to MPn formed exclusively from all determinants

in graph G_v. Given that we are able to generate graphs randomly with a computable *normalized* generation probability (see Appendix A of Ref. [5]), $p_{\text{gen}}[G_v]$, we can rewrite the sum as an expectation value,

$$E = \sum_{v=1}^{n} \sum_{G_v} \frac{E_{v\to n}[G_v]}{p_{\text{gen}}[G_v]} p_{\text{gen}}[G_v] \tag{5}$$

$$= \sum_{v=1}^{n} \left\langle \frac{E_{v\to n}[G_v]}{p_{\text{gen}}[G_v]} \right\rangle_{p_{\text{gen}}[G_v]}. \tag{6}$$

This formulation allows expectation values to calculated stochastically, and at each step, a graph G randomly generated. Note that this is not a Metropolis MC calculation: each generated graph is accepted, and the value $E_{v\to n}[G]/p_{\text{gen}}[G]$ averaged. Letting $V_{\mathbf{ij}} = \langle D_{\mathbf{i}}|\hat{V}|D_{\mathbf{j}}\rangle$, for MP2 and MP3 the required energy contributions are

$$E_{2\to 2}[0\mathbf{i}] = \frac{V_{0\mathbf{i}} V_{\mathbf{i}0}}{E_0 - E_{\mathbf{i}}}, \tag{7}$$

$$E_{2\to 3}[0\mathbf{i}] = \frac{V_{0\mathbf{i}} V_{\mathbf{i}0}}{E_0 - E_{\mathbf{i}}} + \frac{V_{0\mathbf{i}} V_{\mathbf{ii}} V_{\mathbf{i}0}}{(E_0 - E_{\mathbf{i}})^2}, \tag{8}$$

$$E_{3\to 3}[0\mathbf{ij}] = \frac{V_{0\mathbf{i}} V_{\mathbf{ij}} V_{\mathbf{j}0} + V_{0\mathbf{j}} V_{\mathbf{ji}} V_{\mathbf{i}0}}{(E_0 - E_{\mathbf{i}})(E_0 - E_{\mathbf{j}})}. \tag{9}$$

A simple example of this is afforded by MP2, which consists of 2-vertex graphs between the Hartree-Fock determinant and connected double excitations. Not taking into account any symmetry restrictions, for N electrons and V virtual spin-orbitals, there are $\binom{N}{2}\binom{V}{2}$ such double excitations; therefore, for an unweighted generation, the normalized generation probability is $p_{\text{gen}} = \binom{N}{2}^{-1}\binom{V}{2}^{-1}$.

To reduce the variance of an MC calculation, a weighted graph generation scheme was used, where the weighting is based on the generation of connected excitations of a given determinant. Weighting the generation of excitations with a known normalized probability requires some care, as to naïvely normalize an arbitrary weighting would require summing over all possible excitations at every step [12]. For double excitations, we chose to split the excitation routine into two parts. First, a pair of occupied orbitals (i, j) was selected with uniform probability. Next, the weights of all possible symmetry-allowed virtual pairs were calculated, and a pair (a, b) selected from this set with probability proportional to its weight. Based upon the form of the MP2 contribution, the weight of a double excitation was set proportional to the inverse of the energy difference of excitation $w \propto |\Delta\varepsilon|^{-1}$, where $\Delta\varepsilon = \varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j$. Where $|\Delta\varepsilon| < 10^{-2}$ Har, it was set to 10^{-2} Har to avoid the singularity. Single excitations were not explicitly weighted, and were generated with same likelihood as if there were no weighting.

To compare between weighted and unweighted schemes we have performed calculations on the N_2 molecule with a bond length of $2.1a_0$. Hartree-Fock calculations were first performed with DALTON [13] using a cc-pVQZ basis [14], precomputing the required one- and two-electron integrals between molecular orbitals. This was followed by MP2 and MP3 MC runs of 10^7 steps, freezing the four core electrons. Running averages for these can be seen in Fig. 2. There is a clear decrease in the expected errors using the weighting scheme, and a good convergence toward the exact values. Our graph generation scheme is not optimal, however: for this system all generated 2-vertex graphs contributed to the MP2 energy, but only 1% of the 3-vertex graphs generated contained a connected 3-cycle, the requirement for having a contribution to the MP3 energy. This leaves plenty of scope for more efficient sampling, which would lower both the weighted and unweighted errors. In Table I, we show the results of MC calculations on some small molecules.

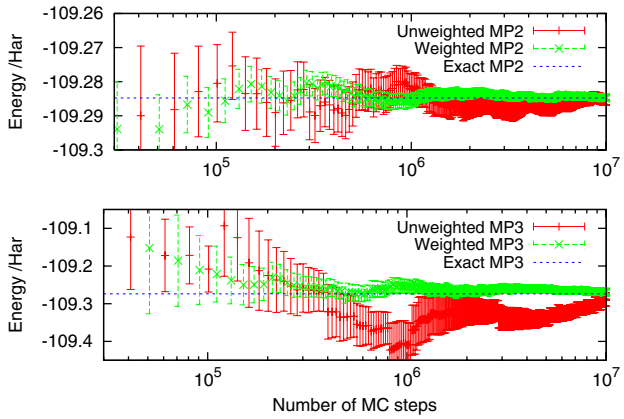

FIG. 2 (color online). MC running averages for cc-pVQZ N_2. Energy and error estimators have been plotted every 20 000 steps in a 10^7 step simulation for both unweighted and weighted generation schemes showing that the weighted schemes exhibit a faster convergence to the exact calculated values.

Three systems were chosen to investigate the effect of system size on the scaling of the method: (i) a linear array of He atoms, $10a_0$ apart; (ii) a linear array of H_2 molecules with bond distance $r_e = 0.742a_0$, spaced $3r_e$ apart; (iii) a similar array with bond distance $1.5r_e$ and separation

TABLE I. Some sample MP3 MC energies for closed-shell molecules close to equilibrium using a cc-pVQZ basis set. The geometries used were $r_{\text{NN}} = 2.1a_0$ and $r_{\text{OH}} = 1$ Å, $\theta_{\text{HOH}} = 104.5°$. All energies are in Hartrees with MC errors in the last figure in brackets. MC runs had 10 and 40×10^6 steps at the 2- and 3-vertex levels, respectively. Calculations were all electron, except for N_2 where 4 electrons were frozen.

	MP2	MP2 MC	MP3	MP3 MC
Ne	−128.869 73	−128.870 1(5)	−128.8703	−128.868(2)
Ar	−527.110 45	−527.110 6(8)	−527.1272	−527.125(5)
N_2	−109.284 76	−109.284 7(6)	−109.2739	−109.273(4)
H_2O	−76.374 65	−76.374 7(11)	−76.3763	−76.381(5)

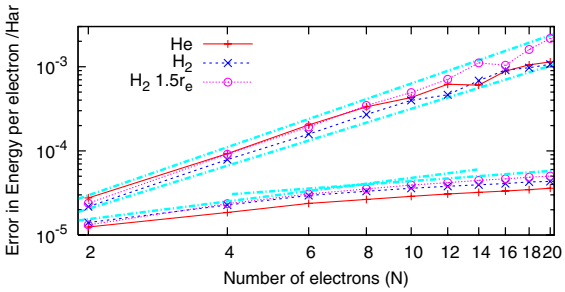

FIG. 3 (color online). The estimated error per electron from a 10^6 step MC run is shown against the number of electrons for chains of H_2 and He using a weighted graph generation algorithm. MP2 errors (lower lines) are bounded by scalings (thick lines) $N^{0.7}$ and $N^{0.4}$ for small and large system scaling regime. MP3 errors (upper lines) have scalings between $N^{1.7}$ and $N^{1.9}$.

$2.5r_e$, i.e., with partial bond equalization, where the nature of the interaction is very different to (ii). In both H_2 systems, the centers of masses of neighboring molecules were $4r_e$ apart. Hartree-Fock calculations were performed with DALTON [13] using a Dunning cc-pVTZ basis [15] for He and a cc-pVDZ basis [14] for H, and the required one- and two-electron integrals between molecular orbitals were precomputed for systems with 1 to 10 molecules. MC calculations were performed for all systems, and we investigate three types of scaling for each of MP2 and MP3 theory: decrease in errors with number of steps, and both step-length and error per electron as system size increases. As the number of steps increases, the errors follow the expected $1/\sqrt{N_{\text{steps}}}$ dependence.

For step-length scaling, the graph generation algorithm has a projected scaling of at most $\mathcal{O}[N^2]$, owing to the number of pairs of virtuals possible, but the greatest scaling seen is independent of the level of theory at $\mathcal{O}[N^{1.8}]$, which may be due to the relatively small system size. Next, we investigate how the errors in the MC calculations change with system size. It is reasonable to expect them to increase as the number of graphs available with n vertices, and thus the space in which the MC calculation is performed, scales as $\mathcal{O}[(NM)^{2(n-1)}]$; however, it is not clear what the exponent of the error scaling will be. Figure 3 shows that the scaling of the error for these three different systems is rather similar. Table II collates the scalings of MP2 and MP3, giving estimated total scalings.

In formulating an arbitrary level of Møller-Plesset theory in terms of graphs, we have been able to sample these stochastically owing to a graph generation algorithm with normalized probabilities. Our current weighting scheme is certainly not optimal as it does not include the magnitude of the coupling between determinants in the generation. However, as this Letter is conceived as a proof-of-concept, we have not investigated the effects of different weighting functions for graph generation, and leave this investigation to future work. Similarly, exploiting spatial locality in computing the matrix elements should aid this, and furthermore reduce the expected scaling of the graph generation. We should, however, stress that, for graphs of size n to have an MPn contribution, they must contain a cycle of length n. We have not currently built this restriction into our graph generation process, and so for $n \geq 3$ calculations most of the graphs generated contribute nothing to the energy, and so are being needlessly sampled; work is in progress to include the restriction, which should considerably reduce the scaling of the higher levels of theory.

TABLE II. The various scalings within an MC run. The total scaling is the time taken to achieve a given expected error. The number of steps required will scale as (error scaling)^2 because of the $1/\sqrt{N_{\text{steps}}}$ dependence of the MC errors, which is multiplied by the step timing to give the total scaling.

Method	Step timing	Error scaling	Total scaling
MP2	$N^{1.8}$	$N^{0.4}$	$N^{2.6}$
MP3	$N^{1.8}$	$N^{1.9}$	$N^{5.6}$

More generally, we believe Monte Carlo methods can be usefully applied to diagrammatic theories whose term-by-term evaluation is prohibitive, but which may be formulated, as in the present Letter, in a manner suitable to a stochastic sampling, with error bars which can be reliably estimated from the simulation data.

The authors would like to thank the EPSRC for their support.

*ajwt3@cam.ac.uk

[1] H.-J. Werner, F. R. Manby, and P. J. Knowles, J. Chem. Phys. **118**, 8149 (2003).
[2] J. E. Subotnik, A. Sodt, and M. Head-Gordon, J. Chem. Phys. **125**, 074116 (2006).
[3] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, Chichester, 2000).
[4] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[5] A. J. W. Thom and A. Alavi, J. Chem. Phys. **123**, 204106 (2005).
[6] S. Wilson, *Electron Correlation in Molecules* (Clarendon Press, Oxford, U.K., 1984).
[7] K. A. Brueckner, Phys. Rev. **100**, 36 (1955).
[8] B. H. Brandow, Rev. Mod. Phys. **39**, 771 (1967).
[9] J. Goldstone, Proc. R. Soc. A **239**, 267 (1957).
[10] N. M. Hugenholtz, Physica (Amsterdam) **23**, 481 (1957).
[11] J. Paldus and J. Čížek, Adv. Quantum Chem. **9**, 105 (1975).
[12] A. J. W. Thom, Ph.D. thesis, University of Cambridge, 2006.
[13] DALTON, Release 2.0, a molecular electronic structure program (2005), see http://www.kjemi.uio.no/software/dalton/dalton.html.
[14] T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).
[15] D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **100**, 2975 (1994).