CHAPTER 2

# Applications of Post-Hartree–Fock Methods: A Tutorial

**Rodney J. Bartlett and John F. Stanton**

*Quantum Theory Project, Departments of Chemistry and Physics, University of Florida, Gainesville, Florida 32611*

## INTRODUCTION

This chapter is meant to be a condensed(!) tutorial on the intelligent use of post-Hartree–Fock (correlated) methods for the determination of molecular structure and spectra.[1] The content is directed at users of the ACES II, GAUSSIAN, CADPAC, HONDO, or GAMESS[2] type of ab initio program, or anyone who would appreciate a broader knowledge of the modern treatment of electron correlation for molecules. The ready availability and applicability of these programs might be said to be one of the principal contributions of ab initio quantum chemistry to science, inasmuch as these systems provide a probe of structural, spectral, and reactivity characteristics frequently unavailable from experiment or, alternatively, facilitate interpretation of available experimental data. However, the ease of using these programs (and their official-looking output!) can mask many difficult situations that might not be recognized by users, resulting in misinterpretation of computational results. Theoreticians who develop such program systems are aware of the potential shortcomings of each approximation, but many users lack a full understanding of the

*Reviews in Computational Chemistry*, Volume V
Kenny B. Lipkowitz and Donald B. Boyd, Editors
VCH Publishers, Inc. New York, © 1994

theoretical foundations of the various methods. We try to explain some of these considerations, and how they affect the choice of method and basis set for a given application. Clearly, no single chapter can replace several courses in quantum chemistry and years of hands-on experience in the use of correlated electronic structure methods, but we think it is possible to provide a guide that provides much of the essential information required for intelligent application of the current level of post-Hartree–Fock methods without becoming "too theoretical."

At the same time, we must recognize that quantum chemistry is an ongoing research area. Though it is mature enough to provide useful tools for a wide range of applications, even a year ago many of the tools described here and recently incorporated into the ACES II[3] program system did not exist. Among other things, this means that the useful lifetime of any tutorial is finite, but by including some recent developments not widely known to the current user community, we hope that this chapter will have an effective lifetime that is relatively long. No unique recipe for doing post-Hartree–Fock calculations can be offered, but many of the considerations pertinent to such calculations can be identified and emphasized. In the application areas, by following the sequence from self-consistent field (SCF) to the highest current levels of correlated theory, we hope to show how approaches to different problems have evolved, and continue to progress, to provide better and more comprehensive results.

After a discussion of the independent particle or Hartree–Fock (SCF) approximation[4] to introduce some basic concepts, we discuss the correlation problem and available methods for its treatment.[5-10] We then address several topics that cover most of the applications of interest to users of quantum chemical program systems. These include molecular geometries, vibrational spectra, photoelectron spectra, electronic spectra, first- and second-order properties, and nuclear magnetic resonance spectra. In each category of application, we address the range of applicability of SCF methods and the degree of correlation correction required, and we state what can be expected, numerically, from increasing levels of rigor. Our objective is to provide the reader with both a conceptual understanding and some appreciation of the approximate uncertainty for a given level of theory. Our intent is not to present the voluminous results on which these "error bars" are based—that is left to the literature—but instead to pick a few standard and a few difficult molecules that are representative of results one should expect. We do not derive equations, but we do present basic equations to help the reader better appreciate the nature of the approach. Just as a picture can be worth a thousand words, an equation will often do the job of hundreds, and only in this way can we attempt to maintain some brevity while minimizing impreciseness. We assume only that the reader has some familiarity with Hartree–Fock theory.[4] On the other hand, nothing we say should unduly depend on appreciation of mathematical arguments—more conceptual, complementary explanations are employed whenever possible.

To be most useful, we necessarily address the correlated methods that are

most readily applicable to the majority of problems of interest to chemists, thereby imposing several requirements that a method must meet. These include treating different molecular geometries with effectively equal accuracy; equivalent applications to open- and closed-shell systems; treatment of excited, ionized, or electron-attached states as well as ground states; and predictions of other properties such as moments and polarizabilities. Also the term "post-Hartree–Fock" implies a single determinant starting point. Consequently, because of their wide applicability and ease of use, we emphasize methods built upon a single reference when possible. These include many-body perturbation theory[5] (MBPT, also known as MP in some program systems[8]), and its infinite-order coupled-cluster (CC) generalization,[6,11,12] [including quadratic configuration interaction (QCI)[13] as a special case] as well as some configuration interaction (CI) methods.[9] However, we also mention situations for which multi (instead of single) determinant references are appropriate,[14] such as excited states and certain transition states, and provide some alternative readily applicable tools that are no more difficult to understand and use than standard single reference theory.

To avoid being discouraged by discussions of theory, some readers might prefer a "results-oriented" sequence. For those readers, a reasonable sequence would be to start with the section on molecular geometries, continue through the applications areas, and treat the preceding sections as appendices that define the various levels of theory. The sections on theory, however, were written to be clear and thorough.

Most of the results reported here were obtained with the ACES II program system.[3] The program provides CC/MBPT results for molecules, open or closed shell, using UHF, ROHF, RHF and QRHF reference functions (see later for acronyms). It offers analytical gradients for most such methods including MBPT (2), (3) and (4), CCSD, CCSD(T), and analytical Hessians for MBPT(2) with RHF, UHF or ROHF reference functions. ACES II requires only simple Z-matrix (bond distance and angle) input plus key words that identify basis sets (most modern basis sets are catalogued), level of correlation, degrees of freedom to optimize, and other options. The program automatically introduces the available molecular symmetry and employs the largest Abelian subgroup to simplify and speed CC/MBPT calculations. ACES II provides CC methods for excited states (UV-vis spectroscopy), ionized states (photoelectron spectroscopy), molecular properties (ESR, NMR), and relativistic corrections. The program has a number of analysis tools to facilitate interpretation, locate optimum geometries and transition states, and evaluate vibrational spectra.

## INDEPENDENT PARTICLE MODEL

The reference framework for most theoretical descriptions of molecular electronic structure and spectra is the independent particle or molecular orbital

(MO) model. In quantum mechanics, the energy of a molecule is based on the Hamiltonian operator

$$\mathcal{H} = -\frac{1}{2}\sum_i \nabla^2(\mathbf{r}_i) - \sum_\alpha \sum_i Z_\alpha / r_{\alpha i} + \frac{1}{2}\sum_{i,j}{}' \frac{1}{r_{ij}} + \frac{1}{2}\sum_{\alpha,\beta}{}' \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} \qquad [1]$$

(the prime eliminates the $i = j$ and $\alpha = \beta$ term). The first term is the kinetic energy, which is given by a sum over one-particle operators corresponding to each electron $\mathbf{r}_i$. The second one-electron term is the Coulombic electron–nuclear attraction term, where $Z_\alpha$ is the atomic number for the nucleus $\alpha$, and $r_{\alpha i}$ measures the distance of electron $i$ from the $\alpha$th nucleus. The third term is the two-electron Coulomb repulsion term, where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between the $i$th and $j$th electrons. The final term is the nuclear–nuclear repulsion, where $R_{\alpha\beta} = |\mathbf{R}_\alpha - \mathbf{R}_\beta|$. In the Born–Oppenheimer "clamped nuclei" approximation that underlies nearly all molecular calculations, this term is a constant determined by the nuclear coordinates and is simply added at the end of a calculation. It introduces the characteristic repulsive wall for a potential curve and, thus, is important in the solution of the vibrational Schrödinger equation.

Hence for the electronic Schrödinger equation, our objective is the solution to $\mathcal{H}\Psi_k = E_k\Psi_k$ for the spectrum of electronic states $\{\Psi_k\}$, where

$$\mathcal{H} = \sum_i h(i) + \frac{1}{2}\sum_{i,j}{}' \frac{1}{r_{ij}} \qquad [2]$$

with the two one-electron parts combined into $h(i)$.

The delightful thing about one-electron operators is that we can exactly solve the Schrödinger equation if the Hamiltonian is approximated by its one-electron part ($\mathcal{H} \approx \Sigma_i h(i) = H_0$) since a "separable" wavefunction can be constructed as a product of one-particle functions, $\phi_j(\mathbf{r}_j)$,

$$\Phi_0(r;R) = \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\cdots\phi_n(\mathbf{r}_n) \qquad [3]$$

where the notation $\Phi_0(r;R)$ means a function of all electron coordinates collectively indicated by $r$, that through the Born–Oppenheimer approximation is parametrically dependent on the collective location of the nuclei, $R$. Using Eq. [3] we would have $[\Sigma_j h(j)]\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\cdots\phi_n(\mathbf{r}_n) = [h(\mathbf{r}_1)\phi_1(\mathbf{r}_1)]\phi_1(\mathbf{r}_2)\cdots\phi_n(\mathbf{r}_n) + \phi_1(\mathbf{r}_1)[h(\mathbf{r}_2)\phi_2(\mathbf{r}_2)]\phi_3(\mathbf{r}_3)\cdots + \cdots\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\cdots[h(\mathbf{r}_n)\phi_n(\mathbf{r}_n)]$. Then by knowing that each one-particle function was an eigenfunction to $h(\mathbf{r}_j)$ with energy $\epsilon_j$,

$$h(\mathbf{r}_1)\phi_j(\mathbf{r}_1) = \epsilon_j\phi_j(\mathbf{r}_1) \qquad [4]$$

we would have

$$H_0\Phi_0 = E_0\Phi_0 \qquad [5]$$

$$E_0 = \sum_{j=1}^{n} \epsilon_j \qquad [6]$$

However, we can do even better without sacrificing this convenient separability. If we are able to approximate the two-particle term as some "average" one-particle operator, $u(\mathbf{r}_j)$, where, for example, we might average over the repulsion of all other electrons in the system toward electron $j$, then we obtain the expression

$$\begin{aligned}\mathcal{H} &= \sum_{j=1}^{n} [h(\mathbf{r}_j) + u(\mathbf{r}_j)] + \frac{1}{2}\sum_{i,j}{}' \frac{1}{r_{ij}} - \sum_{j=1}^{n} u(\mathbf{r}_j) \\ &= \sum_{j=1}^{n} \mathcal{F}(\mathbf{r}_j) + V = H_0 + V\end{aligned} \qquad [7]$$

If $V$ is small (and it is clearly smaller than the two-particle term itself), then we might expect that replacement of $\mathcal{H}$ by $H_0$ (which is a separable sum of one-particle Hamiltonians, $\mathcal{F}(\mathbf{r}_j)$) to be a good approximation. This simplification allows us to reduce the $n$-particle problem to a set of one-particle eigenvalue problems,

$$\mathcal{F}(\mathbf{r}_j)\phi_j(\mathbf{r}_j) = \epsilon_j\phi_j(\mathbf{r}_j) \qquad [8]$$

and is the basis of molecular orbital theory. An MO is an eigenfunction of an effective one-particle Hamiltonian, and the electron in the MO is considered to have an orbital energy of $\epsilon_j$. Eigenfunctions to a Hermitian operator $\{\phi_j\}$ are either orthonormal or may be chosen to be orthonormal in degenerate cases, so $\langle\phi_i|\phi_j\rangle = \delta_{ij}$.

Although the basic concept will not change, two additional modifications must be made to this simple theory to make it consistent with other aspects of quantum mechanics. First, to fully characterize electrons we need to include spin. Consequently, this requires that to each spatial one-particle function $\phi_j(\mathbf{r}_j)$ we attach $\alpha$ (spin-up) or $\beta$ (spin-down). Hence, each one-particle function becomes $\phi_j(\mathbf{r}_j)\alpha(\omega_j)$ or $\phi_j(\mathbf{r}_j)\beta(\omega_j)$, where $\omega_j$ indicates a coordinate in spin space. We can use an even shorter notation by specifying $\varphi_j(x_j) = \phi_j(\mathbf{r}_j)\alpha(\omega_j)$ to be spin orbitals, where the combined space-spin coordinate is $x_j$. It is still true that $\mathcal{F}(x_j)\varphi_j(x_j) = \epsilon_j\varphi_j(x_j)$.

Second, suitable solutions to the Schrödinger equation for electrons must have appropriate permutational symmetry. That is, an interchange of the space-spin coordinates $x_i$ and $x_j$ must not alter the probability density $|\Psi|^2$.[15] For this to be true, the interchange operator $P_{ij}$ has to have the effect, $P_{ij}\Psi = \pm\Psi$. For

electrons and other fermions, the minus applies (the Pauli exclusion principle), requiring that our wavefunctions be antisymmetrized products that assume the determinantal form

$$\Phi_0 = \frac{1}{\sqrt{n!}} \begin{vmatrix} \varphi_1(x_1) & \varphi_2(x_2) \ldots & \varphi_n(x_n) \\ \varphi_2(x_1) & \ldots & \ldots \\ \vdots & & \vdots \\ \varphi_n(x_1) & \ldots & \varphi_n(x_n) \end{vmatrix} = \mathcal{A}(\varphi_1(x_1) \ldots \varphi_n(x_n)) \qquad [9]$$

The operator $\mathcal{A} = (1/\sqrt{n!}) \sum_P (-1)^p P$, where $P$ presents a permutation of electron labels. $\mathcal{A}$ forms a determinant by permuting all the space-spin labels in the simple product multiplied by the parity factor $(-1)^p$ and normalizing the wavefunction.

It is fairly easy to see that each of the $n!$ products introduced by the determinant will behave in exactly the same way (i.e., $H_0$ and $\mathcal{A}$ commute) resulting in the same one-particle equations (suppressing the $x$),

$$\mathcal{F}(1)\varphi_j(1) = \epsilon_j \varphi_j(1) \qquad [10]$$

and in particular

$$\begin{aligned} H_0 \mathcal{A}(\varphi_1(1) \cdots \varphi_n(n)) &= E_0 \mathcal{A}(\varphi_1(1) \cdots \varphi_n(n)) \\ H_0 \Phi_0 &= E_0 \Phi_0 \end{aligned} \qquad [11]$$

In this simple scheme, such other eigenstates of $H_0$ as $\Phi_j^a$ are formed by replacing one of the molecular orbitals $\varphi_j(j)$ occupied by electron $j$ by some orbital unoccupied in the ground state, call it $\varphi_a(j)$, that is also an eigenfunction of $\mathcal{F}(i)$. Thus we would have

$$\begin{aligned} H_0 \Phi_j^a &= H_0 \mathcal{A}(\varphi_1(1) \cdots \varphi_a(j) \cdots \varphi_n(n)) \\ &= E_j^a \mathcal{A}(\varphi_1(1) \cdots \varphi_a(j) \cdots \varphi_n(n)) \\ &= E_j^a \Phi_j^a \end{aligned} \qquad [12]$$

where the eigenvalue is written

$$\begin{aligned} E_j^a &= \left( \sum_{k(\neq j)} \epsilon_k \right) + \epsilon_a \\ &= E_0 + \epsilon_a - \epsilon_j \end{aligned} \qquad [13]$$

The general independent particle model described above becomes the Hartree–Fock, self-consistent field (SCF) approximation when $\Phi_0$ is the ener-

getically "best" single determinant approximation to $\Psi$. (In some SCF approaches, $\Phi_0$ is energetically optimal subject to certain constraints on the orbitals.) That is, we use the variational principle, $E = \langle\Phi_0|\mathcal{H}|\Phi_0\rangle \geq E_{\text{exact}}$ to vary the form of the $n$-spin orbitals $\{\varphi_i\}$ subject to maintaining their orthogonality $\langle\varphi_i|\varphi_j\rangle = \delta_{ij}$ until we obtain the lowest possible energy, $E_{\text{SCF}}$. In the course of the energy variation, we find that

$$\mathcal{F}(1) = h(1) + u(1) = h(1) + \sum_{j=1}^{n} \int \varphi_j^*(2) \frac{1}{r_{12}} (1 - P_{12})\varphi_j(2) d\tau_2 \qquad [14]$$

which defines the operator, $u(1)$. The volume element in space-spin coordinates is $d\tau$. Self-consistency in the solution of Eq. [10] is necessary because the form of $\mathcal{F}$ is dependent on the form of the spin orbitals, $\{\varphi_j\}$ via the $u(1)$ operator of Eq. [14].

Furthermore,

$$\begin{aligned} E_{SCF} &= \langle\Phi_0|\mathcal{H}|\Phi_0\rangle = Tr(\mathcal{H}\rho) \\ &= \sum_i \langle\varphi_i|h(1)|\varphi_i\rangle + \frac{1}{2}\sum_{i,j}\langle\varphi_i\varphi_j\|\varphi_i\varphi_j\rangle \\ &= \sum_i \langle i|h|i\rangle + \sum_{i<j} \langle ij\|ij\rangle \\ &= \sum_i \epsilon_i - \sum_{i<j} \langle ij\|ij\rangle \\ &= E_0 + \langle\Phi_0|V|\Phi_0\rangle \end{aligned} \qquad [15]$$

so $E_{\text{SCF}}$ is not a simple sum of orbital energies. The double bar means $\langle ij\|ij\rangle = \langle ij|ij\rangle - \langle ij|ji\rangle$, where $\langle ij|ij\rangle = \int \varphi_i^*(1)\varphi_j^*(2)(1/r_{12})\varphi_i(1)\varphi_j(2)d\tau_1 d\tau_2 = (ii|jj)$.

These particular spin-orbital equations are termed the unrestricted Hartree-Fock (UHF) equations, since variation here does not assume that all spatial orbitals are doubly occupied. That is, we permit $\varphi_j = \phi_j\alpha$ and $\varphi_{j+1} = \phi_j'\beta$, where $\phi_j \neq \phi_j'$. For a closed-shell system, we can insist on getting the best energy while having $\phi_j = \phi_j'$, which defines the restricted Hartree–Fock (RHF) solution. Hence, UHF is formally the same for open and closed shells, whereas RHF applies only to the latter. For high spin open shells, we can also insist that our solutions have maximum double occupancy as shown in Figure 1. These are referred to as restricted open-shell Hartree–Fock (ROHF) solutions. For such a case there is a more complicated $\mathcal{F}(1)$ operator,[4] but the orthonormal orbitals that emerge are degenerate. These different SCF models may be succinctly compared by inspection of Figure 1. We collectively refer to any of the foregoing approaches as "SCF" methods.

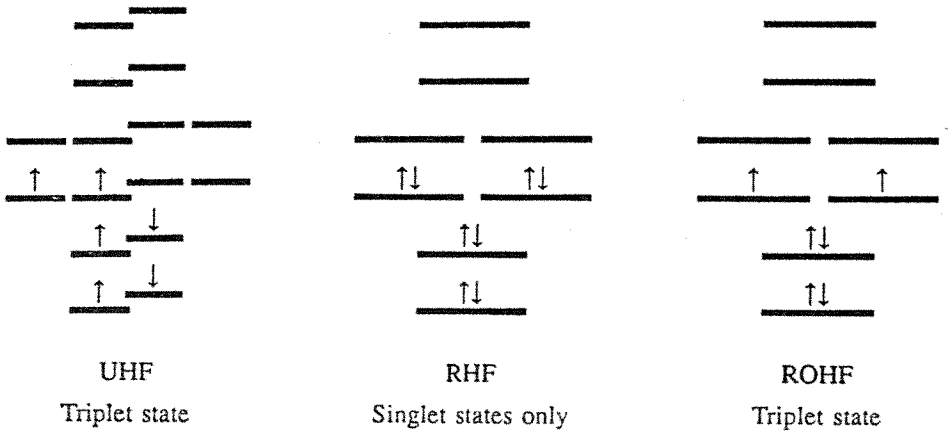


**Figure 1** Orbital energy level diagrams for UHF, RHF, and ROHF reference functions.

The result of an SCF calculation is the wavefunction $\Phi_0 = \Phi_{SCF}$, consisting of the MOs $\{\varphi_i\}$, the density $\rho = \Phi_i \Phi_i^*$, from which properties for some operator $\theta$ are given by $Tr(\theta\rho)$; the total energy $E_{SCF}$; and in the canonical UHF or RHF case, the orbital energies $\epsilon_i$, which provide an (unrelaxed) estimate of the ionization potential via Koopmans' theorem, $\epsilon_i = -I(i)$, for an electron in the $i$th MO. The approximation of interelectronic repulsion by the effective one-particle operator, $u(1)$ in Eq. [14], represents the potential one electron feels when moving in an average field generated by the remaining electrons. This follows because in the sum over $j$ in Eq. [15], when $j = i$, we have $\langle ii \| ii \rangle = 0$.

All the SCF methods mentioned above are independent particle models. Since they serve as the reference for the post-Hartree–Fock methods discussed in this chapter, we should be aware of their properties. RHF applies only to closed-shell systems and is a singlet eigenfunction of spin. (Some authors insist that "restricted" must pertain to spatial symmetry as well as spin symmetry, but GAUSSIAN, ACES II, and other program systems can in principle converge to broken spatial symmetry solutions even for closed shells, so we will not insist that RHF solutions be restricted to spatial symmetry eigenfunctions, too.) As is well known, and shown in Figure 2 for $H_2$, an RHF solution does not separate correctly to open-shell fragments on a potential energy curve. (RHF separates correctly for closed-shell products such as $B_2H_6 \rightarrow 2BH_3$, or $Be_2 \rightarrow 2Be$, or $HeH^+ \rightarrow He + H^+$.) That is because RHF uses equal orbitals for different spins at all internuclear separations, which does not allow the orbitals to localize on the individual atoms. The RHF wavefunction for $H_2$ is $|1\sigma_g\alpha 1\sigma_g\beta|$. Yet to get two H(1s) atoms at separation we must have $1s_A = 1\sigma_g + 1\sigma_u$ and $1s_B = 1\sigma_g - 1\sigma_u$, where $g$ and $u$ refer to *gerade* (even) and *ungerade* (odd). Hence, correct separation would require the two determinants $C_1|1\sigma_g\alpha 1\sigma_g\beta| \pm C_2|1\sigma_u\alpha 1\sigma_u\beta|$, where $C_1 = C_2$ at separation but $C_1 \gg C_2$ at $R_e$. This two-determinant wave-

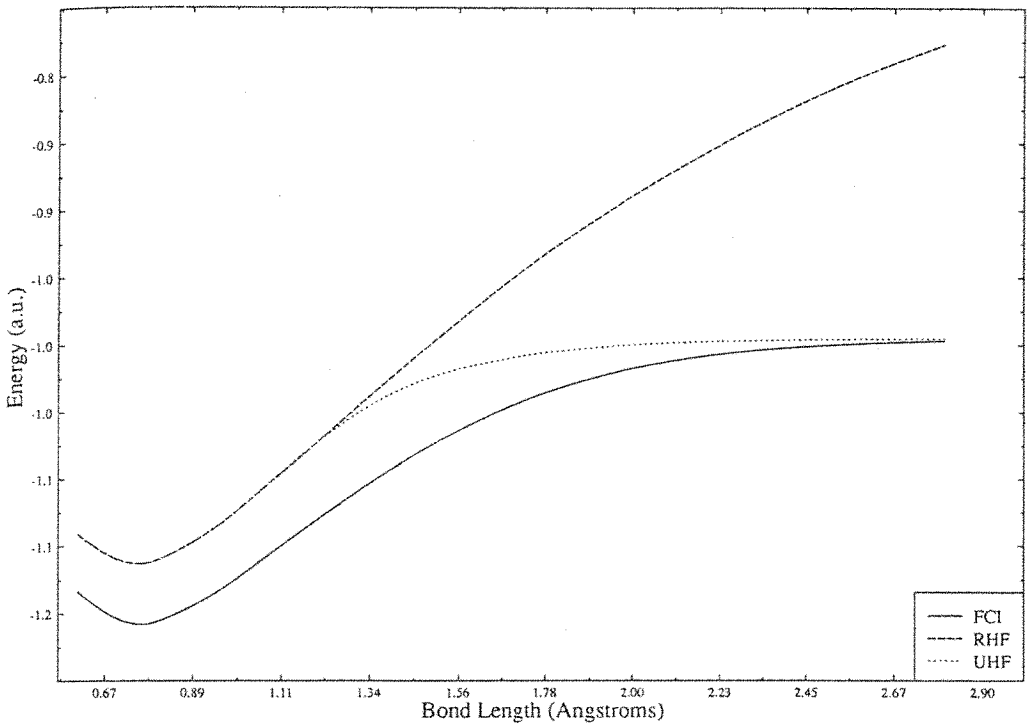


**Figure 2** Potential energy curves for $H_2$.

function is an example of a generalized valence bond (GVB) function. Compare the GVB result with the two SCF results in Figure 2.

A high spin ROHF wavefunction is also an eigenfunction of spin: that is, $\hat{S}^2|\text{ROHF}\rangle = S(S+1)|\text{ROHF}\rangle$. The designation "high spin" is meant to eliminate such situations as an open-shell singlet (illustrated in Figure 3), which cannot be written as a single determinant. We would require two determinants, that is, $(|\phi_1\alpha\phi_1\beta\phi_2\alpha\phi_3\beta| - |\phi_1\alpha\phi_1\beta\phi_2\beta\phi_3\alpha|)$, to be a spin eigenfunction. Similar situations pertain to the low-spin doublets also shown in Figure 3, whereas the "maximum" double-occupancy doublet is an eigenfunction of spin. Like RHF, an ROHF wavefunction usually will not separate correctly on a potential energy surface, as shown for $O_2(^3\Sigma_g^-)$ in Figure 4. (An open-shell example like


Open-shell singlet

Low-spin doublet

**Figure 3** Orbital energy level diagrams for the open-shell singlet and low-spin doublet cases.

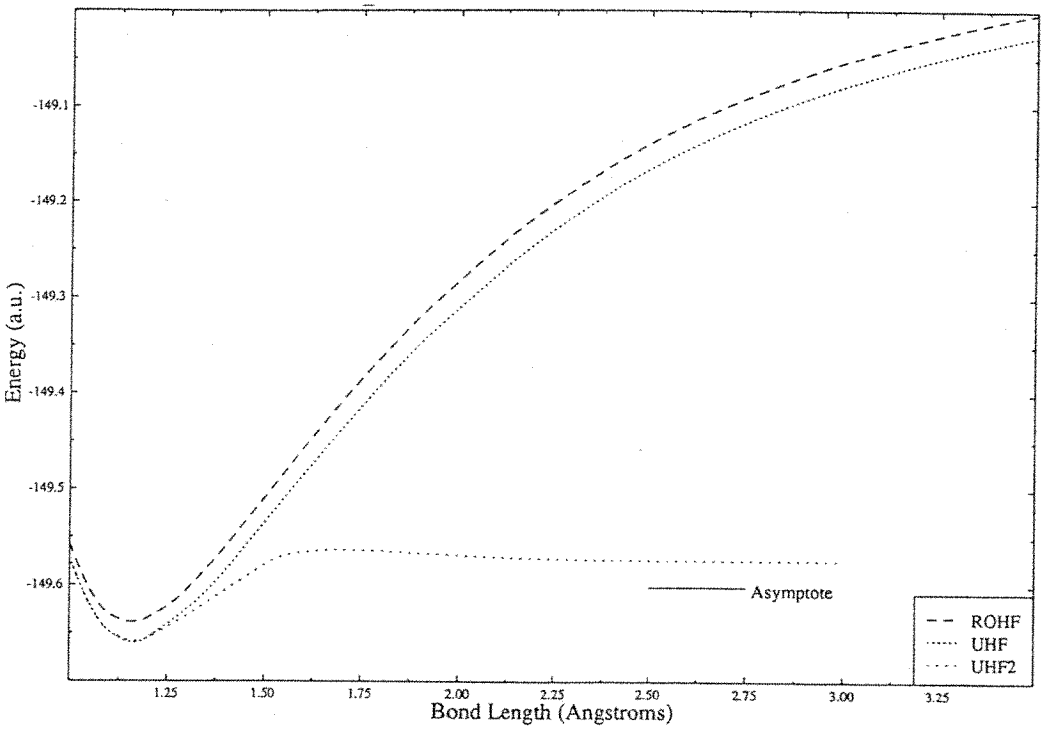


**Figure 4** SCF potential curves for $O_2$.

$AB^+ \rightarrow A^+ + B$, where B is a closed-shell system, will separate correctly in ROHF.)

UHF is a fully unconstrained SCF method and actually represents several different solutions, depending on what symmetries are broken. First, it allows the wavefunctions to be impure spin states. Also, because the various kinds of spatial and spin symmetry are frequently broken in UHF wavefunctions, we can often obtain correct asymptotic separation on a potential energy curve as in Figure 2. For example, the UHF solution for $H_2$ breaks g and u symmetry to give localized $1s_A$ and $1s_B$ functions, but, unlike the GVB solution, the UHF wavefunction does not have the symmetry of the molecule (see Figure 5; it is half of the correct solution!). To construct a proper wavefunction, the UHF solution must be mixed with its phase-reversed partner. The change in energy brought about by this process of symmetrization is related to the singlet–triplet splitting. However, this gap goes to zero in the limit of infinite separation, since the two UHF wavefunctions are degenerate. As a result, UHF and GVB energies are asymptotically degenerate.

The UHF solutions for $O_2$ in Figure 4 illustrate another aspect as the higher energy UHF only breaks spin symmetry, whereas the lower energy solution, UHF2, breaks spatial symmetry, too. However, unlike $H_2$, there is no UHF solution that correctly separates to two UHF $O(^3P)$ atoms for $O_2(^3\Sigma_g^-)$.

In making practical SCF calculations we introduce a basic set $\{\chi_\mu\}$, which usually consists of contracted Gaussian-type orbitals centered on each atom $\gamma$,

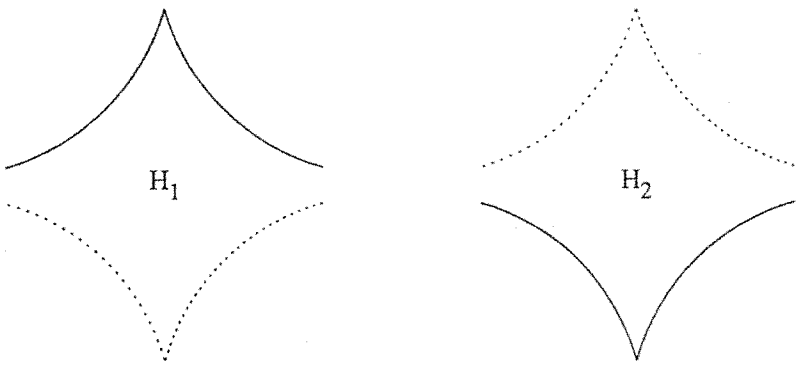


Figure 5 Solid lines, UHF solution for $H_2$ at $R = \infty$; dashed lines, degenerate UHF solution.

that is, $\chi_s(R_\gamma) = \Sigma_k \eta_{s_k} \exp(-\xi_{s_k} r_\gamma^2)$, $\chi_{p_x}(R_\gamma) = \Sigma_k \eta_{x_k} x \exp(-\xi_{x_k} r_\gamma^2)$, etc., where the coefficients $\eta_k$ and $\xi_k$ are fixed typically from prior atomic Hartree–Fock calculations, but other prescriptions are possible. Using UHF as an example, spatial orbitals are associated with each spin via

$$\begin{aligned} \phi_j^\alpha &= \sum_{\mu=1}^{M} \chi_\mu c_{\mu_j}^\alpha = |\tilde{\chi}\rangle \mathbf{c}_j^\alpha \\ \phi_j^\beta &= \sum_{\mu=1}^{M} \chi_\mu c_{\mu_j}^\beta = |\tilde{\chi}\rangle \mathbf{c}_j^\beta \end{aligned} \qquad [16]$$

Inserting Eq. [16] into the canonical SCF equation $\mathcal{F}\varphi_i = \epsilon_i \varphi_i$, leads to the matrix formulation of the UHF equations

$$\begin{aligned} \mathbf{F}^\alpha \mathbf{c}_p^\alpha &= \mathbf{S}\mathbf{c}_p^\alpha \epsilon_p^\alpha \\ \mathbf{F}^\beta \mathbf{c}_p^\beta &= \mathbf{S}\mathbf{c}_p^\beta \epsilon_p^\beta \end{aligned} \qquad [17]$$

where $\mathbf{F} = \langle \tilde{\chi} | \mathcal{F} | \tilde{\chi} \rangle$ and $\mathbf{S} = \langle \tilde{\chi} | \tilde{\chi} \rangle$, and the $\alpha$ or $\beta$ dependence in $\mathbf{F}$ arises from the explicit spin integrated $u(1)$ part of $\mathcal{F}$. For RHF $\mathcal{F}^\alpha = \mathcal{F}^\beta$, of course. For ROHF there are two different forms of $\mathcal{F}$ to consider, one for the $\alpha$-spin open-shell orbitals, and one for the $\alpha$- or $\beta$-spin doubly occupied orbitals.

Because the matrix $\mathbf{F}^\alpha$ or $\mathbf{F}^\beta$ has dimension $M \times M$, we obtain $M$ $\alpha$-spin and $M$ $\beta$-spin spatial orbitals via diagonalization. But we only use $n_\alpha$ and $n_\beta$ of these orbitals to construct the SCF solution. This leaves us with $M - n_\alpha = N_\alpha$ and $M - n_\beta = N_\beta$ spin orbitals unoccupied in the SCF solution. We identify these orbitals collectively by the letters $a, b, c, \ldots$, whereas $i, j, k, \ldots$ indicate orbitals occupied in the SCF solution, and $p, q, r, s, \ldots$ indicate any orbital. The orbitals $a, b, c, \ldots$ are frequently called "virtual" orbitals because they

are obtained as a by-product of diagonalizing the **F** matrix. They are not determined by the variational principle as are the occupied SCF orbitals. Physically, an electron in a virtual orbital feels the average field of all $n$ electrons instead of the $n - 1$ felt by an electron in an occupied orbital. The virtual orbitals are, however, a convenient orthogonal set of one-particle functions that satisfy $\mathcal{F}(1)\varphi_a(1) = \epsilon_a\varphi_a(1)$ for all $a$ (like the occupied orbitals) and, combined with the occupied orbitals, span the space represented by the contracted Gaussian functions. The virtual orbitals play a critical role in correlated methods.

## CORRELATION PROBLEM

An investigation of the potential curves for $N_2$ (Figure 6) and $F_2$ (Figure 7) tells us a great deal about the deficiencies of a model of the Hartree–Fock, independent particle type. The inability of a restricted (RHF or ROHF) solution to correctly separate to open shells causes the curve to rise too quickly, leading to a gross overestimate of the dissociation energy. Hence, RHF usually will underestimate equilibrium bond lengths ($R_e$). For the same reason, the curvature (force constant) at $R_e$ will be larger than it should be, resulting in an overprediction of the vibrational frequency. UHF does not rectify this problem, even though it will often separate correctly, as in the $N_2$ example. The reason is

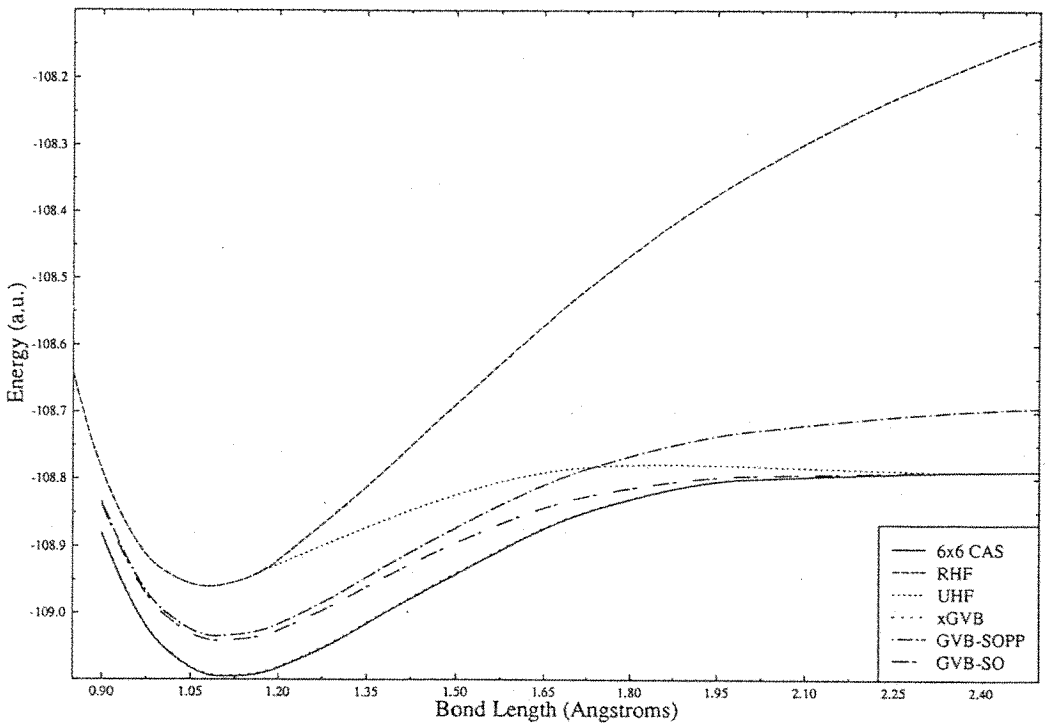


**Figure 6** SCF, GVB, and CASSCF potential curves for $N_2$.

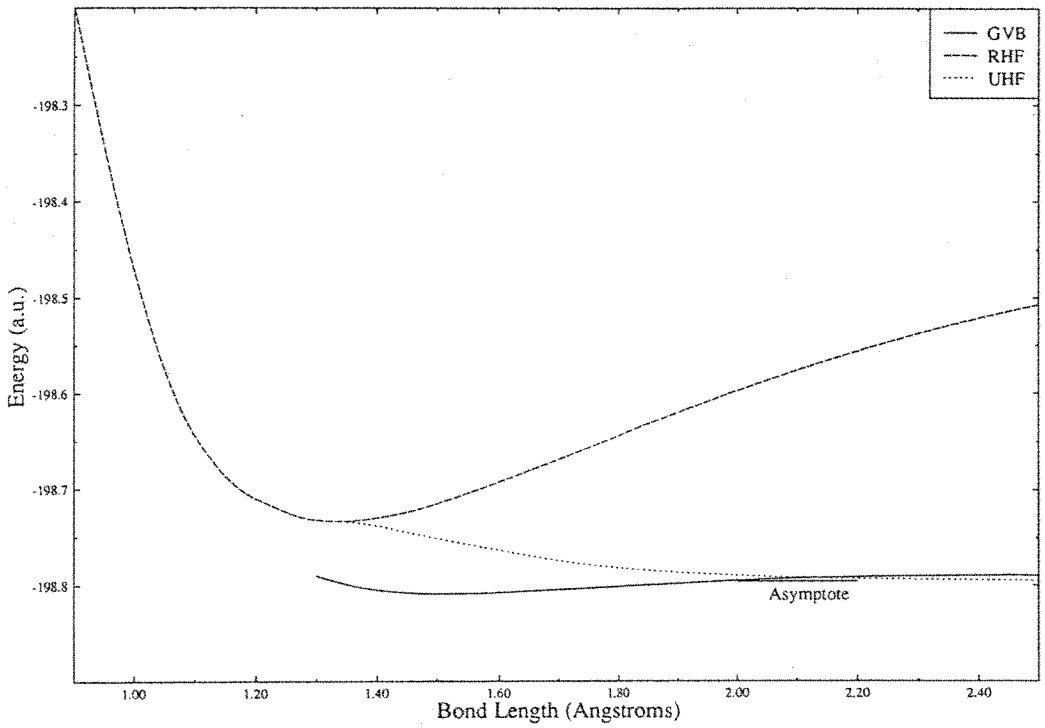


**Figure 7** SCF and GVB potential curves for $F_2$.

that for closed-shell systems, the UHF solution and the RHF are usually the same at $R_e$, so the RHF deficiencies apply. (For some closed-shell examples a UHF solution which is necessarily below the RHF solution can be obtained, but that will not change our conclusion.) Furthermore, even when the UHF differs from the RHF, its lack of spin symmetry can affect the shape of the potential curve. For closed-shell cases in particular, where UHF is being used to effect correct separation, all triplet, quintet, and higher odd-spin states, which have higher energies, contaminate the UHF solution, causing the curve to rise too steeply slightly beyond equilibrium. (Because the UHF solution is not an eigenfunction of spin, we can resolve the UHF solution into its pure spin eigenstates, via $|\text{UHF}\rangle = \Sigma_{s_k} |s_k\rangle c_{s_k}$ and $c_{s_k} = \langle s_k|\text{UHF}\rangle$. If $S_z = 0$, as for a closed shell, then only triplet, quintet, etc., can have $S_z = 0$ components.) Because the UHF solution is separating correctly in this case, the curve can even show a slight artificial barrier to separation for some examples (see Figure 6). On the other hand, at the UHF separated atom limit, the several spin states all become degenerate. Hence, even though the spin multiplicity is erroneous, the energy is correct at the asymptote, as in the $H_2$ example (Figure 2).

Another failing of UHF is less frequent but pertinent for some examples. This is demonstrated by the $F_2$ UHF curve (Figure 7), which is not bound. That is, two F atoms are lower in energy than is the $F_2$ molecule, causing UHF not to predict the existence of the $F_2$ molecule!

To correct the foregoing deficiencies of a Hartree-Fock reference, we have to introduce "electron correlation." What is electron correlation? As discussed above, in the SCF method we attempt to describe the contribution of the two-electron part of the Hamiltonian by one electron moving in an average field of the $n - 1$ other electrons. Overall, this is not a bad approximation, inasmuch as the SCF energy is usually more than 99% of the total (nonrelativistic) electronic energy of an atom or molecule. However, chemists are interested in much smaller energy differences, such as bond dissociation energies, and we see from Figures 4, 6, and 7 that these are poorly described by SCF methods. Other quantities of interest (ionization potentials, electron affinities, electronic excitation energies, vibrational energies, etc.) are sensitive in varying degrees to the "average" electron interaction approximation of SCF theory. An average field neglects the fact that electrons, as charged particles, have instantaneous Coulombic interactions that keep them apart. That is, the motion of electrons is "correlated." The movement of one affects the movement of all others, so we have a complicated many-body system to describe. Because the instantaneous interactions serve to keep electrons apart better than an average interaction, the interelectronic repulsion is reduced, causing the energy for the molecule to be lower than the SCF value. The energy difference is called the *correlation energy:*

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{SCF}} \qquad [18]$$

where $E_{\text{SCF}}$ can be any of our independent particle models, RHF, UHF, or ROHF. [Some authors use $E_{\text{RHF}}$ (or $E_{\text{ROHF}}$) as the reference, so that $E_{\text{UHF}}$ already includes some degree of electron correlation by relaxing the double-occupancy restriction on the spatial orbitals. For our purposes, "correlation" will be the correction for any SCF reference.]

To improve on a single determinant reference we must develop a superior treatment of the two-electron interaction. Logically, this would involve explicit use of the two-particle operator $r_{ij}$ in a trial wavefunction, and such Hylleraas-type trial wavefunctions have been used to obtain the best results (for, e.g., He, Li, Be, $H_2$, and LiH).[15] Analysis of certain other analytic properties of the correlation cusp $1/(|\mathbf{r}_i - \mathbf{r}_j|)$[16] have also been exploited to develop better descriptions without having to use wavefunctions that are explicitly dependent on $r_{ij}$, but such methods also have many computational restrictions.

Electron correlation can be introduced solely with one-particle functions if we construct a wavefunction with the flexibility to allow electrons to stay away from each other. That means we need functions in regions of space different from just that sampled by the SCF calculation. As the SCF wavefunction spans only the occupied space, we can introduce different regions of space by admitting the "virtual" orbitals that are unoccupied in the SCF into the wavefunction. We do this by constructing a configuration interaction (CI) wavefunction

$$\Psi_{CI} = \Phi_0 + \sum_{i,a} C_i^a \Phi_i^a + \sum_{\substack{i<j \\ a<b}} C_{ij}^{ab} \Phi_{ij}^{ab} + \cdots \qquad [19]$$

which in addition to the SCF solution introduces single, double, and higher excitations. $\Phi_0$ is called the *reference state* wavefunction. The single excitation $\Phi_i^a = \mathcal{A}(\varphi_1(1) \cdots \varphi_a(i) \cdots \varphi_n(n))$ corresponds to replacing the occupied SCF orbital $\varphi_i$ by the unoccupied orbital, $\varphi_a$. $\Phi_{ij}^{ab}$ represents the double excitation, $\mathcal{A}(\varphi_1(1) \cdots \varphi_a(i) \cdots \varphi_b(j) \cdots \varphi_n(n))$, and so forth through triple $\Phi_{ijk}^{abc}$, quadruple $\Phi_{ijkl}^{abcd}$, . . . all the way to $n$-tuple excitations $\Phi_{ijk\ldots n}^{abc\ldots z}$ for $n$ electrons. The CI coefficients ($C_i^a$, $C_{ij}^{ab}$, etc.) are then optimized to give the lowest variational energy,

$$E_{CI} = \frac{\langle \Psi_{CI} | \mathcal{H} | \Psi_{CI} \rangle}{\langle \Psi_{CI} | \Psi_{CI} \rangle} \qquad [20]$$

Because the CI wavefunction is a much more flexible wavefunction than is $\Phi_0$, $E_{CI} < E_{SCF}$ and we lower the energy in line with the effect of correlating electron motions. [If we add only single excitations, there can be no improvement of the RHF or UHF energy, since singles do not directly mix with these solutions (Brillouin's theorem). Double excitations have to be included to introduce electron correlation.]

The "full CI" is defined as the wavefunction that includes all possible excitations through $n$-fold for $n$ electrons. If the $\{\chi_\mu\}$ basis used to represent the MOs were complete (which usually means infinite), then the full CI would be the "complete" CI, which is the exact solution to the nonrelativistic Schrödinger equation. In a practical case $\{\chi_\mu\}$ is finite, so the full CI is the best that can be done in the basis set. Different correlated methods are recommended according to how closely they approach full CI results.

In the full CI wavefunction we have three (not mutually exclusive) categories of correlation corrections relative to $\Phi_0$ that should be distinguished:

1. Excitations whose individual contributions are small, but because of an excessive number, their cumulative contributions are quite significant.
2. Excitations required to provide a correct zeroth-order description as dictated by spin-symmetry considerations.
3. Excitations whose coefficients will be comparatively large, as required to enable a molecule to separate correctly into fragments, in some excited states, and for certain other cases.

Category 1 is called *dynamic correlation* because its function is to enable the electrons to stay apart. It usually represents the largest part of the total correlation energy.

Category 2 occurs for the open-shell singlet example discussed earlier. A single determinant is not an eigenfunction of spin, so such states should always

be preferentially described using a two-determinant reference. We will refer to this as symmetry required, or *static correlation*. When these states are described by a single UHF determinant $|\phi_1\alpha(1)\phi_2\beta(2)|$ the wavefunction corresponds to a 50:50 mixture of the $|\phi_1\alpha(1)\phi_2\beta(2)| \pm |\phi_1\beta(1)\phi_2\alpha(2)|$ singlet (minus) and triplet (plus) state. Because the latter state can also be described by the high spin single determinant $|\phi_1\alpha(1)\phi_2\alpha(2)|$, the triplet energy can be obtained, and, aided by that information, approximate methods can be used to subtract the triplet contamination from the energy for the single UHF determinant,[17] but this is far less desirable than a treatment using the two-determinant reference.

Category 3 is often called *nondynamic correlation* and is perhaps best described by considering the simplest possible correlated wavefunction for $F_2$ and $N_2$ that can correctly dissociate. In the $F_2$ case (Fig. 7) this is a GVB wavefunction,[10] which in the orthogonal orbital form for $F_2$ consists of the two-determinant wavefunction $C_1|\text{core } 2p\sigma_g\alpha 2p\sigma_g\beta| + C_2|\text{core } 2p\sigma_u\alpha 2p\sigma_u\beta|$.

As discussed above for $H_2$, a $\sigma_u$ must be mixed with a $\sigma_g$ function to allow the electrons to localize on the two separated atoms in a model that, unlike UHF, preserves spin and spatial symmetry. Furthermore, the GVB result requires that the form of the orbitals $2p\sigma_g$ and $2p\sigma_u$ be determined together with the configuration coefficients $C_1$ and $C_2$. GVB is the simplest example of a multiconfiguration (MCSCF) calculation.[18,19] The GVB result for $F_2$, shown in Figure 7 is qualitatively correct as a function of internuclear distance $R$.

Similarly, in Figure 6 we show results for $N_2$ with analogous GVB wavefunctions for the triple bond,[20] which requires a product of wavefunctions consisting of two orbitals for each of the three bonds. This means we mix the $2p\sigma_g$ and $2p\sigma_u$ orbitals and also mix the $2p\pi_{xu}$, $2p\pi_{xg}$, $2p\pi_{yu}$, and $2p\pi_{yg}$. This results in a set of determinants starting with $\Phi_0 = |\text{core } 2p\sigma_g\alpha 2p\sigma_g\beta 2\pi_{xu}\alpha\pi_{xu}\beta 2\pi_{yu}\alpha 2\pi_{yu}\beta|$ through all other symmetry-allowed combinations including the hextuple excitation $|\text{core } 2p\sigma_u\alpha 2p\sigma_u\beta 2\pi_{xg}\alpha 2\pi_{xg}\beta 2\pi_{yg}\alpha 2\pi_{yg}\beta|$. If we allow the three two-determinant products to be singlet coupled (GVB-SO-PP) (8 configurations) only, we do not get the correct asymptote. Allowing singlet and triplet coupling among the three bonds, GVB-SO (14 configurations) and $x$GVB (20 configurations), both of which arise from singlet and triplet coupled combinations including the open-shell singlets, correctly separate $N_2$ into two $^4S$ N atoms as shown in Figure 6. The complete active space SCF (CASSCF)[14] takes this one step further. It recognizes that $N_2$ requires six orbitals, $3\sigma_g$, $3\sigma_u$, the pair $1\pi_u$, and the pair $1\pi_g$, and distributes the six electrons among them in all possible ways. This gives the $x$GVB set plus others (32 configurations) and clearly effects proper separation to two N($^4$S) atoms. In Fig. 6, the CASSCF is nearly indistinguishable from the $x$GVB result.

This nondynamic correlation does not pertain to keeping electrons apart, which is a dynamic effect, but instead emphasizes having the appropriate sets of excitations in the wavefunction that requires a "large coefficient" to permit a more proper zeroth-order description. ["Large" is not well defined. In practice a coefficient for a determinant greater than 0.2 in an intermediately normalized wavefunction (see later) is frequently cause for concern, while for well-defined

dynamical correlation effects all determinants have coefficients $< 0.1$.] Other examples include $O_3$, where the competing resonance structures require at least two determinants in its zeroth-order solution (see section entitled Electronic Spectra).

These different categories of correlation effects are not mutually exclusive because the full CI is the exact result in a basis for any situation. Hence, even when a single determinant SCF model is a poor approximation to the right answer (e.g., RHF for $N_2$ at a large internuclear separation), the full CI solution would properly weight the other important configurations to give the right result. Similarly, even for the open-shell singlet example we could start with just the $|\phi_1\alpha\phi_2\beta|$ determinant, and since the double excitation $|\phi_1\beta\phi_2\alpha|$ is in the full CI solution, we will obtain both terms with the same coefficient but opposite sign in the full CI for the singlet solution.

Thus, we have two operational choices:

1. We can take the dynamic correlation route and always start with a single determinant reference, even when it is a poor approximation, and introduce higher categories of excitations until we believe our wavefunction is flexible enough to obtain satisfactory results.
2. We can insist on first introducing the relevant static or nondynamic correlation to have a qualitatively correct zeroth-order solution and then augment it by dynamic correlation to obtain quantitative results.

The preferable route will depend on the problem, the ease of implementation, and the computational efficiency.

The single reference (dynamic correlation) approaches will lead us to CI approximations such as all single and double excitations from a single determinant (CISD);[9] or many-body perturbation theory (MBPT)[5] [also known as Møller–Plesset (MP) perturbation theory[8]]; or coupled-cluster (CC) methods[6] for single and double excitations like CCSD[21] and its approximation termed quadratic CI (i.e., QCISD);[13] or coupled-cluster single, double, and triple excitations (CCSDT).[22] The second operational choice above will introduce multireference (MR) methods like MCSCF,[18,19] GVB,[10] or CASSCF[14] [the latter is also known as full-optimized reaction space (FORS)[23]]. Each of these references may then be augmented by additional dynamic correlation via CI, as when single and double excitations from a selection of reference determinants are included to define a MR-CISD wavefunction.[9] More specific designations are GVB-CISD and CASSCF-CISD and so on. Furthermore, there are similar MR-MBPT[24-26] and MR-CC methods.[29-32]

Whereas CI methods are variational, giving upper bounds to the total energy, the other discriminating factor in choice of correlated method is whether a method scales correctly with size, or, equivalently, the number of electrons. This (size) extensive property,[33] which can be quite significant numerically, is satisfied by many-body methods (e.g., MBPT, CC, and QCI methods) but not by CI methods other than full CI. "Truncated" CI ap-

proaches are often subject to semiempirical extensivity corrections[34] in an attempt to alleviate this failing. We will further discuss the consequences of inextensivity in the next section. MR-CC has the attraction of being extensive while including nondynamic correlation. One particularly useful development in MR-CC is the recently developed two-determinant reference open-shell singlet (TD-CCSD) method[31]. Another is the Fock space MR-CC[27-29] for excited, ionized, and electron-attached states, both in ACES II.[29,30]

## METHODS FOR ELECTRON CORRELATION

### Methods

Given a set of $n = n_\alpha + n_\beta$ molecular spin orbitals $i, j, k, \ldots$ occupied in $\Phi_0$, and a set of $N = N_\alpha + N_\beta$ molecular spin orbitals $a, b, c, \ldots$ that are unoccupied in $\Phi_0$, to allow for electron correlation our objective is to introduce the excited molecular orbitals $a, b, c, \ldots$ that do not appear in $\Phi_0$ to make an improved wavefunction

$$\Psi = \Phi_0 + \sum_{i,a} C_i^a \Phi_i^a + \sum_{\substack{i<j \\ a<b}} C_{ij}^{ab} \Phi_{ij}^{ab} + \sum_{\substack{i<j<k \\ a<b<c}} C_{ijk}^{abc} \Phi_{ijk}^{abc} + \cdots$$

$$= (1 + \hat{C}_1 + \hat{C}_2 + \hat{C}_3 + \cdots)\Phi_0 \qquad [21]$$

In the limit of all possible $n$ excitations (since $m_s$ is a good quantum number—we can only excite to orbitals with the same spin), we have the full CI (FCI) solution. Notice that this CI wavefunction applies equally well for open and closed shells and, for the latter, permits either ROHF, UHF, or other orbitals. The full CI solution has several attractive properties:

1. It gives the best solution for the energy and density in the contracted (AO) basis set for any states.
2. It is "extensive" (i.e., it scales properly with molecular size and the number of electrons).
3. It is invariant to any change in the form of the MOs, so any orbitals are satisfactory.
4. It provides an upper bound to the exact electronic energy for any state.

The full CI is impractical except as a reference because the number of determinants is asymptotically $\sim N^n$. That means its application is possible only for small molecules and small basis sets. In practice, calculations using $\sim 25 \times 10^6$ determinants for $H_2O$ in a double zeta plus polarization (DZP) basis have been made[35] and $\sim 10^8$ determinants for $CH_3$ in a triple zeta plus polarization (TZP) basis.[36]

The three most popular methods for introducing electron correlation are (truncated) CI such as CISD, MBPT, and CC theory. Other correlated methods (e.g., propagator methods,[37,38] which are related to the last two) have also been widely used for excitation and ionization energies. All can be viewed as different ways of building in some of the excitations that are present in the FCI. The comparative merits of the methods depend on how rapidly they converge to the FCI solutions for energies, densities, and other quantities of interest.

***Configuration Interaction***

CI methods are straightforward. Once truncated to some subset of excitations, the coefficients, $C_i^a, C_{ij}^{ab}, \ldots$ are determined from the variational principle $E_{CI} = \langle \Psi_{CI}|\mathcal{H}|\Psi_{CI}\rangle / \langle \Psi_{CI}|\Psi_{CI}\rangle \geq E_{exact}$ to give the lowest energy for the ground state (or lowest state of a given symmetry). Introducing only double excitations defines CID, whose wavefunction is $\Psi_{CID} = (1 + \hat{C}_2)\Phi_0$. The variational condition is equivalent to the matrix equation

$$\begin{bmatrix} \mathbf{H}_{OO} & \mathbf{H}_{OD} \\ \mathbf{H}_{DO} & \mathbf{H}_{DD} \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{C}_D \end{bmatrix} = \begin{bmatrix} 1 \\ \mathbf{C}_D \end{bmatrix} E_{CID} \qquad [22]$$

$$E_{CID} = \langle \Phi_0 | H | \Psi_{CID} \} \qquad [23]$$

where **D** in

$$\begin{aligned} \mathbf{H}_{OD} &= \langle \Phi_0 | H | \mathbf{D} \rangle \\ \mathbf{H}_{DD} &= \langle \mathbf{D} | H | \mathbf{D} \rangle \end{aligned} \qquad [24]$$

indicates all double excitations, $\Phi_{ij}^{ab}$. Furthermore, in CI, each member of a subset of excitations will improve the energy somewhat, so adding singles, $E(\text{CISD}) < E(\text{CID})$. Similarly, if the most important few triples are added (call it CISDT′) then $E(\text{CISDT}')$ might be a good approximation to the full $E(\text{CISDT})$. Truncated CI methods also provide upper bounds to all electronic states. However, in chemistry we are interested in energy differences (dissociation energies, excitation energies, activation barriers, etc.), and there is seldom a bound on them. Furthermore, unlike MBPT and CC methods (see below), CI methods are not "extensive," and this can cause numerically significant errors.

Because the full CI (FCI) is the ultimate result in a given basis set, we can partly assess the comparative quality of different correlated methods by comparing with the limited number of full CI results available for molecules.[35,39,40] These are available for BH, FH, and $H_2O$ at the equilibrium bond length $R_e$, 1.5 $R_e$, and 2.0 $R_e$ in a DZP basis (see later). As discussed earlier, the RHF reference function separates incorrectly when going to open-shell fragments, so as bonds are stretched, it offers a particularly poor approximation to the correct answer. Because the FCI is the best possible answer, this behavior has no effect on those results, but for a truncated CI, MBPT, or CC calculation, a poor

**Table 1** Differences (millihartrees[a]) Between FCI Energies and Various Truncations at Three Internuclear Separations (DZP Basis)

| | BH | | | FH | | | $H_2O$ | | | Mean absolute error |
|---|---|---|---|---|---|---|---|---|---|---|
| CI method | $R_e$ | 1.5 $R_e$ | 2.0 $R_e$ | $R_e$ | 1.5 $R_e$ | 2.0 $R_e$ | $R_e$ | 1.5 $R_e$ | 2.0 $R_e$ | |
| CID | 6.02 | 9.63 | 20.8 | 10.3 | 18.6 | 35.5 | 13.7 | 34.5 | 84.8 | 25.9 |
| CISD | 5.21 | 7.51 | 14.5 | 9.38 | 14.9 | 27.6 | 12.9 | 30.4 | 75.6 | 22.0 |
| CISDT | 3.60 | 5.19 | 10.1 | 7.01 | 11.1 | 19.2 | 10.6 | 23.5 | 60.3 | 16.7 |
| CISDTQ | 0.03 | 0.06 | 0.16 | 0.28 | 0.49 | 0.92 | 0.40 | 1.55 | 6.29 | 1.13 |

[a]One millihartree is 0.6275 kcal/mol.

reference places extreme demands on the method. The ability of a correlated method to overcome such a poor reference is a testament to its accuracy and applicability. The errors in the energy relative to FCI for various truncated CI methods are shown in Table 1.

The correlation energy error in these examples ranges from 102 to 370 millihartrees, with an average of 213. CID accounts for about 88% of the correlation.

We can make a few observations. The variational condition guarantees that the error is above the exact result. Because of incorrect separation of the RHF reference, the errors are greatest at twice the equilibrium $R_e$. Comparing the different CI truncations, we see that on average single excitations account for ~4, triples for ~5, and quadruples for a whopping ~15 millihartrees. The vast improvement of CISDTQ compared to CISDT emphasizes the numerical importance of quadruple excitations in CI. Very accurate methods should include these effects, and MBPT and CC methods do so in a convenient way.

### *Many-Body Perturbation Theory*

The next method we consider is many-body perturbation theory (MBPT).[41,42] This approach exploits the form of the Hamiltonian in Eq. [7]

$$\mathcal{H} = H_0 + V \qquad [25]$$

$$H_0 = \sum_i \mathcal{F}(i) \qquad [26a]$$

$$V = \frac{1}{2} \sum_{i,j} \frac{1}{r_{ij}} - \sum_i u(i) \qquad [26b]$$

so that we attempt to solve the correlation problem by introducing the perturbation, $V$. In other words, the exact (FCI) wavefunction can be written as a perturbation expansion

$$\Psi = \Phi_0 + \psi^{(1)} + \psi^{(2)} + \cdots \qquad [27]$$

instead of the form in Eq. [21]. Similarly, we have the eigenvalue $E = E_0 + E^{(1)} + E^{(2)} + \cdots$. Inserting into the Schrödinger equation, with $\Phi_0 = \psi_0$, we obtain

$$\begin{aligned}(H_0 + V)(\psi_0 + \psi^{(1)} + \psi^{(2)} + \cdots) \\ = (E_0 + E^{(1)} + E^{(2)} + \cdots)(\psi_0 + \psi^{(1)}) \cdots\end{aligned} \qquad [28]$$

All the terms in such an expansion are linearly independent, so this equation must be satisfied for each order $m$. Therefore,

$$0th: (E_0 - H_0)\psi_0 = 0 \qquad [29a]$$

$$1st: (E_0 - H_0)\psi^{(1)} = (V - E^{(1)})\psi_0 \qquad [29b]$$

$$2nd: (E_0 - H_0)\psi^{(2)} = (V - E^{(1)})\psi^{(1)} - E^{(2)}\psi_0 \qquad [29c]$$

$$3rd: (E_0 - H_0)\psi^{(3)} = (V - E^{(1)})\psi^{(2)} - E^{(2)}\psi^{(1)} - E^{(3)}\psi_0 \qquad [29d]$$

$$\vdots$$

$$mth: (E_0 - H_0)\psi^{(m)} = (V - E^{(1)})\psi^{(m-1)} - \sum_{k=2}^{m} E^{(k)}\psi^{(m-k)} \qquad [29e]$$

By multiplying on the left by $\psi_0$ and using the zeroth-order equation and ($\langle\psi_0|\psi_0\rangle = 1$, $\langle\psi_0|\psi^{(m)}\rangle = 0$), we find that

$$\begin{aligned} E^{(1)} &= \langle\psi_0|V|\psi_0\rangle \\ E^{(2)} &= \langle\psi_0|V|\psi^{(1)}\rangle \\ &\vdots \\ E^{(m+1)} &= \langle\psi_0|V|\psi^{(m)}\rangle \end{aligned} \qquad [30]$$

In the energy expression we are using intermediate normalization. Remember that all wavefunctions are determined only up to an arbitrary constant, which is frequently chosen to make $\langle\Psi|\Psi\rangle = 1$ to provide a convenient probability interpretation. In perturbation theory (and in CI as we are doing it here), it is more convenient to choose the intermediate normalization $\langle\psi_0|\Psi\rangle = 1$, which is equivalent to $\langle\psi_0|\psi^{(k)}\rangle = 0$ for all $k > 0$. The relationship between a fully normalized wavefunction, $\langle\Psi|\Psi\rangle = 1$, and an intermediately normalized one is straightforward. If $\Psi = \Sigma_k C_k\Phi_k$ and $\langle\Psi|\Psi\rangle = 1$, for $\{\Phi_k\}$ a set of functions orthogonal to $\Phi_0$, then $\Psi' = \Phi_0 + \Sigma_{k\neq 0}(C_k/C_0)\Phi_k$ is intermediately normalized, since $\langle\Phi_0|\Psi'\rangle = 1$. This normalization has the effect of simplifying the energy expression to that in Eq. [30].

Consistent with the intermediate normalization condition, we want all our perturbed $n$-particle functions $\Psi^{(m)}$ to be orthogonal to $\Psi_0$. Just as in the case of CI, a natural set to represent these functions is the set of single, double, etc., excitations.

To obtain $\psi^{(1)}$ consider its expansion in double excitations,

$$\psi^{(1)} = \sum_{\substack{i>j \\ a>b}} C_{ij}^{ab(1)}\,\Phi_{ij}^{ab} \qquad [31]$$

where the $^{(1)}$ on the coefficient emphasizes its first-order character. Because every excitation like $\Phi_{ij}^{ab}$ is an eigenfunction of $H_0$ (i.e., $H_0\Phi_{ij}^{ab} = E_{ij}^{ab}\Phi_{ij}^{ab}$),

$$E_{ij}^{ab} = \sum_{k\neq i,j}^{n} \epsilon_k + \epsilon_a + \epsilon_b$$

$$= E_0 - \epsilon_i - \epsilon_j + \epsilon_a + \epsilon_b = E_0 - \varepsilon_{ij}^{ab}$$

it follows that

$$(E_0 - H_0)\Phi_{ij}^{ab} = \varepsilon_{ij}^{ab}\Phi_{ij}^{ab} \qquad [32]$$

Consequently, inserting $\psi^{(1)}$ from Eq. [31] into Eq. [29b] and left-multiplying by another double excitation $\Phi_{kl}^{cd}$ and integrating, via orthonormality of the excitations, we obtain

$$\varepsilon_{kl}^{cd} C_{kl}^{cd(1)} = \langle \Phi_{kl}^{cd} | V | \Phi_0 \rangle \qquad [33]$$

or

$$\psi^{(1)} = \sum_{\substack{i>j \\ a>b}} \frac{\langle \Phi_{ij}^{ab} | V | \Phi_0 \rangle}{\varepsilon_{ij}^{ab}} \Phi_{ij}^{ab} \qquad [34]$$

and

$$E^{(2)} = \sum_{\substack{i>j \\ a>b}} |\langle \Phi_{ij}^{ab} | V | \Phi_0 \rangle|^2 / \varepsilon_{ij}^{ab} = \sum_{\substack{i>j \\ a>b}} |\langle ij \| ab \rangle|^2 / \varepsilon_{ij}^{ab} \qquad [35]$$

This second-order correction defines MBPT(2), and it is the simplest correlation correction. It should be clear that triple and higher excitations cannot contribute to $\psi^{(1)}$ because V includes only a two-particle operator. Following the same procedure as above for single excitation contributions, $\Phi_i^a$, the results depend on $\langle \Phi_0 | V | \Phi_i^a \rangle = \langle \Phi_0 | H | \Phi_i^a \rangle = \langle i | \mathcal{F} | a \rangle = \mathcal{F}_{ia} = 0$, which is why no single excitations can mix with an RHF or UHF $\Phi_0$. For an ROHF reference MBPT, however, $\mathcal{F}_{ia} \neq 0$. This introduces an additional term in $E^{(2)}$ and various additional terms in higher order. These have to be included in ROHF-MBPT applications, but that has been accomplished in our ROHF-MBPT method.

Remembering that the average correlation energy correction for the full CI examples is 213 millihartrees MBPT(2) typically accounts for 85–95% of the correlation energy obtainable in a given basis set, as seen in Table 2. Because MBPT(2) depends only on the relatively small subset of two occupied

Table 2 Differences (millihartrees) Between FCI and MBPT Energies (DZP Basis)

| Perturbation theory | BH | | | FH | | | $H_2O$ | | | Mean absolute error |
|---|---|---|---|---|---|---|---|---|---|---|
| | $R_e$ | 1.5 $R_e$ | 2.0 $R_e$ | $R_e$ | 1.5 $R_e$ | 2.0 $R_e$ | $R_e$ | 1.5 $R_e$ | 2.0 $R_e$ | |
| MBPT(2) | 28.6 | 36.1 | 52.8 | 7.80 | 10.6 | 24.0 | 13.0 | 23.3 | 53.7 | 27.8 |
| MBPT(3) | 11.1 | 15.7 | 27.1 | 5.44 | 11.9 | 27.0 | 7.22 | 26.4 | 74.6 | 22.9 |
| SDQ-MBPT(4) | 5.69 | 8.20 | 15.0 | 2.75 | 5.39 | 12.5 | 4.40 | 13.3 | 34.2 | 11.3 |
| MBPT(4) | 5.06 | 7.23 | 13.3 | −0.26 | 0.77 | 4.84 | 0.92 | 5.76 | 14.9 | 5.89 |
| MBPT(5)[a] | 2.52 | 3.60 | 6.07 | 0.81 | 2.29 | 8.10 | 0.70 | 4.98 | 17.0 | 5.12 |

[a] Reference 44.

and two virtual orbital two-electron integrals $\langle ij\|ab\rangle$ (most integrals would be four-virtual $\langle ab\|cd\rangle$ type) and the SCF orbital energies, it is computationally easy to evaluate $E^{(2)}$. $E^{(2)}$ itself is an $\sim n^2N^2$ procedure but the transformed integral $\langle ij\|ab\rangle$ requires an $\sim Mn^2N^2$ step, where $M = n + N$. Because MBPT(2) offers such a large percentage of the correlation correction, it typically provides much better results than SCF alone. We will see that in many applications.

Higher orders of perturbation theory are not conceptually too different from $E^{(2)}$. To get third order we have to consider $\psi^{(2)}$, but for RHF and UHF cases, it depends only on double excitations. We would have to evaluate terms that depend on integrals like $\langle ab\|cd\rangle$, $\langle ia\|jb\rangle$, and $\langle ij\|kl\rangle$ in addition to the $\langle ij\|ab\rangle$ used in $E^{(2)}$. For example, the particle–particle ladder (PPL) part of $E^{(3)}$ is

$$E^{(3)}_{PPL} = \sum_{\substack{a>b\\c>d\\i>j}} \langle ij\|ab\rangle\langle ab\|cd\rangle\langle cd\|ij\rangle / \varepsilon^{ab}_{ij}\,\varepsilon^{cd}_{kl}$$

This term causes MBPT(3) to require an $\sim n^2N^4$ algorithm, like CID (or CISD). In fact, the CID coefficients are $C^{ab}_{ij} \approx C^{ab(1)}_{ij} + C^{ab(2)}_{ij}$ through second order; so we would expect CID and MBPT(3) to usually give similar results. From Table 2 we see that the average absolute error of MBPT(3) is 22.9 millihartrees, which is close to the CID error of 25.9.

Some important distinctions emerge in fourth order. First, even for RHF or UHF $E^{(4)}$ has contributions from single, triple, and quadruple excitations in addition to double excitations.

$$E^{(4)} = \langle\Phi_0|V|\psi^{(3)}\rangle = E^{(4)}_{\mathrm{D}} + E^{(4)}_{\mathrm{S}} + E^{(4)}_{\mathrm{T}} + E^{(4)}_{\mathrm{Q}} - E^{(2)}\langle\psi^{(1)}|\psi^{(1)}\rangle$$

Now we see the consequences of (size)-extensivity for the first time. The last term, which depends on $E^{(2)}$, arises from the second term on the right in Eq. [29c] after multiplying on the left by $\langle\psi^{(1)}|$ and using the fourth-order energy formula. This term plays a critical role in distinguishing MBPT from CI. If we continued approximating the CID coefficients by higher order perturbation theory, we would have

$$C^{ab}_{ij} \approx C^{ab(1)}_{ij} + C^{ab(2)}_{ij} + C^{ab(3)}_{ij} - E^{(2)}C^{ab(1)}_{ij}$$

because $E^{(2)}$ arises only from double excitations.

To understand what is wrong with this, it is convenient to consider a noninteracting system of $m\mathrm{H}_2$ molecules. For such a system, the exact wavefunction for $m\mathrm{H}_2$ would be $\Psi(m\mathrm{H}_2) = [\psi(\mathrm{H}_2)]^m$. Furthermore, the exact en-

ergy $E(mH_2) = mE(H_2)$. Since the exact energy must be written in terms of perturbation corrections, it follows that $E_{SCF}(mH_2) = mE_{SCF}(H_2)$, $E^{(2)}(mH_2) = mE^{(2)}(H_2)$, $E^{(3)}(mH_2) = mE^{(3)}(H_2)$, etc. But notice, the negative (renormalization) term depends on $E^{(2)}$ and $\langle\psi^{(1)}|\psi^{(1)}\rangle$ and since $E^{(2)}(mH_2) = mE^{(2)}(H_2)$ and since $\psi^{(1)}(mH_2) = [\psi^{(1)}(H_2)]^m$, $\langle\psi^{(1)}(mH_2)|\psi^{(1)}(mH_2)\rangle = m\langle\psi^{(1)}(H_2)|\psi^{(1)}(H_2)\rangle$, the quantity $E^{(2)}\langle\psi^{(1)}|\psi^{(1)}\rangle$ depends on $m^2$! How can this be, because we must have $E^{(4)}(mH_2) = mE^{(4)}(H_2)$? This can happen only if somehow the $m^2$-dependent part of the $E^{(2)}\langle\psi^{(1)}|\psi^{(1)}\rangle$ term cancels another part of $E^{(4)}$ to eliminate the $m^2$ dependence. This is exactly what happens. Part of $E_Q^{(4)}$ cancels $E^{(2)}\langle\psi^{(1)}|\psi^{(1)}\rangle$. (There are actually two parts to $E^{(2)}\langle\psi^{(1)}|\psi^{(1)}\rangle$; an $m^2$-dependent "disjoint part" and an $m$-dependent "conjoint or EPV part." The former is canceled while the latter contributes to $E_Q^{(4)}$.) However, because $E^{(2)}$ and $\langle\psi^{(1)}|\psi^{(1)}\rangle$ depend only on double excitations, the cancellation can occur only when quadruple excitations are allowed to be in the wavefunction! Hence, CID is not extensive[5] because it does not scale correctly with $m$ units because of the absence of quadruple excitations. Higher orders would similarly require even higher excitations until the FCI is reached! This failing of CI causes the paradoxical situation that a finite-order approximation to CID [e.g., MBPT(2) or MBPT(3)] frequently can be superior to the converged $E$(CID) result because the CI retains such nonextensive terms as $E^{(2)}\langle\psi^{(1)}|\psi^{(1)}\rangle$.

Although illustrated by CID, the inextensivity failing obviously pertains to any truncated CI, since the cancellations that must occur always involve the higher excitations that are not included in the truncated CI. The numerical effects of these nonextensive terms is important even for small molecules; and, today, most CI calculations include an estimate for these terms that allows their value to be subtracted from the CI energy (Davidson's approximation[34]). Many-body methods, instead, make the intelligent decision to eliminate all such nonphysical terms from the equations before calculation. This makes the method formally suited to describing many-bodies (i.e., many electrons). The term *many-body perturbation theory* emphasizes this cancellation of terms (also known as unlinked diagrams) that are not in the full CI solution. This leaves just the appropriate (linked) ones, in the energy and wavefunction: the so-called *linked-diagram theorem.*[45] Now, by building approximations on that equation we are already closer to the exact solution. This is a vastly better approach formally, operationally, and in terms of numerical accuracy for a given level of computation. Though extensive, MBPT is not variational, however, as negative corrections in Table 2 show.

The SDQ-MBPT(4) method, which includes all terms except the triple excitations, requires only an $\sim n^2N^4$ algorithm compared to $\sim n^3N^4N_{it}$ ($N_{it}$ = number of iterations) for CISDT. Allowing for its nonvariational character, it is generally better than CISDT (Table 2) because it has already incorporated most of the effects of quadruple excitations. Full MBPT(4) also benefits from the effect of triple excitations, is a noniterative $\sim n^3N^4$ procedure, and is considerably closer to the full CI than is CISDT. In fifth order the additional "con-

nected" effects of quadruple excitations are introduced, but these are comparatively small compared to those already in SDQ-MBPT(4). MBPT(5)[44] is an $\sim n^3N^5$ method but shows little numerical improvement over MBPT(4), despite its much greater expense. CISDTQ, which is also correct through fifth order, is still better than MBPT(5) because it includes all orders of perturbation theory among those categories of excitations.

Clearly, extensivity pays important dividends numerically, but it should be remembered that extensivity is only one possible source of error, and others can be equally or more important for a given problem. In particular, for some problems we might prefer the nonextensive multireference CI method instead of single reference MBPT or CC.

### *Coupled-Cluster Theory*

The failing of MBPT is that it is basically an order-by-order perturbation approach. For difficult correlation problems it is frequently necessary to go to high orders. This will be the case particularly when the single determinant reference function offers a poor approximation for the state of interest, as illustrated by the foregoing examples at 2.0 $R_e$. A practical solution to this problem is coupled-cluster (CC) theory.[46] In fact, CC theory simplifies the whole concept of extensive methods and the linked-diagram theorem into one very simple statement: the exponential wavefunction ansatz,

$$\Psi_{CC} = \exp(\hat{T})\Phi_0 = (1 + \hat{T} + \frac{1}{2}\hat{T}^2 + \frac{1}{3!}\hat{T}^3 + \cdots)\Phi_0 \qquad [36]$$

where

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots + \hat{T}_n$$

Just as in $\hat{C}_p$ in a CI wavefunction, $\hat{T}_p$ generates $p$-fold excitations. Hence, CCD gives

$$\begin{aligned}\psi_{CCD} &= (1 + \hat{T}_2 + \frac{1}{2}\hat{T}_2^2 + \cdots)\,\Phi_0 \\ &= \Phi_0 + \sum_{\substack{i<j \\ a<b}} t_{ij}^{ab}\,\Phi_{ij}^{ab} + \sum_{\substack{i<j<k<l \\ a<b<c<d}} t_{ij}^{ab}\,t_{kl}^{cd}\,\Phi_{ijkl}^{abcd} + \cdots\end{aligned} \qquad [37]$$

Unlike CID, even CCD[33] introduces quadruple excitations $\Phi_{ijkl}^{abcd}$ and hextuples $\Phi_{ijklmn}^{abcdef}$, etc., up to all $n$-tuples for $n$ electrons. However, the number of coefficients to determine in CCD $(t_{ij}^{ab})$ is exactly the same as in CID $(C_{ij}^{ab})$, because the coefficients of the higher excitations are products of the same coefficients. Since they arise from $\hat{T}_2^2/2$, $\hat{T}_2^3/3!$, etc., we call the latter "disconnected" terms (not unlinked; all terms in the exact wavefunction are linked!), which will lead us to a nonlinear set of equations for the determination of the $t_{ij}^{ab}$. "Connected" quadruples come from $\hat{T}_4$.

The correct scaling or extensive property is manifestly obvious in CC methods because for noninteracting $H_2$ molecules (described with localized orbitals),

$$\Psi(mH_2) = [\psi(H_2)]^m \qquad [38]$$

and in CC theory

$$\begin{aligned} \Psi_{CC} &= \exp(T(1) + T(2) + \cdots T(m))\Phi_0(mH_2) \\ &= \exp\,(mT)[\Phi_0(H_2)]^m \\ &= [\exp(T)\Phi_0(H_2)]^m \end{aligned} \qquad [39]$$

This means

$$E_{CC}(mH_2) = mE_{CC}(H_2) \qquad [40]$$

Note that Eq. [40] is true regardless of the subset of CC excitations $(T_p)$ included (CCD,[33] CCSD,[21] CCSDT,[22] etc.). Similarly, it is true for such approximations as CCSD + T(CCSD)[47] and CCSD(T)[48,49] (see below), where the initial contributions of triple excitations are computed noniteratively after a CCSD calculation, as well as essentially any other approximation based on CC and MBPT theory. [The CEPA (coupled electron-pair approximations)[50] are approximations to CC theory but do not necessarily retain all the invariance properties required to ensure rigorous extensivity.]

The use of the noninteracting limit above is pedagogical, but the consequences of the exponential ansatz clearly pertain to the interacting case, since there are no "unlinked" terms. It is important to recognize that it pertains to a high density electron gas, too, which mimics a metal. Then the exponential ansatz guarantees correct scaling with the number of electrons. *Extensivity,* whose rigorous definition is the absence of "unlinked diagrams,"[5,33] is a very general and important consistency condition for a correlated method, and it is absolutely essential for applications of quantum chemistry to extended systems.

The CC equations are obtained by inserting the wavefunction from Eq. [36] into the Schrödinger equation projected by the reference function $\Phi_0$ and single, double, and higher excitations. Using

$$H_N = H - \langle\Phi_0|H|\Phi_0\rangle$$

for CCD we have

$$\begin{aligned} &\langle\Phi_0|H_N \exp\,(T)|\Phi_0\rangle = \Delta E \\ &\langle\Phi_{ij}^{ab}|H_N \exp\,(T)|\Phi_0\rangle = \Delta E\langle\Phi_{ij}^{ab}|e^T|\Phi_0\rangle \text{ for all } i,j;\ a,b \end{aligned} \qquad [41]$$

where $\Delta E$ is the correlation energy. The equations projected by excitations provide equations for the coefficients, $\{t_{ij}^{ab}\}$. The actual equations are

$$\Delta E = \langle \Phi_0 | H_N \hat{T}_2 | \Phi_0 \rangle = \sum_{\substack{i>j \\ a>b}} t_{ij}^{ab} \langle \Phi_{ij}^{ab} | H_N | \Phi_0 \rangle = \sum_{\substack{i>j \\ a>b}} t_{ij}^{ab} \langle ab \| ij \rangle \qquad [42]$$

$$\Delta E t_{ij}^{ab} = \langle \Phi_{ij}^{ab} | H_N | \Phi_0 \rangle + \langle \Phi_{ij}^{ab} | H_N \hat{T}_2 | \Phi_0 \rangle + \langle \Phi_{ij}^{ab} | H_N \hat{T}_2^2 | \Phi_0 \rangle \qquad [43]$$

Notice, no higher terms than $\hat{T}_2^2$ can contribute to the CC equations because $\hat{T}_2^3$ would generate hextuple excitations, which would have vanishing matrix elements with $\Phi_{ij}^{ab}$. Therefore, even though the exponential CC wavefunction consists of all terms up to $n$-fold excitations, the equations for the coefficients $\{t_{ij}^{ab}\}$ are lower order.

Explicitly, the canonical orbital CCD equations can be further simplified to

$$\begin{aligned} \varepsilon_{ij}^{ab} t_{ij}^{ab} &= \langle ab \| ij \rangle + \frac{1}{2} \sum_{c,d} \langle ab \| cd \rangle t_{ij}^{cd} + \frac{1}{2} \sum_{k,l} \langle ij \| kl \rangle t_{kl}^{ab} \\ &+ \sum_{k,c} (\langle ak \| jc \rangle t_{ik}^{cb} - \langle bk \| jc \rangle t_{ik}^{ca} - \langle ak \| ic \rangle t_{jk}^{cb} + \langle bk \| ic \rangle t_{jk}^{ca}) \\ &+ \frac{1}{4} \sum_{\substack{k,l \\ c,d}} \langle kl \| cd \rangle [t_{kl}^{ab} t_{ij}^{cd} - 2(t_{ij}^{ac} t_{kl}^{bd} + t_{ij}^{bd} t_{kl}^{ac}) \\ &- 2(t_{ik}^{ab} t_{jl}^{cd} + t_{ik}^{cd} t_{jl}^{ab}) + 4(t_{ik}^{ac} t_{jl}^{bd} + t_{ik}^{bd} t_{jl}^{ac})] \end{aligned} \qquad [44]$$

Notice the elimination of $\Delta E$ from the CCD equations, unlike those for CID. Having *algebraic* instead of eigenvalue equations is a consequence of extensivity. If we include $T_1$ in addition to $T_2$, we would obtain coupled equations for $t_i^a$ and $t_{ij}^{ab}$. Including $T_3$ and its coefficients $t_{ijk}^{abc}$, we have coupled equations for $t_i^a$, $t_{ij}^{ab}$, $t_{ijk}^{abc}$, etc. Given a set of transformed two-electron integrals, the solution of the nonlinear CC equations is obtained by iteration exploiting various acceleration techniques. Table 3 compares results for CCD, CCSD, CCSDT, and CCSDTQ[51] to full CI. These are clearly the most accurate results we have yet considered. Like MBPT, CC results are not variational, and negative errors compared to full CI are possible.

The low order iterations of the CC equations recover MBPT approximations. For example, taking the first approximation to Eq. [44] and inserting it into Eq. [42],

**Table 3** Differences (millihartrees) Between FCI Energies and Coupled-Cluster Results (DZP Basis)

| Method | BH | | | FH | | | $H_2O$ | | | Mean absolute error |
|---|---|---|---|---|---|---|---|---|---|---|
| | $R_e$ | 1.5 $R_e$ | 2.0 $R_e$ | $R_e$ | 1.5 $R_e$ | 2.0 $R_e$ | $R_e$ | 1.5 $R_e$ | 2.0 $R_e$ | |
| CCD | 2.72 | 5.00 | 12.8 | 3.76 | 8.13 | 21.9 | 5.01 | 15.9 | 40.2 | 12.8 |
| CCSD | 1.79 | 2.64 | 5.05 | 3.01 | 5.10 | 10.2 | 4.12 | 10.2 | 21.4 | 7.06 |
| CCSD(T)[a] | 0.41 | 0.55 | 0.41 | 0.40 | 0.88 | −0.26 | 0.72 | 2.09 | 4.63 | 1.15 |
| CCSDT-1[b] | 0.46 | 0.62 | 0.96 | 0.16 | 0.48 | 0.56 | 0.59 | 1.96 | −2.88 | 0.96 |
| CCSDT[c] | 0.07 | 0.03 | −0.09 | 0.27 | 0.65 | 1.13 | 0.53 | 1.78 | −2.47 | 0.78 |
| CC5SD(TQ*)[d] | 0.05 | 0.01 | −0.59 | 0.33 | 0.56 | −0.18 | 0.19 | 0.13 | −1.96 | 0.44 |
| CCSDTQ[e] | 0.00 | 0.00 | +0.00 | 0.02 | 0.04 | 0.06 | 0.02 | 0.14 | −0.02 | 0.03 |

[a] Reference 49.
[b] Reference 52.
[c] Reference 22.
[d] Reference 49, 53.
[e] Reference 51.

$$t_{ij}^{ab(1)} = \frac{\langle ab \| ij \rangle}{\varepsilon_{ij}^{ab}}$$
$$E^{(2)} = \sum_{\substack{i>j \\ a>b}} t_{ij}^{ab(1)} \langle ab \| ij \rangle \qquad [45]$$

Plugging $t_{ij}^{ab(1)}$ into the second, third, and fourth summations on the right of Eq. [44] gives $t_{ij}^{ab(1)}$ and $E^{(3)} = \Sigma_{\substack{i>j \\ a>b}} t_{ij}^{ab(2)} \langle ab \| ij \rangle$. Fourth-order quadruple excitations arise from inserting $t_{ij}^{ab(1)}$ into the term quadratic in $t$, along with $t_{ij}^{ab(2)}$ inserted into the terms linear in $t$, because both terms contribute to $t_{ij}^{ab(3)}$. This order defines the DQ-MBPT(4) approximation, which continuing to convergence must give CCD. Including singles via the CCSD equations gives SDQ-MBPT(4), which at convergence gives CCSD. Similarly, if we also introduce the triple excitation $T_3$ equations, we obtain SDTQ-MBPT(4) or just MBPT(4) as the fourth-order approximation to CCSDT.

Computationally, CCD and CCSD are $\sim n^2N^4N_{it}$ methods, whereas CCSDT is $\sim n^3N^5N_{it}$ and if $T_4$ were added to give CCSDTQ, we would require an $\sim n^4N^6N_{it}$ algorithm. The corresponding CI methods have the same computational dependence but do not introduce the higher excitations that are found in the CC methods.

It is possible to combine some aspects of MBPT with CC theory to retain much of the infinite-order aspect of CC while saving significantly in the computational cost for higher cluster operators. This may be done iteratively or noniteratively. Taking just the most important contributions of $T_3$ (as judged by order in perturbation theory) and coupling to all $T_1$ and $T_2$ terms, we obtain CCSDT-1.[52] This still requires an $\sim n^3N^4N_{it}$ algorithm, though. Results are shown in Table 3.

If we further consider only these $T_3$ terms in CCSDT-1 noniteratively (i.e., we do not let the $T_3$ coefficients change), we obtain highly accurate but considerably less expensive methods. This is done as follows. The fourth-order triple excitation correction may be written as

$$\sum_{\substack{i>j>k \\ a>b>c}} (t_{ijk}^{abc(2)})^2 \varepsilon_{ijk}^{abc} \qquad [46]$$

where $t_{ijk}^{abc(2)}$ would consist of several terms derived from $T_2$ coefficients of the form $\Sigma_e \langle bc \| ei \rangle t_{jk}^{ae(1)}$ and $\Sigma_m \langle ma \| jk \rangle t_{mi}^{bc(1)}$. The former requires an $n^3N^4$ evaluation. If we simply use the converged $t_{jk}^{ae}$ and $t_{mi}^{bc}$ coefficients from CCSD instead of their first-order approximations (Eq. [45]) in Eq. [46], we would have higher than fourth-order terms. In this manner we evaluate the $n^3N^4$ step just one time and obtain CCSD + T(CCSD)[47]. If, in addition, we introduce a term like $t_i^a \langle bc \| jk \rangle$ that arises in CCSDT-1 from considering the contribution of $T_3$ to the $T_1$ equation, (a fifth-order term), we obtain CCSD(T).[48] Both are

clearly correct through fourth order and should be an improvement over MBPT(4). In this manner much of the often important effect of $T_3$ can be introduced without solving the full CCSDT equations, which is a substantial savings. Similar noniterative methods have been developed that include the initial $T_4$ contributions and are correct through fifth order, CCSD + TQ*(CCSD) = CC5SD(TQ*),[49] whose rate-determining step is a single $n^3N^5$ step.

Again recalling that unlike CI, CC methods are not variational, we can compare various CC methods with full CI and the prior approximations in Table 4. In terms of the energies, it is apparent that CC methods are closer to full CI than the corresponding CI or MBPT results, as must be clear from the theory. For example, CCSDT is closer than even CISDTQ. Even CCSD(T) is competitive with CISDTQ, yet the latter requires an $n^4N^6N_{it}$ computational dependence, while CCSD(T) is $n^2N^4N_{it}$ with a single $n^3N^4$ step. Also, CISDTQ is still not extensive. Similarly, the extensive fourth-order MBPT methods, SDQ-MBPT(4), which is the fourth-order approximation to CCSD, and MBPT(4), which is the fourth-order approximation to CCSDT or CCSD(T), differ from the infinite-order methods by ~5 millihartrees. Notice the improvement by a factor of better than 3 by CCSD compared to CISD and factor of ~16 improvement of CCSD(T) compared to CISDT. Whereas triples are important in CC theory, their improvement in CISDT is small. Although Table 4 is for just the energy, the overall relative quality of the various methods tends to

**Table 4** Comparison of Mean Absolute Errors for Different Correlated Methods Versus Full CI for BH, FH, and $H_2O$[a]

| Method | Mean absolute error (millihartrees) |
|---|---|
| CID | 25.9 |
| CISD | 22.0 |
| CISDT | 16.7 |
| CISDTQ | 1.13 |
| MBPT(2) | 27.8 |
| MBPT(3) | 22.9 |
| SDQ-MBPT(4) | 11.3 |
| MBPT(4) | 5.89 |
| MBPT(5) | 5.12 |
| CCD | 12.8 |
| CCSD | 7.06 |
| CCSD(T) | 1.15 |
| CCSDT-1 | 0.96 |
| CCSDT | 0.78 |
| CC5SD(TQ*) | 0.44 |
| CCSDTQ | 0.03 |

[a]Summary of data from Tables 1–3.

follow the errors shown. Hence we expect CCSD(T) > MBPT(4) > CCSD > SDQ-MBPT(4) > CISD > MBPT(2).

The Brueckner orbital variant of CC should also be mentioned.[54] CCSD puts in all single excitation effects via the wavefunction $\exp(T_1 + T_2)\Phi_0$. We can instead change the orbitals $\{\varphi_i\}$ in $\Phi_0$ in this wavefunction until $T_1 = 0$. These orbitals are called Brueckner orbitals and define a single determinant reference B instead of $\Phi_0$ that has maximum overlap with the correlated wavefunction. Since B-CCD[54] (or BD)[55] effectively puts in $T_1$, it will give results similar but not identical to those from CCSD (they differ in fifth order). For BH, the corresponding B-CCD errors are 1.81, 2.88, and 5.55, compared to 1.79, 2.64 and 5.05, for CCSD as a function of $R_e$. See also B-CCD for symmetry breaking problems.[56]

Quadratic configuration interaction (QCISD)[13] is an approximation to CCSD methods (it is also the same up to fifth-order terms) and will give similar results. Generally $E$(QCISD) < $E$(CCSD) because the higher order nonlinear terms in CCSD tend to be positive. For the purposes of the present discussion, we do not need to consider it separately. The details and numerical results compared to full CI are presented in the Appendix.

## Numerical Results for Potential Energy Curves

Besides FCI comparisons discussed above, correlated results for the $N_2$ and $F_2$ potential energy curves are informative. Because of the importance of nondynamic correlation in correctly separating the $N_2$ potential curve to two N($^4$S) atoms, the reference result shown in Figures 8 and 9 is the MR-CI, which is the same as CASSCF-CISD. Recall from Figure 6 that the CASSCF solution is qualitatively correct as a function of $R$, so augmenting it with all the single and double excitations that can be formed from excitations among the 32 spin-adapted configurations in the CASSCF should be excellent (even though not an extensive) approximation to the FCI. In this case it consists of 77,558 configurations. For the correctly separating UHF-based correlated methods, all the purely single reference, purely dynamic correlation methods go to the correct asymptotic limit except for MBPT(2) (Figure 8). The characteristic spin contamination of the UHF that persists into the correlated calculations is apparent from the erroneous, too steep, curvature near 1.3 Å. The latter is never alleviated by finite-order MBPT results, but is once the CCSD result has been obtained. That is, despite no explicit consideration of nondynamic correlation, UHF-CCSD offers a perfectly reasonable approximation to the $N_2$ curve all the way to separation. Adding triples either noniteratively, CCSD(T), or completely, CCSDT, provides results that follow the CASSCF-CISD curve very accurately. Notice that the CCSD(T) and CCSDT energies (along with MBPT(4)) are actually lower in the vicinity of equilibrium and, if we had the full CI result, would be closer to it than the CI. This reflects the remaining

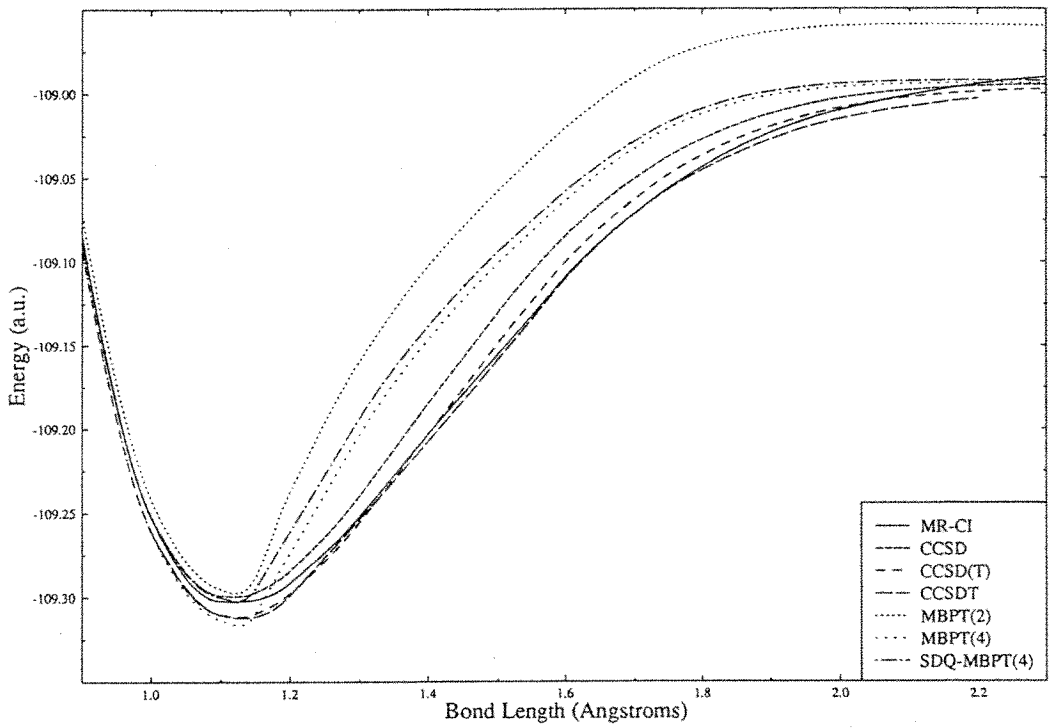


**Figure 8** Correlated potential curves for $N_2$ using a UHF reference.

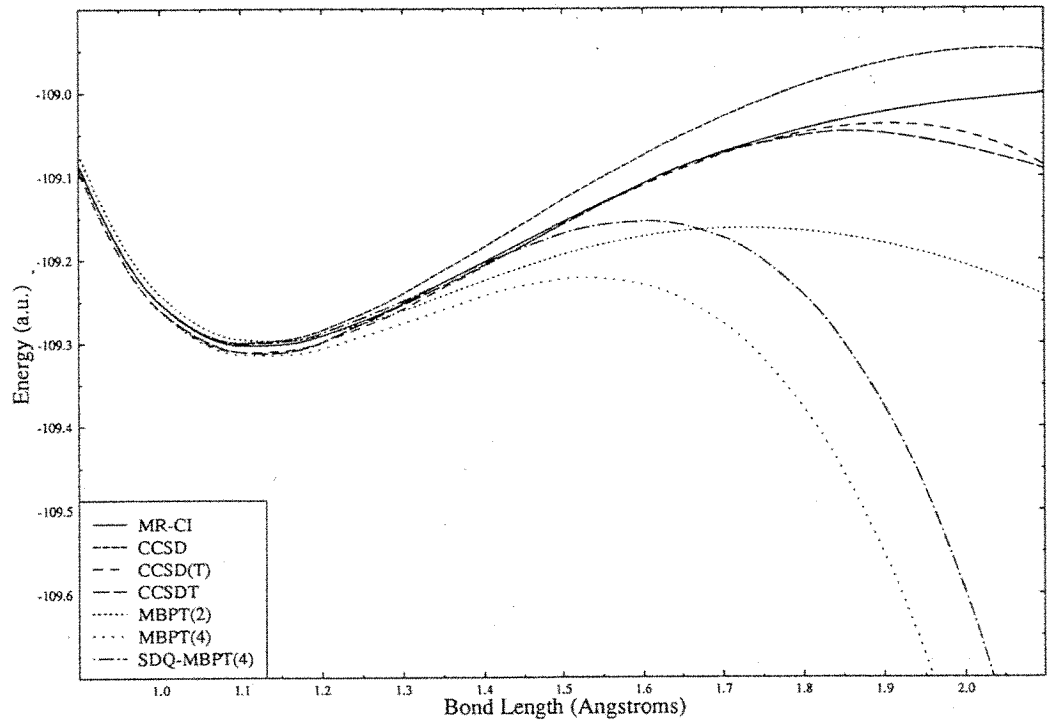


**Figure 9** Correlated potential curves for $N_2$ using an RHF reference.

inextensivity error even in a large CASSCF-CISD calculation. The CI still does not have all the benefit of triple and quadruple excitations, which is why most such calculations will still be augmented by a semiempirical estimate of the extensivity correction. It is clear from the above that whereas CC methods should be able to describe the curvature accurately, MBPT frequencies will suffer from spin contamination.

Considering the RHF-based CC and MBPT methods (Figure 9), to obtain a reasonable curve, the CC and MBPT methods must overcome a reference that is incorrectly separating. From Figure 6 that means correcting for about 0.8 hartree (or ~500 kcal/mol!). The fact that CCSD(T) and CCSDT can largely do that is rather remarkable! In practice that means that CCSD and CCSD(T) are fairly good to about 1.5 Å and CCSDT is nearly superimposable upon the CASSCF-CISD result until about 1.9 Å. Once again near the minimum where UHF and RHF coincide, the many-body methods with triples have lower energies. Here, finite-order MBPT shows the characteristic turnover that results when an energy denominator $\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b$ approaches zero where $\epsilon_i + \epsilon_j \approx \epsilon_a + \epsilon_b$, as some orbitals are exactly degenerate at separation. The denominator is negative, which causes the result to go to negative infinity. Unlike the UHF-based results, finite-order RHF-MBPT is qualitatively correct near equilibrium, however.

Figures 10 and 11 show a similar set of curves for $F_2$. Figure 12 shows the differences in dissociation energies. $F_2$ is complicated by its small dissociation energy and its unbound UHF solution. However, it is simpler than $N_2$ in that only a single bond is being broken. For single bond breaking, CC methods will not show the turnover observed for the $N_2$ RHF-based examples. Comparisons with the GVB-CISD = CASSCF-CISD demonstrate the general accuracy of nondynamic CC methods for single bond breaking. In particular, results from the simplest iterative triples model, CCSDT-1[52] (Figure 12) are in good agreement with the GVB-CISD at separation and better than those from its noniterative counterpart, CCSD(T). As for $N_2$, the many-body results give deeper minima and better dissociation energies for $F_2$ than does CASSCF-CISD. Here, of course, the CAS space consists of only two orbitals and the CI contains only 6620 configurations. The RHF-MBPT curves still turn over (Figure 10), but farther out on the potential curve than for $N_2$. Because of UHF's qualitative failure for $F_2$, UHF-MBPT shows no binding either (Figure 11). However, once again, CCSD and CCSD(T) provide qualitatively correct potential energy curves (Figure 12) although the former is too steep near 1.8 Å as it is for $N_2$.

To summarize, CC/MBPT methods are usually more accurate near equilibrium geometries than are nondynamic correlated methods, but less accurate asymptotically. These two examples are extreme cases chosen to show failures. Correctly breaking a C—H bond in $CH_4$ or C—C bond in ethane is not too demanding for these methods. Because chemistry is mostly accomplished by breaking single bonds, most situations involving bond cleavage are amendable to single-reference CC methods. In particular, there are no particular problems