

Accelerated Metropolis Method

C. J. Umrigar

Theory Center and Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853
(Received 19 April 1993)

It is shown that the freedom in the choice of the proposal matrix in the generalized Metropolis method can be used to greatly enhance the efficiency of the method. For example, the difficulties associated with the existence of multiple length scales in electronic structure calculations can be avoided by making an intelligent choice. Results are presented for Ne, Ar, and Li₂.

PACS numbers: 71.10.+x, 31.15.+q

Introduction.—In many branches of physics, simulations of systems with multiple length scales are very time consuming. Accelerated methods, that involve making collective moves of the degrees of freedom, have been developed for lattice problems [1–3]. Here we present an accelerated method generally applicable to continuum problems and demonstrate its efficiency by applying it to calculations of the electronic energy of atoms and molecules. In this case, much of the interesting physics or chemistry is related to the valence electrons but the size of the Monte Carlo (MC) moves is restricted by the much shorter length scale of the 1s core electrons [4]. It is shown that an accelerated Metropolis algorithm, wherein each electron attempts moves that are proportional to its distance from the nearest nucleus, enhances greatly the rate at which the system evolves.

The generalized Metropolis algorithm.—The importance of the Metropolis method [5] in computational science is due to the fact that it is a simple and powerful method for sampling *any* known distribution $f(\mathbf{R})$, where \mathbf{R} labels the degrees of freedom of the system, which may be continuous or discrete. We review here a generalization [6,7] of the Metropolis algorithm which yields an infinite family of algorithms depending on the choice of the proposal matrix $T(\mathbf{R}_f|\mathbf{R}_i)$. We show that the efficiency of the method can be enhanced greatly by a suitable choice of $T(\mathbf{R}_f|\mathbf{R}_i)$.

Let $T(\mathbf{R}_f|\mathbf{R}_i) = S(\mathbf{R}_f|\mathbf{R}_i)/I(\mathbf{R}_i)$, where $I(\mathbf{R}_i) = \int d\mathbf{R}_f \times S(\mathbf{R}_f|\mathbf{R}_i)$, be the probability for an attempted move from \mathbf{R}_i to \mathbf{R}_f and let $A(\mathbf{R}_f|\mathbf{R}_i)$ be the probability for the move to be accepted. Assuming ergodicity, the equilibrium distribution is $f(\mathbf{R})$, provided that the ratio of acceptance probabilities is chosen to satisfy detailed balance

$$\begin{aligned} \frac{A(\mathbf{R}_f|\mathbf{R}_i)}{A(\mathbf{R}_i|\mathbf{R}_f)} &= \frac{f(\mathbf{R}_f)}{f(\mathbf{R}_i)} \frac{T(\mathbf{R}_i|\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i)} \\ &= \frac{f(\mathbf{R}_f)}{f(\mathbf{R}_i)} \frac{S(\mathbf{R}_i|\mathbf{R}_f)}{S(\mathbf{R}_f|\mathbf{R}_i)} \frac{I(\mathbf{R}_i)}{I(\mathbf{R}_f)}. \end{aligned} \quad (1)$$

Note here that we have complete freedom [7] in the choice of the attempt probabilities T . For a given choice of T , the optimal choice for the acceptance probabilities, is given by

$$A(\mathbf{R}_f|\mathbf{R}_i) = \min \left\{ \frac{f(\mathbf{R}_f)}{f(\mathbf{R}_i)} \frac{T(\mathbf{R}_i|\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i)}, 1 \right\}. \quad (2)$$

For this choice, $A(\mathbf{R}_f|\mathbf{R}_i)$ is largest and therefore the system evolves the fastest. Many applications of the Metropolis method (and this is the form in which it was originally formulated [5]), make the simple but inefficient choice that T is symmetric, in which case the factors of T in Eq. (2) cancel.

The drawback of the Metropolis method is that the points sampled are sequentially correlated, resulting in a loss of computational efficiency. If it takes, on the average, T_{corr} MC moves for an observable to be decorrelated, then the effective number of independent observations in a MC run of length N is only N/T_{corr} . It is clearly advantageous to reduce the autocorrelation time T_{corr} . This can be done by either increasing the average size of the proposed moves or by increasing the acceptance of the moves. In order to prevent the acceptance from getting too small, it is common practice to restrict the moves to be in the neighborhood of \mathbf{R}_i by choosing $S(\mathbf{R}_f|\mathbf{R}_i)$ to be nonzero only within a domain $D(\mathbf{R}_i)$ of volume $\Omega(\mathbf{R}_i)$ around \mathbf{R}_i . For a given functional form of $S(\mathbf{R}_f|\mathbf{R}_i)$ the acceptance decreases as $\Omega(\mathbf{R}_i)$ (and therefore the average size of the proposed moves) increases, so, there exists an optimal $\Omega(\mathbf{R}_i)$ for which the system evolves the fastest.

In this Letter we propose a $S(\mathbf{R}_f|\mathbf{R}_i)$ with both large moves and large acceptance. $S(\mathbf{R}_f|\mathbf{R}_i)$ should be viewed as being a function of \mathbf{R}_f which depends parametrically on \mathbf{R}_i . Our task is to find a functional form for $S(\mathbf{R}_f|\mathbf{R}_i)$ such that $\int d\mathbf{R}_f S(\mathbf{R}_f|\mathbf{R}_i)$ is known, $T(\mathbf{R}_f|\mathbf{R}_i)$ can be sampled directly, the resulting proposed moves are large and the acceptance in Eq. (2) is large on average. It was observed in Ref. [7] that if $I(\mathbf{R}_i)$ is independent of \mathbf{R}_i then the choice $S(\mathbf{R}_f|\mathbf{R}_i) \sim f(\mathbf{R}_f)$ results in $A(\mathbf{R}_f|\mathbf{R}_i)/A(\mathbf{R}_i|\mathbf{R}_f) \approx 1$. However, the only way to have $I(\mathbf{R}_i)$ be independent of \mathbf{R}_i and $S(\mathbf{R}_f|\mathbf{R}_i) \sim f(\mathbf{R}_f)$ is to have $S(\mathbf{R}_f|\mathbf{R}_i)$ be independent of \mathbf{R}_i . However, for most $f(\mathbf{R})$ of interest, it is not possible to find a function that approximates it sufficiently well over the entire domain of $F(\mathbf{R})$ and which can be sampled directly. Hence, as mentioned before, $S(\mathbf{R}_f|\mathbf{R}_i)$ is chosen to be

nonzero only in some domain $D(\mathbf{R}_i)$ of volume $\Omega(\mathbf{R}_i)$ around \mathbf{R}_i , so that the proposed moves are within this domain. In that case $I(\mathbf{R}_i) \approx S(\mathbf{R}_f|\mathbf{R}_i)\Omega(\mathbf{R}_i)$ and Eq. (1) becomes

$$\frac{A(\mathbf{R}_f|\mathbf{R}_i)}{A(\mathbf{R}_i|\mathbf{R}_f)} \approx \frac{f(\mathbf{R}_f)}{f(\mathbf{R}_i)} \frac{S(\mathbf{R}_i|\mathbf{R}_f)}{S(\mathbf{R}_f|\mathbf{R}_i)} \frac{S(\mathbf{R}_i|\mathbf{R}_i)}{S(\mathbf{R}_f|\mathbf{R}_f)} \frac{\Omega(\mathbf{R}_i)}{\Omega(\mathbf{R}_f)}, \quad (3)$$

from which it is apparent that the choice

$$S(\mathbf{R}_f|\mathbf{R}_i) = g_i(\mathbf{R}_f)/\sqrt{\Omega(\mathbf{R}_f)} \quad (4)$$

yields $A(\mathbf{R}_f|\mathbf{R}_i)/A(\mathbf{R}_i|\mathbf{R}_f) \approx 1$, if $g_i(\mathbf{R}_f) \sim \sqrt{f(\mathbf{R}_f)}$ within $D(\mathbf{R}_i)$ [8]. To be more precise, it can be shown [9] by Taylor expansion that if the logarithmic derivatives of $g_i(\mathbf{R}_f)$ at \mathbf{R}_i equal those of $\sqrt{f(\mathbf{R}_f)}$ then the average acceptance goes as $1 + \mathcal{O}(\Delta^m)$, where Δ is the linear dimension of $D(\mathbf{R}_i)$. In general, $m=2$, but if $D(\mathbf{R}_i)$ is inversion symmetric with \mathbf{R}_i at its center, then $m=3$. This is a considerable improvement compared to using a symmetric $S(\mathbf{R}_f|\mathbf{R}_i)$ or choosing $S(\mathbf{R}_f|\mathbf{R}_i) \sim f(\mathbf{R}_f)$ for either of which the average acceptance goes as $1 + \mathcal{O}(\Delta)$.

In the case of electronic structure calculations the probability distribution $f(\mathbf{R})$ is $|\Psi(\mathbf{R})|^2$, where $\Psi(\mathbf{R})$ is a trial wave function and \mathbf{R} specifies the $3n$ electron coordinates. Hence Eq. (4) becomes

$$S(\mathbf{R}_f|\mathbf{R}_i) = |\Phi_i(\mathbf{R}_f)|/\sqrt{\Omega(\mathbf{R}_f)}, \quad (5)$$

where $\Phi_i(\mathbf{R}_f) \sim \Psi(\mathbf{R}_f)$ in $D(\mathbf{R}_i)$ and has logarithmic derivatives at \mathbf{R}_i that match those of $\Psi(\mathbf{R}_f)$. We now discuss the explicit forms of $S(\mathbf{R}_f|\mathbf{R}_i)$ for which we present results here. For the first two choices $\Omega(\mathbf{R}_i)$ is constant, independent of \mathbf{R}_i .

Simple metropolis.—The simplest form is $S(\mathbf{R}_f|\mathbf{R}_i)$ equal to a constant when \mathbf{R}_f is within a $3n$ -dimensional hypercube (box) of linear dimension 2Δ centered at \mathbf{R}_i and zero elsewhere. Then $T(\mathbf{R}_f|\mathbf{R}_i) = (2\Delta)^{-3n}$ within the box and zero elsewhere. Aside from the question of whether one or all particles are moved in a single MC update, this is the original Metropolis method [5] and it is the form in which it has most frequency been used.

Cartesian coordinate directed Metropolis.—According to Eq. (5), an improved form of S is $S(\mathbf{R}_f|\mathbf{R}_i) = |\Psi_i(\mathbf{R}_f)|$ within the box. The simplest choice of $\Phi_i(\mathbf{R}_f)$ is a linear approximation (in each of the $3n$ Cartesian coordinate directions) to $\Psi(\mathbf{R}_f)$ at \mathbf{R}_i .

Two features of the wave function restrict the size of the attempted moves. First, probable electron configurations have two core electrons at a distance of $\mathcal{O}(1/Z)$ Bohr radii from each nucleus, whereas valence electrons are typically a distance $\mathcal{O}(1)$ from the nearest nucleus [10]. Hence, if the same value of Δ is used for all the electrons, as is usually the case, then the core electrons set the length scale for all the electrons, else a large fraction of the proposed moves are rejected. Second, $\Psi(\mathbf{R})$

has a derivative discontinuity when an electron is at a nucleus, which renders any Taylor series approximation, in Cartesian coordinates, of $\Psi(\mathbf{R})$, very inaccurate. (There is also a derivative discontinuity when two electrons overlap, but this is less problematic since electrons repel each other.) A natural solution to both these problems is provided by the use of spherical polar rather than Cartesian coordinates in proposing the Monte Carlo moves.

Spherical polar coordinate directed Metropolis.—We now describe a choice for $S(\mathbf{R}_f|\mathbf{R}_i)$ which allows each electron to make a move appropriate to its length scale and which avoids the derivative discontinuity in $\Psi(\mathbf{R})$. Each electron moves in a volume which is the intersection of a cone which subtends an angle $2\theta_M$ at the closest nucleus and a concentric spherical annulus. Let $r_{k,i}$ be the initial distance of the k th electron from the nucleus closest to it. The proposed move will use spherical polar coordinates centered at that nucleus. The radial moves are made in the interval $(r_{k,i}/\Delta_r, r_{k,i}\Delta_r)$. Hence the size of the radial moves is proportional to $r_{k,i}$. For atoms, this ensures that if a move is possible, then the reverse move is also possible. We will discuss the necessary changes for molecules and solids later. The advantages of moving in this volume are that the size of the valence electron moves is not restricted by the core electrons and that Taylor expansions of the wave function are valid in the region of interest since $\Psi(\mathbf{R})$ does not have derivative discontinuities in spherical polar coordinates at nuclei.

Since

$$\Omega(\mathbf{R}_f) = \prod_{k=1}^n (2\pi/3)(1 - \cos\theta_M)(\Delta_r^3 - 1/\Delta_r^3)r_{k,f}^3,$$

according to Eq. (5), $S(\mathbf{R}_f|\mathbf{R}_i) = |\Phi_i(\mathbf{R}_f)|/\prod_{k=1}^n r_{k,f}^{3/2}$. In order to simplify the sampling, $\Phi_i(\mathbf{R}_f)$ is chosen to have the form

$$\Phi_i(\mathbf{R}_f) = \prod_{k=1}^n W_i(\phi_{k,f}|\theta_{k,f}, r_{k,f}) V_i(\theta_{k,f}|r_{k,f}) U_i(r_{k,f}).$$

For each electron k , the radial coordinate $r_{k,f}$ is sampled from $|U_i(r_{k,f})|\sqrt{r_{k,f}}$, then $\theta_{k,f}$ is sampled from $|V_i(\theta_{k,f}|r_{k,f})|\sin\theta_{k,f}$ conditional upon $r_{k,f}$ and finally $\phi_{k,f}$ is sampled from $|W_i(\phi_{k,f}|\theta_{k,f}, r_{k,f})|$ conditional on $r_{k,f}$ and $\theta_{k,f}$. The precise choice of the functions W_i, V_i, U_i is described elsewhere [9].

It is possible to make further improvements by making θ_M a function $r_{k,f}$ and $r_{k,i}$. The reason it is advantageous to do so is that most trial functions have a finite discontinuity of magnitude Z in the local energy when two electrons approach a nucleus [11], Z being the nuclear charge. In this limit the local energy is Z hartrees higher when the two electrons and the nucleus lie along a straight line with the electrons on opposite sides of the nucleus than with the electrons on the same side. It is desirable to average over this discontinuity as rapidly as possible by making large angular moves when $r_{av} \ll 1/Z$, where $r_{av} = (r_{k,i} + r_{k,f})/2$. Hence $\cos\theta_M$ is chosen to be

$$\cos\theta_M = \cos\theta_m - \frac{1 + \cos\theta_m}{1 + (Zr_{av})^2}, \quad (6)$$

where θ_m is fixed, which has the limiting behaviors $\theta_M = \theta_m$ when $r_{av} \gg 1/Z$ and $\theta_M = \pi$ when $r_{av} \ll 1/Z$. In spite of the fact that when an electron is close to a nucleus, the angular move is made over the entire surface of a sphere, the acceptance of these moves is close to one [9]. On the other hand when moves are made in Cartesian coordinates, the acceptance for the electrons close to the nucleus is very low [9].

Finally we note that although each of the above algorithms assumes that all the electrons are moved during each Monte Carlo update, it is trivial to modify the algorithms to move only one electron at each MC step or in fact any number in between.

Generalization to molecules and solids.—The additional complication for molecules and solids is that the closest nuclei to each of the n electrons at \mathbf{R}_i need not be the closest nuclei to the corresponding electrons at \mathbf{R}_f . For some fraction of these the reverse move from \mathbf{R}_f to \mathbf{R}_i is not possible, i.e., $S(\mathbf{R}_i|\mathbf{R}_f) = 0$, because whereas \mathbf{R}_f lies in $D(\mathbf{R}_i)$, \mathbf{R}_i may not lie in $D(\mathbf{R}_f)$. In that case detailed balance demands that the move from \mathbf{R}_i to \mathbf{R}_f be rejected, i.e., $A(\mathbf{R}_f|\mathbf{R}_i) = 0$. Since these rejections may be performed on purely geometrical grounds (they do not require calculating the wave function or its derivatives at \mathbf{R}_f), this does not result in an appreciable loss of efficiency.

Results.—The efficiency of the algorithms is inversely proportional to the autocorrelation time of observables of interest. Table I shows the autocorrelation time of the energy for four algorithms and four wave functions. The four algorithms are (1) the simple Metropolis algorithm

moving all electrons at each MC step, (2) the Cartesian coordinate directed Metropolis algorithm moving all electrons at each MC step, (3) the Cartesian coordinate directed Metropolis algorithm moving only one electron at each MC step, and (4) the spherical-polar coordinate directed Metropolis algorithm moving only one electron at each MC step. The four wave functions used are (1) a simple Ne wave function, (2) a good Ne wave function, (3) a simple Ar wave function, and (4) a simple Li_2 wave function. The simple wave functions consist of a determinant multiplied by a simple Jastrow function which is a function of the interelectron distances only, whereas the good wave function consists of a determinant multiplied by a more complicated Jastrow function which is a function of both the interelectron distances and the electron-nuclear distances [12].

The measure of efficiency of the algorithm is the autocorrelation time which is determined as follows. The entire MC run (after discarding the equilibration updates) consists of N MC updates that are divided into N_b blocks each consisting of N_s MC steps for each of the n electrons. The local energy is measured after each MC update. The autocorrelation time is given by $T_{\text{corr}} = N_s(\sigma_b/\sigma)^2$, where σ and σ_b are the rms fluctuations of the individual energies and the block average energies, respectively. N_s must be chosen such that $N_s \gg T_{\text{corr}}$. The autocorrelation times presented in Table I were obtained using values of N_s that were at least 100 times greater than T_{corr} . It was found that using $N_s \approx 10 T_{\text{corr}}$ resulted in estimates of T_{corr} that were too low by as much as 20%. When MC moves consist of moving one electron at a time it takes twice the computer time to move all the electrons as compared to when they are all moved at once [7]. Hence Table I has values of $T_{\text{corr}}^* = 2T_{\text{corr}}$ for algo-

TABLE I. Autocorrelation times for the four wave functions and the four algorithms discussed in the text. In order to have a fair comparison, $T_{\text{corr}}^* = T_{\text{corr}}$ for algorithms 1 and 2 and $T_{\text{corr}}^* = 2T_{\text{corr}}$ for algorithms 3 and 4. In algorithm 4, $\theta_m = \pi/2$ for Ne and Ar and $\theta_m = \pi$ for Li_2 . \bar{A} is the average acceptance. The uncertainty in T_{corr}^* is typically 10% of its value.

Wave function	Algorithm	Δ	Δ_r	\bar{A}	T_{corr}^*
$\bar{E} = -128.716$ hartree $E_{\text{corr}} = 43\%$ $\sigma = 1.8$ hartree	Ne simple 1	0.25	...	0.288	84
	2	0.3	...	0.661	28
	3	0.8	...	0.769	13
	4	...	5	0.708	2.0
$\bar{E} = -128.901$ hartree $E_{\text{corr}} = 91\%$ $\sigma = 0.91$ hartree	Ne good 1	0.25	...	0.290	37
	2	0.3	...	0.663	11
	3	0.8	...	0.771	7.2
	4	...	5	0.708	1.7
$\bar{E} = -527.1$ hartree $E_{\text{corr}} = 37\%$ $\sigma = 4.0$ hartree	Ar simple 1	0.12	...	0.307	190
	2	0.2	...	0.474	44
	3	0.5	...	0.813	13
	4	...	5	0.620	2.2
$\bar{E} = -14.9476$ hartree $E_{\text{corr}} = 61\%$ $\sigma = 0.41$ hartree	Li_2 simple 1	0.75	...	0.268	210
	2	1	...	0.614	56
	3	2	...	0.616	32
	4	...	8	0.775	5.8

gorithms 3 and 4. We note that σ and the variational energy \bar{E} depend on the trial wave function but not on the algorithm used. T_{corr} and the acceptance \bar{A} depend on both the algorithm and the trial wave function.

For each of the algorithms the values of Δ or Δ_r and θ_m were optimized to yield the smallest possible T_{corr} . Table I shows the values of T_{corr}^* for approximately optimal values of the parameters. Moderate variations of the parameters about their optimal values affects the efficiency of the algorithm only slightly. For example, changing the value of Δ_r from 5 to 4 altered the value of T_{corr} by less than 10% for each of the wave functions. For the four wave functions, the autocorrelation times get smaller by factors of 42, 22, 86, and 36, respectively, as we progress from algorithm 1 to 4 and by factors of 6.5, 4.2, 5.9, and 5.5 in going from algorithm 3 to 4.

Note that for each of the four algorithms T_{corr} is smaller for the good Ne wave function than for the simple Ne wave function. The reason for this is that the good Ne wave function has a more rapidly varying local energy (but with a smaller amplitude of course) and consequently it takes fewer Monte Carlo steps to wander from a region where the local energy is too high to one where it is too low and vice versa. Hence the gain in efficiency from improving the wave function is greater than would be supposed by merely comparing their respective values of σ .

The variational energy for the good Ne wave function is -128.9005 ± 0.0005 hartree, corresponding to 91% of the correlation energy. This is the lowest energy calculated to date by variational Monte Carlo for Ne. The fact that the energy is good is due to the high quality of the wave function, but the fact that the energy could be determined with a small statistical error, in just a few hours on a workstation, is testimony to the efficiency of the new algorithm.

Comparison of the results for algorithms 2 and 3 shows, as has been noted before [7], that for systems with many electrons, and for this class of algorithms, it is more efficient to move one electron at a time rather than all at once.

For algorithm 1 there is a considerable increase in T_{corr} in going from Ne to Ar. On the other hand for algorithm 4, T_{corr} increases very little. This does not mean that calculations of heavy atoms can be performed as rapidly as those of light atoms. The time for evaluating determinants in the wave functions scales as Z^3 and the fluctuation in the local energy σ is empirically found to scale

roughly as Z or $Z^{1.5}$. Consequently the computer time required to obtain results with a fixed statistical uncertainty increases as Z^5 or Z^6 if T_{corr} is independent of Z and yet more rapidly otherwise.

In conclusion it was shown that an efficient choice of the proposal matrix in the generalized Metropolis method is $S(\mathbf{R}_f|\mathbf{R}_i) \sim \sqrt{f(\mathbf{R}_f)/\Omega(\mathbf{R}_f)}$. Large gains in efficiency were demonstrated in calculations of the electronic energy of atoms and molecules.

I thank Malvin Kalos for his encouragement and Malvin Kalos, Peter Nightingale, Karl Runge, and Kevin Schmidt for useful discussions. This work was supported by the Office of Naval Research. The computations were performed on IBM RS-6000 workstations provided by the Cornell National Supercomputer Facility and Cornell-IBM Computing for Scientific Research Joint Study.

-
- [1] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86 (1987).
 - [2] U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
 - [3] See review by A. D. Sokal, in *Computer Simulation Studies in Condensed Matter Physics, Recent Developments*, edited by D. P. Landau, K. K. Mon, and H. B. Schüttler, Springer Proceedings in Physics Vol. 33 (Springer, Berlin, 1988).
 - [4] P. J. Reynolds, Int. J. Quant. Chem. Symp. **24**, 679 (1990); G. G. Batrouni and P. J. Reynolds (to be published) have attacked this problem by a different method to that used here.
 - [5] N. Metropolis *et al.*, J. Chem. Phys. **21**, 1087 (1953).
 - [6] W. K. Hastings, Biometrika **57**, 97 (1970).
 - [7] D. Ceperley, G. V. Chester, and M. H. Kalos, Phys. Rev. B **16**, 3081 (1977).
 - [8] M. Rao and B. J. Berne, J. Chem. Phys. **71**, 129 (1979), introduced an adjustable parameter λ in their force-bias method. Although they do not describe their work in these terms, we note that they obtained better results for $\lambda = 1/2$ corresponding to $S(\mathbf{R}_f|\mathbf{R}_i) \sim \sqrt{f(\mathbf{R}_f)}$ than for $\lambda = 1$ corresponding to $S(\mathbf{R}_f|\mathbf{R}_i) \sim f(\mathbf{R}_f)$.
 - [9] C. J. Umrigar (unpublished).
 - [10] The terms *valence electron* or *core electron* are only used to indicate the instantaneous distance of an electron from the nearest nucleus. During the course of a Monte Carlo run a given electron will be in both the valence and core regions part of the time.
 - [11] C. J. Myers, C. J. Umrigar, J. P. Sethna, and J. D. Morgan, Phys. Rev. A **44**, 5537 (1991).
 - [12] C. J. Umrigar, Int. J. Quant. Chem. Symp. **23**, 217 (1989), and references therein.