
4 Electron Correlation Methods

The Hartree–Fock method generates solutions to the Schrödinger equation where the real electron–electron interaction is replaced by an average interaction (Chapter 3). In a sufficiently large basis, the HF wave function is able to account for $\sim 99\%$ of the total energy, but the remaining $\sim 1\%$ is often very important for describing chemical phenomena. The difference in energy between the HF and the lowest possible energy in a given basis set is called the *Electron Correlation* (EC) energy.¹ Physically it corresponds to the motion of the electrons being correlated, on average they are further apart than described by the HF wave function. As shown below, an UHF type of wave function is to a certain extent able to include electron correlation. The proper reference for discussing electron correlation is therefore a restricted (RHF or ROHF) wave function, although many authors use a UHF wave function for open-shell species. In the RHF case all the electrons are paired in molecular orbitals. The two electrons in a MO occupy the same physical space, and differ only in the spin function. The spatial overlap between the orbitals of two such “pair” -electrons is (exactly) one, while the overlap between two electrons belonging to different pairs is (exactly) zero, owing to the orthonormality of the MOs. This not the same as saying that there is no repulsion between electrons in different MOs, since the electron–electron repulsion integrals involve products of MOs ($\langle \phi_i | \phi_j \rangle = 0$ for $i \neq j$, but $\langle \phi_i \phi_j | \mathbf{g} | \phi_i \phi_j \rangle$ and $\langle \phi_i \phi_j | \mathbf{g} | \phi_j \phi_i \rangle$ are not necessarily zero).

Naively it may be expected that the correlation between pairs of electrons belonging to the same spatial MO would be the major part of the electron correlation. However, as the size of the molecule increases, the number of electron pairs belonging to different spatial MOs grows faster than those belonging to the same MO. Consider for example the valence orbitals for CH_4 . There are four intraorbital electron pairs of opposite spin, but there are 12 interorbital pairs of opposite spin, and 12 interorbital pairs of the same spin. A typical value for the intraorbital pair correlation of a single bond is ~ 20 kcal/mol, while that of an interorbital pair (where the two MO are spatially close, as in CH_4) is ~ 1 kcal/mol. The interpair correlation is therefore often comparable to the intrapair contribution.

Since the correlation between opposite spins has both intra- and inter-orbital contributions, it will be larger than the correlation between electrons having the same spin. The Pauli principle (or equivalently the antisymmetry of the wave function) has the consequence that there is no intraorbital correlation from electron pairs with the same spin. The opposite spin correlation is sometimes called the *Coulomb correlation*, while the same spin correlation is called the *Fermi correlation*, i.e. the Coulomb correlation is the largest contribution. Another way of looking at electron correlation is in terms of the electron density. In the immediate vicinity of an electron, there is a reduced probability of finding another electron. For electrons of opposite spin, this is often referred to as the *Coulomb hole*, the corresponding phenomenon for electrons of the same spin is the *Fermi hole*.

The HF method determines the best one-determinant trial wave function (within the given basis set). It is therefore clear that in order to improve on HF results, the starting point must be a trial wave function which contains more than one Slater Determinant (SD) Φ . This also means that the mental picture of electrons residing in orbitals has to be abandoned, and the more fundamental property, the electron density, should be considered. As the HF solution usually gives $\sim 99\%$ of the correct answer, electron correlation methods normally use the HF wave function as a starting point for improvements.

A generic multi-determinant trial wave function can be written as

$$\Psi = a_0 \Phi_{\text{HF}} + \sum_{i=1} a_i \Phi_i \quad (4.1)$$

where a_0 usually is close to 1. Electron correlation methods differ in how they calculate the coefficients in front of the other determinants, a_0 being determined by the normalization condition.

As mentioned in Chapter 5, one can think of the expansion of an unknown MO in terms of basis functions as describing the MO “function” in the “coordinate system” of the basis functions. The multi-determinant wave function (4.1) can similarly be considered as describing the total wave function in a “coordinate” system of Slater determinants. The basis set determines the size of the one-electron basis (and thus limits the description of the one-electron functions, the MOs), while the number of determinants included determines the size of the many-electron basis (and thus limits the description of electron correlation).

4.1 Excited Slater Determinants

How are the additional determinants beyond the HF constructed? With N electrons and M basis functions, solution of the Roothaan–Hall equations for the RHF case will yield $N/2$ occupied MOs and $M - N/2$ unoccupied (virtual) MOs. Except for a minimum basis, there will always be more virtual than occupied MOs. A Slater determinant is determined by $N/2$ spatial MOs multiplied by two spin functions to yield N spinorbitals. By replacing MOs which are occupied in the HF determinant by MOs which are unoccupied, a whole series of determinants may be generated. These can be denoted according to how many occupied HF MOs have been replaced by unoccupied MOs, i.e. Slater determinants which are *singly*, *doubly*, *triply*, *quadruply* etc. *excited* relative to the HF determinant, up to a maximum of N excited electrons. These

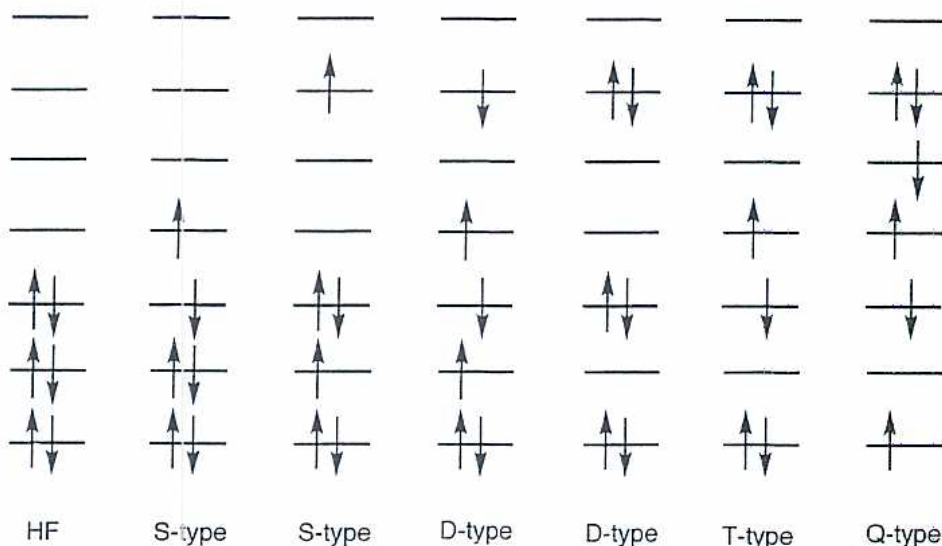


Figure 4.1 Excited Slater determinants generated from a HF reference

determinants are often referred to as *Singles (S)*, *Doubles (D)*, *Triples (T)*, *Quadruples (Q)* etc.

The total number of determinants that can be generated depends on the size of the basis set, the larger the basis, the more virtual MOs, and the more excited determinants can be constructed. If all possible determinants in a given basis set are included, all the electron correlation (in the given basis) is (or can be) recovered. For an infinite basis the Schrödinger equation is then solved exactly. Note that “exact” in this context is not the same as the experimental value, as the nuclei are assumed to have infinite masses (Born–Oppenheimer approximation) and relativistic effects are neglected. Methods which include electron correlation are thus two-dimensional, the larger the one-electron expansion (basis set size) and the larger the many-electron expansion (number of determinants), the better are the results. This is illustrated in Figure 4.2.

In order to calculate total energies with a “chemical accuracy” of ~ 1 kcal/mol, it is necessary to use sophisticated methods for including electron correlation and large basis sets, which is only computationally feasible for small systems. Instead the focus is usually on calculating relative energies, trying to make the errors as constant as possible.

EC \	Basic	Minimum	DZ	DZP	Infinite
HF (0%)						HF limit
10%						
.....						
100%						“Exact”

Figure 4.2 Convergence to the exact solution

The important chemical changes take place in the valence orbitals; the core orbitals are almost constant. In many cases the interest is therefore only in calculating the correlation energy associated with the valence electrons. Limiting the number of determinants to only those which can be generated by exciting the valence electrons is known as the *frozen core* approximation. In some cases the highest virtual orbitals corresponding to the anti-bonding combinations of the core orbitals are also removed from the correlation treatment (*frozen virtuals*). The frozen core approximation is not justified in terms of total energy; the correlation of the core electrons gives a substantial energy contribution. However, it is essentially a constant factor, which drops out when calculating relative energies. Furthermore, if we really want to calculate the core electron correlation, the standard basis sets are insufficient. In order to represent the angular correlation, higher angular momentum functions with the same radial size as the filled orbitals are needed, e.g. p- and d-functions with large exponents for correlating the 1s-electrons, as discussed in Section 5.4.5. Just allowing excitations of the core electrons in a standard basis set does not “correlate” the core electrons.

There are three main methods for calculating electron correlation: *Configuration Interaction (CI)*, *Many Body Perturbation Theory (MBPT)* and *Coupled Cluster (CC)*. A word of caution before we describe these methods in more details. The Slater determinants are composed of spin-MOs, but since the Hamilton operator is independent of spin, the spin dependence can be factored out. Furthermore, to facilitate notation, it is often assumed that the HF determinant is of the RHF type. Finally, many of the expressions below involve double summations over identical sets of functions. To ensure only the unique terms are included, one of the summation indices must be restricted. Alternatively, both indices can be allowed to run over all values, and the overcounting corrected by a factor of 1/2. Various combinations of these assumptions result in final expressions which differ by factors of 1/2, 1/4 etc. from those given here. In the present book the MOs are always spin-MOs, and conversion of a restricted summation to an unrestricted is always noted explicitly.

Finally a comment on notation. The quality of a calculation is given by the level of theory (i.e. how much electron correlation is included) and the size of the basis set. In a commonly used /-notation, introduced by J. A. Pople, this is denoted as “level/basis”. If nothing further is specified, this implies that the geometry is optimized at this level of theory. As discussed in Section 5.5, the geometry is usually much less sensitive to the theoretical level than relative energies, and high-level calculations are therefore often carried out using geometries optimized at a lower level. This is denoted as “level2/basis2//level1/basis1”, where the notation after the // indicates the level at which the geometry is optimized.

4.2 Configuration Interaction

This is perhaps the easiest method to understand. It is based on the variational principle (Appendix B), analogous to the HF method. The trial wave function is written as a linear combination of determinants with the expansion coefficients determined by requiring that the energy should be a minimum (or at least stationary), a procedure known as *Configuration Interaction (CI)*. The MOs used for building the excited Slater determinants are taken from a Hartree–Fock calculation and held fixed. Subscripts S, D, T etc. indicate determinants which are singly, doubly, triply etc. excited relative to the

HF configuration.

$$\Psi_{\text{CI}} = a_0 \Phi_{\text{SCF}} + \sum_{\text{S}} a_{\text{S}} \Phi_{\text{S}} + \sum_{\text{D}} a_{\text{D}} \Phi_{\text{D}} + \sum_{\text{T}} a_{\text{T}} \Phi_{\text{T}} \dots = \sum_{i=0} a_i \Phi_i \quad (4.2)$$

This is an example of a constrained optimization, the energy should be minimized under the constraint that the total CI wave function is normalized. Introducing a Lagrange multiplier (Section 14.6), this can be written as

$$L = \langle \Psi_{\text{CI}} | \mathbf{H} | \Psi_{\text{CI}} \rangle - \lambda [\langle \Psi_{\text{CI}} | \Psi_{\text{CI}} \rangle - 1] \quad (4.3)$$

The first bracket is the energy of the CI wave function, the second bracket is the norm of the wave function. In terms of determinants (eq. (4.2)), these can be written as

$$\begin{aligned} \langle \Psi_{\text{CI}} | \mathbf{H} | \Psi_{\text{CI}} \rangle &= \sum_{i=0} \sum_{j=0} a_i a_j \langle \Phi_i | \mathbf{H} | \Phi_j \rangle = \sum_{i=0} a_i^2 E_i + \sum_{i=0} \sum_{j \neq i} a_i a_j \langle \Phi_i | \mathbf{H} | \Phi_j \rangle \\ \langle \Psi_{\text{CI}} | \Psi_{\text{CI}} \rangle &= \sum_{i=0} \sum_{j=0} a_i a_j \langle \Phi_i | \Phi_j \rangle = \sum_{i=0} a_i^2 \langle \Phi_i | \Phi_i \rangle = \sum_{i=0} a_i^2 \end{aligned} \quad (4.4)$$

The diagonal elements in the sum involving the Hamilton operator are energies of the corresponding determinants. The overlap elements between different determinants are zero as they are built from orthogonal MOs (eq. (3.20)). The variational procedure corresponds to setting all the derivatives of the Lagrange function (4.3) with respect to the a_i expansion coefficients equal to zero.

$$\begin{aligned} \frac{\partial L}{\partial a_i} &= 2 \sum_j a_j \langle \Phi_i | \mathbf{H} | \Phi_j \rangle - 2\lambda a_i = 0 \\ a_i (\langle \Phi_i | \mathbf{H} | \Phi_i \rangle - \lambda) + \sum_{j \neq i} a_j \langle \Phi_i | \mathbf{H} | \Phi_j \rangle &= 0 \\ a_i (E_i - \lambda) + \sum_{j \neq i} a_j \langle \Phi_i | \mathbf{H} | \Phi_j \rangle &= 0 \end{aligned} \quad (4.5)$$

If there is only one determinant in the expansion ($a_0 = 1$), the last equation shows that the Lagrange multiplier is the (CI) energy, $\lambda = E$.

As there is one equation (4.5) for each i , the variational problem is transformed into solving a set of CI *secular equations*. Introducing the notation $H_{ij} = \langle \Phi_i | \mathbf{H} | \Phi_j \rangle$ the matrix equation becomes

$$\begin{pmatrix} H_{00} - E & H_{01} & \dots & H_{0j} & \dots \\ H_{10} & H_{11} - E & \dots & H_{1j} & \dots \\ \dots & \dots & \dots & \dots & \dots \\ H_{j0} & \dots & \dots & H_{jj} - E & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_j \\ \dots \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \\ \dots \end{pmatrix} \quad (4.6)$$

which in shorthand notation may be written as $(\mathbf{H} - E\mathbf{I}) \mathbf{a} = \mathbf{0}$ or as $\mathbf{H}\mathbf{a} = E\mathbf{a}$. Solving the secular equations is equivalent to diagonalizing the CI matrix, see Chapter 13. The CI energy is obtained as the lowest eigenvalue of the CI matrix, and the corresponding eigenvector contains the a_i coefficients in front of the determinants in eq. (4.2). The second lowest eigenvalue corresponds to the first excited state etc.

4.2.1 CI Matrix Elements

The CI matrix elements H_{ij} can be evaluated by the strategy employed for calculating the energy of a single determinant used for deriving the Hartree–Fock equations, Section 3.3. This involves expanding the determinants in a sum of products of MOs, thereby making it possible to express the CI matrix elements in terms of MO integrals. There are, however, some general features which make many of the CI matrix elements equal to zero.

The Hamilton operator (eq. (3.23)) does not contain spin, thus if two determinants have different total spin the corresponding matrix element is zero. This situation occurs if an electron is excited from an α spin-MO to a β spin-MO, such as the second S-type determinant in Figure 4.1. When the HF wave function is a singlet, this excited determinant is (part of) a triplet. The corresponding CI matrix element can be written in terms of integrals over MOs, and the spin dependence can be separated out. If there are different numbers of α and β spin-MOs, there will always be at least one integral $\langle \alpha | \beta \rangle = 0$. That matrix elements between different spin states are zero may be fairly obvious. If we are interested in a singlet wave function, only singlet determinants can enter the expansion with non-zero coefficients. However, if the Hamilton operator includes for example the spin–orbit operator, matrix elements between singlet and triplet determinants are not necessarily zero, and the resulting CI wave function will be a mixture of singlet and triplet states.

Consider now the case where an electron with α spin is moved from orbital i to orbital a . The first S-type determinant in Figure 4.1 is of this type. Alternatively, the electron with β spin could be moved from orbital i to orbital a . Both of these excited determinants will have an S_z value of 0, but neither are eigenfunctions of the S^2 operator. The difference and sum of these two determinants describe a singlet state and the $S_z = 0$ component of a triplet (which depends on the exact definition of the determinants).

Such linear combinations of determinants, which are proper spin eigenfunctions, are called *Spin-Adapted Configurations* (SAC) or *Configurational State Functions* (CSF). For higher excited states construction of proper CSFs may involve several determinants. The first D-type excitation in Figure 4.1, for example, must be combined with five other determinants corresponding to rearrangement of the electron spins to make a singlet CSF (actually there are two linearly independent CSFs that can be made). The second D-

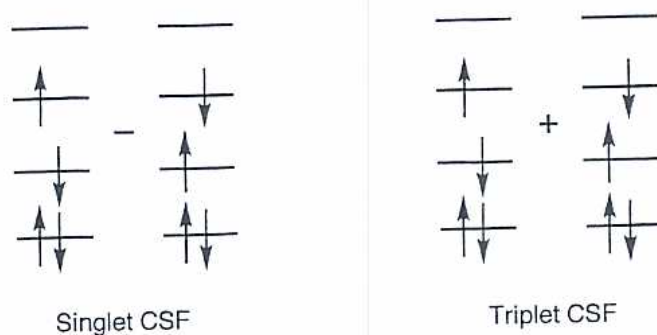


Figure 4.3 Forming configurational state functions from Slater determinants

type determinant in Figure 4.1 is already a proper CSF. By making suitable linear combinations of determinants the number of non-zero CI matrix elements can therefore be reduced.

If the system contains symmetry, there are additional CI matrix elements which become zero. The symmetry of a determinant is given as the direct product of the symmetries of the MOs. The Hamilton operator always belongs to the totally symmetric representation, thus if two determinants belong to different irreducible representations, the CI matrix element is zero. This is again fairly obvious: if the interest is in a state of a specific symmetry, only those determinants which have the correct symmetry can contribute.

The Hamilton operator consists of a sum of one-electron and two-electron operators, eq. (3.24). If two determinants differ by more than two (spatial) MOs there will always be an overlap integral between two different MOs which is zero (same argument as in eq. (3.28)). CI matrix elements can therefore only be non-zero if the two determinants differ by 0, 1, or 2 MOs, and they may be expressed in terms of integrals of one- and two-electron operators over MOs. These connections are known as the *Slater–Condon* rules. If the two determinants are identical, the matrix element is simply the energy of a single determinant wave function, as given by eq. (3.32). For matrix elements between determinants differing by 1 (exciting an electron from orbital i to a) or 2 (exciting two electrons from orbitals i and j to orbitals a and b) MOs, the results can be shown to be (compare with eq. (3.32), where the \mathbf{g} operator is implicit in the notation for the two-electron integrals (eq. (3.56)):

$$\begin{aligned}\langle \Phi_0 | \mathbf{H} | \Phi_i^a \rangle &= \langle \phi_i | \mathbf{h} | \phi_a \rangle + \sum_j (\langle \phi_i \phi_j | \phi_a \phi_j \rangle - \langle \phi_i \phi_j | \phi_j \phi_a \rangle) \\ \langle \Phi_0 | \mathbf{H} | \Phi_{ij}^{ab} \rangle &= \langle \phi_i \phi_j | \phi_a \phi_b \rangle - \langle \phi_i \phi_j | \phi_b \phi_a \rangle\end{aligned}\quad (4.7)$$

The matrix element between the HF and a singly excited determinant is a matrix element of the Fock operator between two different MOs (eq. (3.36)).

$$\langle \phi_i | \mathbf{h} | \phi_a \rangle + \sum_j (\langle \phi_i \phi_j | \phi_a \phi_j \rangle - \langle \phi_i \phi_j | \phi_j \phi_a \rangle) = \langle \phi_i | \mathbf{F} | \phi_a \rangle \quad (4.8)$$

This is an occupied-virtual off-diagonal element of the Fock matrix in the MO basis, and is identical to the gradient of the energy with respect to an occupied-virtual mixing parameter (except for a factor of 4), see eq. (3.67). If the determinants are constructed from optimized canonical HF MOs, the gradient is zero, and the matrix element is zero. This may also be realized by noting that the MOs are eigenfunctions of the Fock operator, eq. (3.41).

$$\begin{aligned}\mathbf{F}\phi_a &= \varepsilon_a \phi_a \\ \langle \phi_i | \mathbf{F} | \phi_a \rangle &= \varepsilon_a \langle \phi_i | \phi_a \rangle = \varepsilon_a \delta_{ia}\end{aligned}\quad (4.9)$$

The disappearance of matrix elements between the HF reference and singly excited states is known as *Brillouins theorem*. The HF reference state therefore only has non-zero matrix elements with doubly excited determinants, and the full CI matrix acquires a block diagonal structure.

In order to evaluate the CI matrix elements one- and two-electron integrals over MOs are needed. These can be expressed in terms of the corresponding AO integrals and the

CI matrix	Φ_{HF}	Φ_{S}	Φ_{D}	Φ_{T}	Φ_{Q}	Φ_{Q}	...
Φ_{HF}	E_{HF}	0		0	0	0	0
Φ_{S}	0				0	0	0
Φ_{D}						0	0
Φ_{T}	0						0
Φ_{Q}	0	0					
Φ_{Q}	0	0	0				
...	0	0	0	0			

Figure 4.4 Structure of the CI matrix

MO coefficients.

$$\begin{aligned} \langle \phi_i | \mathbf{h} | \phi_j \rangle &= \sum_{\alpha}^M \sum_{\beta}^M c_{\alpha i} c_{\beta j} \langle \chi_{\alpha} | \mathbf{h} | \chi_{\beta} \rangle \\ \langle \phi_i \phi_j | \phi_k \phi_l \rangle &= \sum_{\alpha}^M \sum_{\beta}^M \sum_{\gamma}^M \sum_{\delta}^M c_{\alpha i} c_{\beta j} c_{\gamma k} c_{\delta l} \langle \chi_{\alpha} \chi_{\beta} | \chi_{\gamma} \chi_{\delta} \rangle \end{aligned} \quad (4.10)$$

Such MO integrals are required for all electron correlation methods. The two-electron AO integrals are the most numerous and the above equation appears to involve a computational effort proportional to M^8 (M^4 AO integrals each multiplied by four sets of M MO coefficients). However, by performing the transformation one index at a time, the computational effort can be reduced to M^5 .

$$\langle \phi_i \phi_j | \phi_k \phi_l \rangle = \sum_{\alpha} c_{\alpha i} \left(\sum_{\beta} c_{\beta j} \left(\sum_{\gamma} c_{\gamma k} \left(\sum_{\delta} c_{\delta l} \langle \chi_{\alpha} \chi_{\beta} | \chi_{\gamma} \chi_{\delta} \rangle \right) \right) \right) \quad (4.11)$$

Each step now only involves multiplication of M^4 integrals with M coefficients, i.e. the M^8 dependence is reduced to four M^5 operations. In the large basis set limit, all electron correlation methods formally scale as at least M^5 , since this is the scaling for the AO to MO integral transformation. The transformation is an example of a "rotation" of the "coordinate" system consisting of the AOs, to one where the Fock operator is diagonal, the MOs, see Chapter 13. The diagonal system allows a much more compact representation of the matrix elements needed for the electron correlation treatment. The coordinate change is also known as a *four index transformation*, since it involves four indices associated with the basis functions.

4.2.2 Size of the CI Matrix

What is a typical size of the CI matrix? Consider a small system, H_2O with a 6-31G(d) basis. For the purpose of illustration, let us for a moment return to the spinorbital description. There are 10 electrons and 38 spin-MOs, of which 10 are occupied and 28 are empty. There are $K_{10,n}$ possible ways of selecting n electrons out of the 10 occupied orbitals, and $K_{28,n}$ ways of distributing them in the 28 empty orbitals. The number of excited states for a given excitation level is thus $K_{10,n} \cdot K_{28,n}$, and the total number of

Table 4.1 Number of singlet CSFs as a function of excitation level for H₂O with a 6-31G(d) basis

Excitation level <i>n</i>	Number of <i>n</i> th excited CSFs	Total number of CSFs
1	71	71
2	2 485	2 556
3	40 040	42 596
4	348 530	391 126
5	1 723 540	2 114 666
6	5 033 210	7 147 876
7	8 688 680	15 836 556
8	8 653 645	24 490 201
9	4 554 550	29 044 751
10	1 002 001	30 046 752

excited determinants will be a sum over 10 such terms. This is also equivalent to $K_{38,10}$, the total number of ways 10 electrons can be distributed in 38 orbitals.

$$\text{Number of SDs} = \sum_{n=0}^{10} K_{10,n} \cdot K_{28,n} = K_{38,10} = \frac{38!}{10! \cdot (38-10)!} \quad (4.12)$$

The number of excited determinants thus grows factorially with the size of the basis set. Many of these excited determinants will of course have different spin multiplicity (triplet, quintet etc. states for a singlet HF determinant), and can therefore be left out in the calculation. Generating only the singlet CSFs, the number of configurations at each excitation level is shown in Table 4.1.

The number of determinants (or CSFs) that can be generated grows wildly with the excitation level! Even if the C_{2v} symmetry of H₂O is employed, there are still a total of 7 536 400 singlet CSFs with A_1 symmetry. If all possible determinants are included we have a *full CI* wave function, there is no truncation in the many-electron expansion besides that generated by the finite one-electron expansion (size of the basis set). This is the best possible wave function within the limitations of the basis set, i.e. it recovers 100% of the electron correlation in the given basis. In this case it corresponds to diagonalizing a matrix of size $30\,046\,752 \times 30\,046\,752$. This is impossible. However, normally the interest is only in the lowest (or a few of the lowest) eigenvalue(s) and -vector(s), and there are special iterative methods (Section 4.2.4) for determining one (or a few) eigenvector(s) of a large matrix.

In the general case of N electrons and M basis functions the total number of singlet CSFs that can be generated is given by

$$\text{Number of CSFs} = \frac{M!(M+1)!}{\left(\frac{N}{2}\right)! \left(\frac{N}{2}+1\right)! \left(M-\frac{N}{2}\right)! \left(M-\frac{N}{2}+1\right)!} \quad (4.13)$$

For H₂O with the above 6-31G(d) basis there are $\sim 30 \times 10^6$ CSFs ($N = 10$, $M = 19$); with the larger 6-311G(2d,2p) basis there are $\sim 106 \times 10^9$ CSFs ($N = 10$, $M = 41$). For H₂C=CH₂ with the 6-31G(d) basis there are $\sim 334 \times 10^{12}$ CSFs ($N = 16$, $M = 38$).

One of the recent large scale full CI calculations considered H_2O in a DZP type basis with 24 functions.² Allowing all possible excitations of the 10 electrons generates 451 681 246 determinants. The variational wave function thus contains roughly half a billion parameters, i.e. the formal size of the CI matrix is of the order of half a billion squared. Although a determination of the lowest eigenvalue of such a problem can be done in a matter of hours on a modern computer, the result is a single number, the ground-state energy of the H_2O molecule. Owing to basis set limitations, however, it is still some 0.2 a.u. (~ 125 kcal/mol) larger than the experimental value. The computational effort for extracting a single eigenvalue and -vector scales essentially linearly with the number of CSFs, and it is possible to handle systems with up to a few billion determinants. The factorial growth of the number of determinants with the size of the basis set, however, makes the full CI method infeasible for all but the very smallest systems. Full CI calculations are thus not a routine computational procedure for including electron correlation, but they are useful references for developing more approximate methods, as full CI gives the best result that can be obtained in a given basis.

4.2.3 Truncated CI Methods

In order to develop a computationally tractable model, the number of excited determinants in the CI expansion (4.2) must be reduced. Truncating the excitation level at 1 (CI with Singles, CIS) does not give any improvement over the HF result as all matrix elements between the HF wave function and singly excited determinants are zero. CIS is equal to HF for the ground-state energy, although higher roots from the secular equations may be used as approximations to excited states. It has already been mentioned that only doubly excited determinants have matrix elements with the HF wave function different from zero, thus the lowest CI level which gives an improvement over the HF result is that which includes only doubly excited states, yielding the CID (CI with Doubles) model. Compared to the number of doubly excited determinants there are relatively few singly excited determinants (see for example Table 4.1), including these gives the CISD method. Computationally this is only a marginal increase in effort over CID. Although the singly excited determinants have zero matrix elements with the HF reference, they enter the wave function indirectly as they have non-zero matrix elements with the doubly excited determinants. In the large basis set limit the CISD method scales as M^6 . The next level in improvement is inclusion of the triply excited determinants, giving the CISDT method, which is an M^8 method. Taking into account also quadruply excited determinants yields the CISDTQ method which is an M^{10} method. As shown below, the CISDTQ model in general gives results close to the full CI limit, but even truncating the excitation level at 4 produces so many configurations that it can only be applied to small molecules and small basis sets. The only CI method which is generally applicable for a large variety of systems is CISD. For computationally feasible systems (i.e. medium size molecules and basis sets), it typically recovers 80–90% of the available correlation energy. The percentage is highest for small molecules. As the molecule gets larger the CISD method recovers less and less of the correlation energy, as discussed in Section 4.5.

Since only doubly excited determinants have non-zero matrix elements with the HF state, these are the most important. This may be illustrated by considering a full CI

Table 4.2 Weights of excited configurations for the Neon atom

Excitation level	Weight
0	0.9644945073
1	0.0009804929
2	0.0336865893
3	0.0003662339
4	0.0004517826
5	0.0000185090
6	0.0000017447
7	0.0000001393
8	0.0000000011

calculation for the Ne atom in a [5s,4p,3d] basis, where the 1s electrons are omitted from the correlation treatment.³ The contribution to the full CI wave function from each level of excitation is given in Table 4.2.

The weight is the sum of a_i^2 coefficients at the given excitation level, eq. (4.2). The CI method determines the coefficients from the variational principle, thus Table 4.2 shows that the doubly excited determinants are by far the most important in terms of energy. The singly excited determinants are the second most important, then follow the quadruples and triples. Excitations higher than 4 make only very small contributions, although there are actually many more of these highly excited determinants than the triples and quadruples, as illustrated in Table 4.1.

The relative importance of the different excitations may qualitatively be understood by noting that the doubles provide electron correlation for electron pairs. Quadruply excited determinants are important as they primarily correspond to products of double excitations. The singly excited determinants allow inclusion of multi-reference character in the wave function, i.e. they allow the orbitals to “relax”. Although the HF orbitals are optimum for the single determinant wave function, that is no longer the case when many determinants are included. The triply excited determinants are doubly excited relative to the singles, and can then be viewed as providing correlation for the “multi-reference” part of the CI wave function.

While singly excited states make relatively small contributions to the correlation energy of a CI wave function, they are very important when calculating properties (Chapter 10). Molecular properties measure how the wave function changes when a perturbation, such as an external electric field, is added. The change in the wave function introduced by the perturbation makes the MOs no longer optimal in the variational sense. The first-order change in the MOs is described by the off-diagonal elements in the Fock matrix, these are essentially the gradient of the HF energy with respect to the MOs (eq. 3.67). In the absence of a perturbation, these are zero; the HF energy is stationary with respect to an orbital variation (eq. (3.38)). As shown in eqs. (4.7) and (4.8), the Fock matrix off-diagonal elements are CI matrix elements between the HF and singly excited states. For molecular properties, the singly excited states thus allow the CI wave function to “relax” the MOs, i.e. letting the wave function respond to the perturbation.

4.2.4 Direct CI methods

As illustrated above, even quite small systems at the CISD level results in millions of CSFs. The variational problem is to extract one or possibly a few of the lowest eigenvalues and -vectors of a matrix the size of millions squared. This cannot be done by standard diagonalization methods where all the eigenvalues are found. There are, however, iterative methods for extracting one, or a few, eigenvalues and -vectors of a large matrix. The CI problem eq. (4.6) may be written as

$$(\mathbf{H} - E\mathbf{I})\mathbf{a} = 0 \quad (4.14)$$

The \mathbf{H} matrix contains the matrix element between the CSFs in the CI expansion, and the \mathbf{a} vector the expansion coefficients. The idea in iterative methods is to generate a suitable guess for the coefficient vector and calculate $(\mathbf{H} - E\mathbf{I})\mathbf{a}$. This will in general not be zero, and the deviation may be used for adding a correction to \mathbf{a} , forming an iterative algorithm. If the interest is in the lowest eigenvalue, a suitable start eigenvector may be one which contains only the HF configuration, i.e. $\{1,0,0,0,\dots\}$. Since the \mathbf{H} matrix elements essentially are two-electron integrals in the MO basis (eq. (4.7)), the iterative procedure may be formulated as integral driven, i.e. a batch of integrals are read in (or generated otherwise) and used directly in the multiplication with the corresponding \mathbf{a} -coefficients. The CI matrix is therefore not needed explicitly, only the effect of its multiplication with a vector containing the variational parameters, and storage of the entire matrix is avoided. This is the basis for being able to handle CI problems of almost any size, and is known as *direct CI*. Note that it is not "direct" in the sense used to describe the direct SCF method, where all the AO integrals are calculated as needed. The direct CI approach just assumes that the CI matrix elements (e.g. two-electron integrals in the MO basis) are available as required, traditionally stored in a file on a disk. There are several variations on how the \mathbf{a} vector is adjusted in each iteration, the most commonly used versions are based on the *Davidson algorithm*.⁴

4.3 Illustrating how CI Accounts for Electron Correlation, and the RHF Dissociation Problem

Consider the H_2 molecule in a minimum basis consisting of one s-function on each centre, χ_A and χ_B . A RHF calculation will produce two MOs, ϕ_1 and ϕ_2 , being the sum and difference of the two AOs. The sum of the two AOs is a bonding MO, with increased probability of finding the electrons between the two nuclei, while the difference is an antibonding MO, with decreased probability of finding the electrons between the two nuclei.

The HF wave function will have two electrons in the lowest energy (bonding) MO.

$$\begin{aligned} \phi_1 &= N_1(\chi_A + \chi_B) \\ \phi_2 &= N_2(\chi_A - \chi_B) \\ \Phi_0 &= \left| \begin{array}{c} \phi_1(1)\bar{\phi}_1(1) \\ \phi_1(2)\bar{\phi}_1(2) \end{array} \right| \end{aligned} \quad (4.15)$$

Here N_1 and N_2 are suitable normalization constants, and the bar above the MO indicates that the electron has a β spin function, no bar indicates an α spin function. In

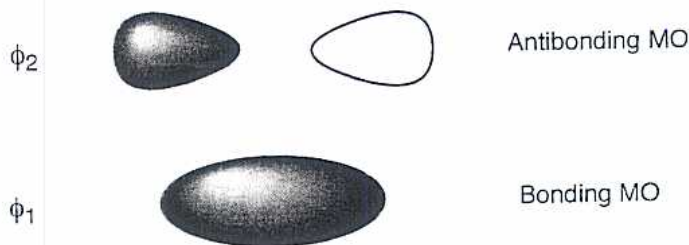


Figure 4.5 Molecular orbitals for H_2

this basis there are one doubly (Φ_1) and four singly excited Slater determinants (Φ_{2-5}).

$$\begin{aligned}
 \Phi_1 &= \begin{vmatrix} \phi_2(1) \bar{\phi}_2(1) \\ \phi_2(2) \bar{\phi}_2(2) \end{vmatrix} \\
 \Phi_2 &= \begin{vmatrix} \phi_1(1) \bar{\phi}_2(1) \\ \phi_1(2) \bar{\phi}_2(2) \end{vmatrix} \\
 \Phi_3 &= \begin{vmatrix} \bar{\phi}_1(1) \phi_2(1) \\ \bar{\phi}_1(2) \phi_2(2) \end{vmatrix} \\
 \Phi_4 &= \begin{vmatrix} \phi_1(1) \phi_2(1) \\ \phi_1(2) \phi_2(2) \end{vmatrix} \\
 \Phi_5 &= \begin{vmatrix} \bar{\phi}_1(1) \bar{\phi}_2(1) \\ \bar{\phi}_1(2) \bar{\phi}_2(2) \end{vmatrix}
 \end{aligned} \tag{4.16}$$

Configurations Φ_4 and Φ_5 are clearly the $S_z = 1$ and $S_z = -1$ components of a triplet state. The plus combination of Φ_2 and Φ_3 is the $S_z = 0$ component of the triplet, and the minus combination is a singlet configuration, Figure 4.3. The H_2 molecule belongs to the $D_{\infty h}$ point group, and the two MOs transform as the $\sigma_g(\phi_1)$ and $\sigma_u(\phi_2)$ representations. The singly excited CSF ($\Phi_2 - \Phi_3$) has overall Σ_u symmetry, while the HF (Φ_0) and doubly excited determinants (Φ_1) have Σ_g . The full 6×6 CI problem therefore blocks into a 2×2 block of singlet Σ_g states, a 1×1 block of singlet Σ_u , and a 3×3 block of triplet Σ_u states, Figure 4.6. Owing to orthogonality of the spin functions the triplet block is already diagonal.

The full CI for the $^1\Sigma_g$ states involves only two configurations, the reference HF and the doubly excited determinant.

$$\begin{aligned}
 \Phi_0 &= \phi_1(1)\bar{\phi}_1(2) - \bar{\phi}_1(1)\phi_1(2) = \phi_1\phi_1(\alpha\beta - \beta\alpha) \\
 \Phi_1 &= \phi_2(1)\bar{\phi}_2(2) - \bar{\phi}_2(1)\phi_2(2) = \phi_2\phi_2(\alpha\beta - \beta\alpha)
 \end{aligned} \tag{4.17}$$

In eq. (4.17) the electron coordinate is given implicitly by the order in which the orbitals are written, i.e. $\phi_1\phi_1[\alpha\beta - \beta\alpha] = \phi_1(1)\phi_1(2)[\alpha(1)\beta(2) - \beta(1)\alpha(2)]$. Ignoring the spin functions (which may be integrated out since \mathbf{H} is spin independent) and normalization, the determinants can be expanded in AOs.

$$\begin{aligned}
 \Phi_0 &= (\chi_A(1) + \chi_B(1))(\chi_A(2) + \chi_B(2)) = \chi_A\chi_A + \chi_B\chi_B + \chi_A\chi_B + \chi_B\chi_A \\
 \Phi_1 &= (\chi_A(1) - \chi_B(1))(\chi_A(2) - \chi_B(2)) = \chi_A\chi_A + \chi_B\chi_B - \chi_A\chi_B - \chi_B\chi_A
 \end{aligned} \tag{4.18}$$

	${}^1\Phi_0(\sum_g)$	${}^1\Phi_1(\sum_g)$	${}^1(\Phi_2 - \Phi_3)(\sum_u)$	${}^3\Phi_4(\sum_u)$	${}^3(\Phi_2 + \Phi_3)(\sum_u)$	${}^3\Phi_5(\sum_u)$
${}^1\Phi_0(\sum_g)$			0	0	0	0
${}^1\Phi_1(\sum_g)$			0	0	0	0
${}^1(\Phi_2 - \Phi_3)(\sum_u)$	0	0		0	0	0
${}^3\Phi_4(\sum_u)$	0	0	0		0	0
${}^3(\Phi_2 + \Phi_3)(\sum_u)$	0	0	0	0		0
${}^3\Phi_5(\sum_u)$	0	0	0	0	0	

Figure 4.6 Structure of the full CI matrix for the H_2 system in a minimum basis.

The first two terms on the right-hand side have both electrons on the same centre, they describe *ionic* contributions to the wave function, H^+H^- . The later two terms describe *covalent* contributions to the wave function, $H'H'$. The HF wave function thus contains equal amounts of ionic and covalent contributions. The full CI wave function may be written in terms of AOs as

$$\begin{aligned}\Psi_{CI} &= a_0\Phi_0 + a_1\Phi_1 = a_0(\phi_1\phi_1) + a_1(\phi_2\phi_2) \\ \Psi_{CI} &= (a_0 + a_1)(\chi_A\chi_A + \chi_B\chi_B) + (a_0 - a_1)(\chi_A\chi_B + \chi_B\chi_A)\end{aligned}\quad (4.19)$$

The optimum values of the a_0 and a_1 coefficients are determined by the variational procedure. The HF wave function constrains both electrons to move in the same bonding orbital. By allowing the doubly excited state to enter the wave function, the electrons can better avoid each other, as the antibonding MO now is also available. The antibonding MO has a nodal plane (where $\phi_2 = 0$) perpendicular to the molecular axis (Figure 4.5), and the electrons are able to correlate their movements by being on opposite sides of this plane. This *left-right* correlation is a molecular equivalent of the atomic radial correlation discussed in Section 5.2.

Consider now the behaviour of the HF wave function Φ_0 (eq. (4.18)) as the distance between the two nuclei is increased toward infinity. Since the HF wave function is an equal mixture of ionic and covalent terms, the dissociation limit is 50% H^+H^- and 50% $H'H'$. In the gas phase all bonds dissociate homolytically, and the ionic contribution should be 0%. The HF dissociation energy is therefore much too high. This is a general problem of RHF type wave functions, the constraint of doubly occupied MOs is inconsistent with breaking bonds to produce radicals. In order for an RHF wave function to dissociate correctly, an even-electron molecule must break into two even-electron fragments, each being in the lowest electronic state. Furthermore, the orbital symmetries must match. There are only a few covalently bonded systems which obey these requirements (the simplest example is HHe^+). The wrong dissociation limit for RHF wave functions has several consequences.

- (1) The energy for stretched bonds is too high. Most transition structures have partly formed/broken bonds, thus activation energies are too high at the RHF level.

- (2) The excessively steep increase in energy as a function of the bond length causes the minimum on a potential energy curve to occur too “early” for covalently bonded systems, and equilibrium bond lengths are too short at the RHF level.
- (3) The excessively steep increase in energy as a function of the bond length causes the curvature of the potential energy surface near the equilibrium to be too large, and vibrational frequencies, especially those describing bond stretching, are in general too high.
- (4) The wave function contains too much “ionic” character, and RHF dipole moments (and also atomic charges) are in general too large.

It should be noted that dative bonds, like metal complexes and charge transfer species, in general have RHF wave functions which dissociate correctly, and the equilibrium bond lengths in these cases are normally too long.

The dissociation problem is solved in the case of a full CI wave function. As seen from eq. (4.19), the ionic term can be made to disappear by setting $a_1 = -a_0$. The full CI wave function generates the lowest possible energy (within the limitations of the chosen basis set) at all distances, with the optimum weights of the HF and doubly excited determinants determined by the variational principle. In the general case of a polyatomic molecule and a large basis set, correct dissociation of all bonds can be achieved if the CI wave function contains all determinants generated by a full CI in the valence orbital space. The latter corresponds to a full CI if a minimum basis is employed, but is much smaller than a full CI if an extended basis is used.

4.4 The UHF Dissociation, and the Spin Contamination Problem

The dissociation problem can also be “solved” by using a wave function of the UHF type. Here the α and β bonding MOs are allowed to “localize”, thereby reducing the MO symmetries to $C_{\infty v}$.

$$\begin{aligned}
 \phi_1 &= N(\chi_A + c\chi_B)\alpha \\
 \bar{\phi}_1 &= N(c\chi_A + \chi_B)\beta \\
 \Phi_0^{\text{UHF}} &= \begin{vmatrix} \phi_1(1) \bar{\phi}_1(1) \\ \phi_1(2) \bar{\phi}_1(2) \end{vmatrix}
 \end{aligned}
 \tag{4.20}$$

The optimum value of c is determined by the variational principle. If $c = 1$, the UHF wave function is identical to RHF. This will normally be the case near the equilibrium distance. As the bond is stretched, the UHF wave function allows each of the electrons to localize on a nucleus; c goes towards 0. The point where the RHF and UHF descriptions start to differ is often referred to as the RHF/UHF *instability* point. This is an example of symmetry breaking, as discussed in Section 3.8.3. The UHF wave function correctly dissociates into two hydrogen atoms, however, the symmetry breaking of the MOs has two other, closely connected, consequences: introduction of electron correlation and spin contamination. To illustrate these concepts, we need to look at the Φ_0 UHF determinant, and the six RHF determinants in eqs. (4.15) and (4.16) in more detail. We will again ignore all normalization constants.

The six RHF determinants can be expanded in terms of the AOs:

$$\begin{aligned}
 \Phi_0 &= [\chi_A\chi_A + \chi_B\chi_B + \chi_A\chi_B + \chi_B\chi_A](\alpha\beta - \beta\alpha) \\
 \Phi_1 &= [\chi_A\chi_A + \chi_B\chi_B - \chi_A\chi_B - \chi_B\chi_A](\alpha\beta - \beta\alpha) \\
 \Phi_2 &= [\chi_A\chi_A - \chi_B\chi_B](\alpha\beta - \beta\alpha) - [\chi_A\chi_B - \chi_B\chi_A](\alpha\beta + \beta\alpha) \\
 \Phi_3 &= [\chi_A\chi_A - \chi_B\chi_B](\alpha\beta - \beta\alpha) + [\chi_A\chi_B - \chi_B\chi_A](\alpha\beta + \beta\alpha) \\
 \Phi_4 &= [\chi_A\chi_B - \chi_B\chi_A](\alpha\alpha) \\
 \Phi_5 &= [\chi_A\chi_B - \chi_B\chi_A](\beta\beta)
 \end{aligned} \tag{4.21}$$

Subtracting and adding Φ_2 and Φ_3 produces a pure singlet ($^1\Phi_-$), and the $S_z = 0$ component of the triplet ($^3\Phi_+$), wave functions.

$$\begin{aligned}
 ^1\Phi_- &= \Phi_2 - \Phi_3 = [\chi_A\chi_A - \chi_B\chi_B](\alpha\beta - \beta\alpha) \\
 ^3\Phi_+ &= \Phi_2 + \Phi_3 = [\chi_A\chi_B - \chi_B\chi_A](\alpha\beta + \beta\alpha)
 \end{aligned} \tag{4.22}$$

Performing the expansion of the Φ_0^{UHF} determinant (4.20) gives

$$\begin{aligned}
 \Phi_0^{\text{UHF}} &= c[\chi_A\chi_A + \chi_B\chi_B](\alpha\beta - \beta\alpha) \\
 &\quad + [\chi_A\chi_B\alpha\beta - c^2\chi_A\chi_B\beta\alpha] \\
 &\quad + [c^2\chi_B\chi_A\alpha\beta - \chi_B\chi_A\beta\alpha]
 \end{aligned} \tag{4.23}$$

Adding and subtracting factors of $\chi_A\chi_B\beta\alpha$ and $\chi_B\chi_A\alpha\beta$ allow this to be written as

$$\begin{aligned}
 \Phi_0^{\text{UHF}} &= [c(\chi_A\chi_A + \chi_B\chi_B) + (\chi_A\chi_B + \chi_B\chi_A)](\alpha\beta - \beta\alpha) \\
 &\quad + (1 - c^2)[\chi_A\chi_B\beta\alpha - \chi_B\chi_A\alpha\beta]
 \end{aligned} \tag{4.24}$$

Since $0 \leq c \leq 1$ the first term shows that UHF orbitals reduce the ionic contribution relative to the covalent structures, compared to the RHF case, eq. (4.18). This is the same effect as for the CI procedure (eq. (4.19)), i.e. the first term shows that the UHF wave function partly includes electron correlation. The first term can be written as a linear combination of the Φ_0 and Φ_1 determinants, and describes a pure singlet state. The last part of the UHF determinant, however, has terms identical to two of those in the triplet $^3\Phi_+$ combination (4.22). If we had chosen the alternative set of UHF orbitals with the alpha spin being primarily on centre B in eq. (4.20), we would have obtained the other two terms in $^3\Phi_+$, i.e. the last term in (4.24) breaks the symmetry. The UHF determinant is therefore not a pure spin state, it contains both singlet and triplet spin states. This feature is known as *spin contamination*. For $c = 1$ the UHF wave function is identical to the RHF, and Φ_0^{UHF} is a pure singlet. For $c = 0$ the UHF wave function only contains the covalent terms, which is the correct dissociation behaviour, but also contains equal amounts of singlet and triplet character. When the bond distance is very large, the singlet and triplet states have identical energies, and the spin contamination has no consequence for the energy. In the intermediate region where the bond is not completely broken, however, spin contamination is important. Compared to full CI, the UHF energy is too high as the higher-lying triplet state is mixed into the wave function. The variational principle guarantees that the UHF energy is lower than or equal to the RHF energy (there are more variational parameters). The full CI energy is the lowest possible (for the given basis set) as it recovers 100% of the correlation energy. The UHF

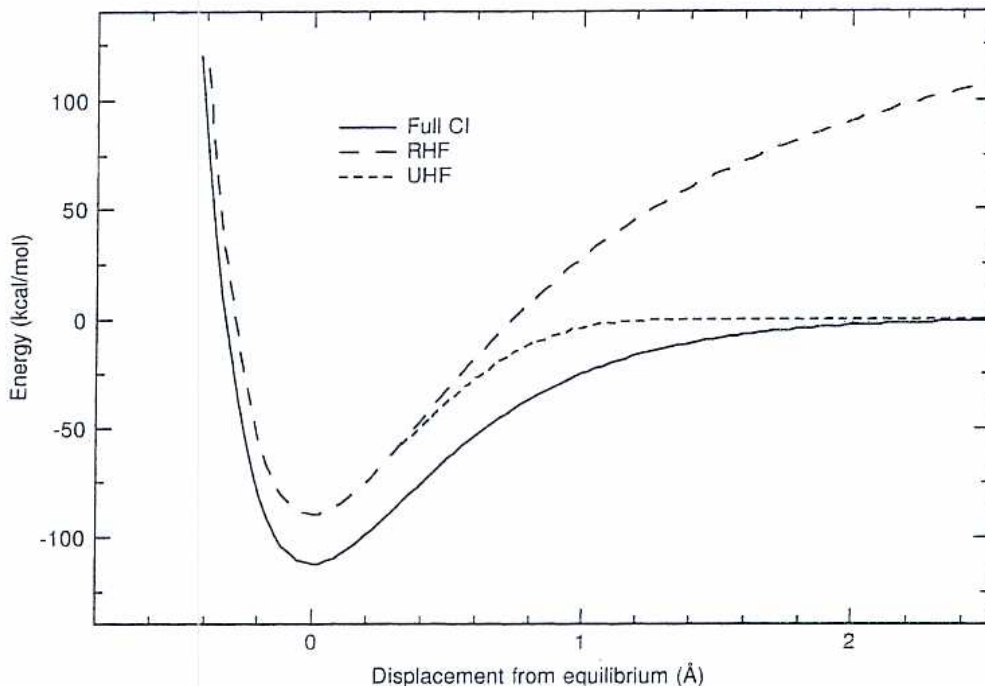


Figure 4.7 Bond dissociation curves for H_2

wave function thus lowers the energy by introducing some electron correlation, but at the same time raises the energy by including higher energy spin states. At the single determinant level (UHF) the variational principle guarantees that the first effect dominates. If the second effect were dominating, the UHF would collapse to the RHF solution. The instability point can thus be viewed as the geometry where the correlation effect becomes larger than the spin contamination. Pictorially the dissociation curves look as shown in Figure 4.7.

Another way of viewing spin contamination is to write the UHF wave function as a linear combination of pure R(O)HF determinants, e.g. for a singlet state.

$${}^1\Phi^{\text{UHF}} = a_1 {}^1\Phi^{\text{RHF}} + a_3 {}^3\Phi^{\text{ROHF}} + a_5 {}^5\Phi^{\text{ROHF}} + \dots \quad (4.25)$$

Since the UHF wave function is multi-determinantal in terms of R(O)HF determinants, it follows that it to some extent includes electron correlation (relative to the RHF reference).

The amount of spin contamination is given by the expectation value of the S^2 operator, $\langle S^2 \rangle$. The theoretical value for a pure spin state is $S_z(S_z + 1)$, i.e. 0 for a singlet ($S_z = 0$), 0.75 for a doublet ($S_z = 1/2$), 2.00 for a triplet ($S_z = 1$) etc. A UHF "singlet" wave function will contain some amounts of triplet, quintet etc. states, increasing the $\langle S^2 \rangle$ value from its theoretical value of zero for a pure spin state. Similarly, a UHF "doublet" wave function will contain some amounts of quartet, sextet etc. states. Usually the contribution from the next higher spin state from the desired is

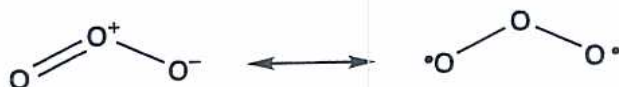


Figure 4.8 Resonance structures for ozone

the most important. The $\langle S^2 \rangle$ value for a UHF wave function is operationally calculated from the spatial overlap between all pairs of α and β spinorbitals

$$\langle S^2 \rangle = S_z(S_z + 1) + N_\beta - \sum_{ij}^{\text{MO}} \langle \phi_i^\alpha | \phi_j^\beta \rangle^2 \quad (4.26)$$

If the α and β orbitals are identical, there is no spin contamination, and the UHF wave function is identical to the RHF.

By including electron correlation in the wave function the UHF method introduces more biradical character into the wave function than RHF. The spin contamination part is also purely biradical in nature, i.e. a UHF treatment in general will overestimate the biradical character. Most singlet states are well described by a closed-shell wave function near the equilibrium geometry, and in those cases it is not possible to generate a UHF solution which has a lower energy than the RHF. There are systems, however, for which this does not hold. An example is the ozone molecule, where two types of resonance structure can be drawn, Figure 4.8.

The biradical resonance structure for ozone requires two singly occupied MOs, and it is clear that an RHF type wave function, which requires all orbitals to be doubly occupied, cannot describe this. A UHF type wave function, however, allows the α and β orbitals to be spatially different, and can to a certain extent incorporate both resonance structures. Systems with biradical character will often have a (singlet) UHF wave function different from an RHF.

As mentioned above, spin contamination in general increases as a bond is stretched. This has some important consequences for transition structures which contain elongated bonds. While most singlet systems have identical RHF and UHF descriptions near the equilibrium geometry, it will normally be possible to find a lower energy UHF solution in the TS region. However, since the spin contamination is not constant along the reaction coordinate, and since the UHF overestimates the biradical character, it is possible that the TS actually becomes a minimum on the UHF energy surface. In other words, the spin contamination may severely distort the shape of the potential energy surface. This may qualitatively be understood by considering the "singlet" UHF wave function as a linear combination of a singlet and a triplet state (eq. (4.25)), as shown in Figure 4.9.

The degree of mixing is determined by the energy difference between the pure singlet and triplet states (as shown for example by second-order perturbation theory, see Section 4.8), which in general decreases along the reaction coordinate. Even if the mixing is not large enough to actually transform a TS to a minimum, it is clear that the UHF energy surface will be much too flat in the TS region. Activation energies calculated at the UHF level will always be lower than the RHF value, but may be either higher or lower than the "correct" value, depending on the amount of spin contamination, since RHF normally overestimates activation energies.

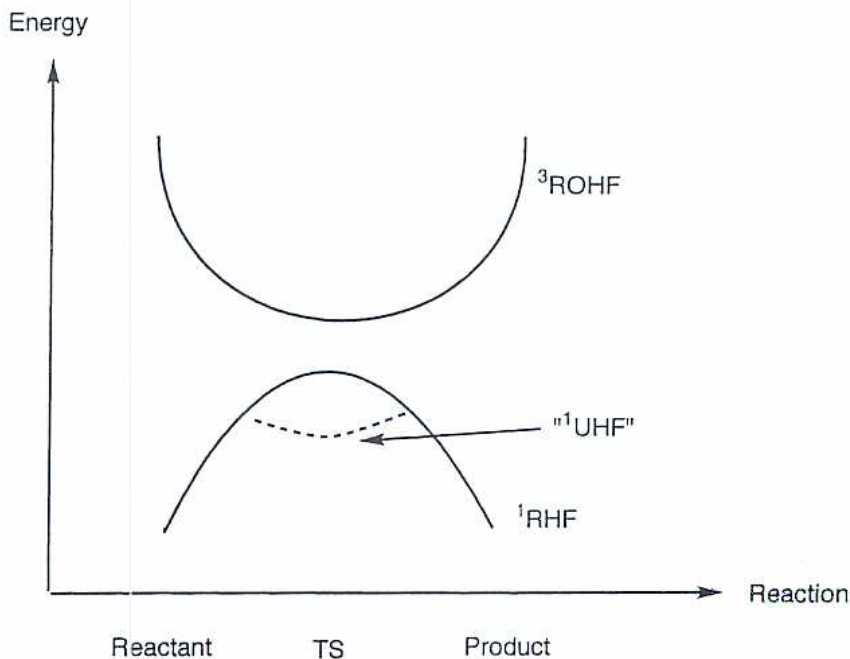


Figure 4.9 Mixing of pure singlet and triplet states may generate artificial minima on the UHF energy surface

From the above it should be clear that UHF wave functions which are spin contaminated (more than a few percent deviation of $\langle S^2 \rangle$ from the theoretical value of $S_z(S_z + 1)$) have disadvantages. For closed-shell systems an RHF procedure is therefore normally preferred. For open-shell systems, however, the UHF method has been heavily used. It is possible to use an ROHF type wave function for open-shell systems, but this leads to computational procedures which are somewhat more complicated than for the UHF case when electron correlation is introduced.

The main problem with the UHF method is the spin contamination, and there have been several proposals on how to remove the unwanted states. There are three strategies which can be considered for removing the contamination: during the SCF procedure, after the SCF has converged, or after electron correlation has been added to the UHF solution. A popular method of removing unwanted states is to project them out with a suitable projection operator (in the picture of the wave function being described in the coordinate system consisting of determinants, the components of the wave function along the higher spin states are removed). As mentioned above, the next higher spin state is usually the most important, and in many cases it is a quite good approximation to only remove this state. After projection, the wave function is then renormalized. If only the first contaminant is removed, this may in extreme cases actually increase the $\langle S^2 \rangle$, value. Performing the projection during the SCF procedure produces a wave function for which it is difficult to formulate a satisfactory theory for including electron correlation by means of perturbation or coupled cluster methods (Sections 4.8 and 4.9). Projections of the converged UHF wave function will lower the energy (although the PUHF energy is no longer variational), since the contributions of the higher-lying states are removed,

and only the correlation effect remains. However, the problems of artificial distortion of the energy surface is even more pronounced at the PUHF level, than with the UHF method itself. For example, it is often found that a false minimum is generated just after the RHF/UHF instability point on a bond dissociation curve. Furthermore, the derivatives of the PUHF energy are not continuous at the RHF/UHF instability point. Projection of the wave function after electron correlation has been added, however, turns out to be a viable pathway. This has mainly been used in connection with perturbation methods, to be described in Section 4.8.2.

4.5 Size Consistency and Size Extensivity

As mentioned above, full CI is impossible, except for very small systems. The only general applicable method is CISD. Consider now a series of CISD calculations in order to construct the interaction potential between two H_2 molecules as a function of the distance between them. Relative to the HF wave function, there will be determinants which correspond to single excitations on only one of the H_2 fragments (S-type determinants), single excitations on both (D-type determinants), and double excitations only on one of the H_2 fragments (also D-type determinants). This will be the case at all intermolecular distances, also when the separation is very large. In that case, however, the system is just two H_2 molecules, and we could consider calculating the energy instead as twice the energy of one H_2 molecule. A CISD calculation on one H_2 molecule would generate singly and doubly excited determinants, and multiplying this by two, would generate determinants which are quadruply excited for the combined H_4 system. A CISD calculation of two H_2 molecules separated by say 100 Å will not give the same energy as twice the results from a CISD calculation on one H_2 molecule (this will be lower). This problem is referred to a *Size Inconsistency*. A very similar, but not identical concept, is *Size Extensivity*. Size consistency is only defined if the two fragments are non-interacting (separated by say 100 Å) while size extensivity implies that the method scales properly with the number of particles, i.e. the fragments can be interacting (separated by say 5 Å). Full CI is size consistent (and extensive), but all forms of truncated CI are not. The lack of size extensivity is the reason why CISD recovers less and less electron correlation as the systems grow larger.

4.6 Multi-configuration Self-consistent Field

The *Multi-configuration Self-consistent Field* (MCSCF) method can be considered as a CI where not only the coefficients in front of the determinants are optimized by the variational principle, but also the MOs used for constructing the determinants are made optimum.⁵ The MCSCF optimization is iterative just like the SCF procedure (if the "multi-configuration" is only one, it is simply HF). Since the number of MCSCF iterations required for achieving convergence tends to increase with the number of configurations included, the size of MCSCF wave function that can be treated is somewhat smaller than for CI methods.

When deriving the Hartree–Fock equations it was only required that the variation of the energy with respect to an orbital variation should be zero. This is equivalent to the first derivatives of the energy with respect to the MO expansion coefficients being equal to zero. The Hartree–Fock equations can be solved by an iterative SCF method, and

there are many techniques for helping the iterative procedure to converge (Section 3.8). There is, however, no guarantee that the solution found by the SCF procedure is a minimum of the energy as a function of the MO coefficients. To determine that, the matrix of second derivatives of the energy with respect to the MO coefficients can be calculated and diagonalized. In order to be a minimum it should have all positive eigenvalues. In practice only the lowest eigenvalue is normally determined, if this is positive the solution is a minimum. This is rarely checked for SCF wave functions, in the large majority of cases the SCF procedure converges to a minimum without problems. MCSCF wave functions, on the other hand, are much harder to make converge, and much more prone to converge on solutions which are not minima. MCSCF wave function optimizations are therefore normally carried out by expanding the energy to second order in the variational parameters (orbital and configurational coefficients), analogously to the second-order SCF procedure described in Section 3.8.1, and using Newton–Raphson based methods described in Chapter 14 to force convergence to a minimum.

MCSCF methods are rarely used for calculating large fractions of the correlation energy. The orbital relaxation usually does not recover much electron correlation, it is more efficient to include additional determinants and keep the MOs fixed (CI) if the interest is just in obtaining a large fraction of the correlation energy. Single determinant HF wave functions normally give a qualitatively correct description of the electron structure; however, there are many examples where this is not the case. MCSCF methods can be considered as an extension of single determinant methods that gives a qualitatively correct description. Consider again the ozone molecule with the two resonance structures shown in Figure 4.8. Each type of resonance structure essentially translates into a different determinant. If more than one non-equivalent resonance structure is important, this means that the wave function cannot be described even qualitatively correctly at the RHF single determinant level (benzene, for example, has two equivalent cyclohexatriene resonance structures, and is adequately described by an RHF wave function). A UHF wave function allows some biradical character, with the disadvantages mentioned above. Alternatively, a second restricted type CSF (consisting of two determinants) with two singly occupied MOs may be included in the wave function. The simplest MCSCF for ozone will contain two configurations (often denoted TCSCF), with the optimum MOs and configurational weights determined by the variational principle. The CSFs entering an MCSCF expansion are pure spin states, and MCSCF wave functions therefore do not suffer from the problem of spin contamination.

Our definition of electron correlation uses the RHF energy as the reference. For ozone both the UHF and the TCSCF wave functions have lower energies, and include some electron correlation. This type of “electron correlation” is somewhat different from the picture presented at the start of this chapter. In a sense it is a consequence of our chosen zero point for the correlation energy, the RHF energy. The energy lowering introduced by adding enough flexibility in the wave function to be able to qualitatively describe the system is sometimes called the *static* electron correlation. This is essentially the effect of allowing orbitals to become (partly) singly occupied instead of forcing double occupation, i.e. describing near-degeneracy effects (two or more configurations having almost the same energy). The remaining energy lowering by correlating the motion of the electrons is called *dynamic* correlation. The problem is that there is no rigorous way of separating these effects. In the ozone example the energy lowering by going from

RHF to UHF, or to a TCSCF, is almost pure static correlation. Increasing the number of configurations in an MCSCF will recover more and more of the dynamical correlation, until at the full CI limit, the correlation treatment is exact. As mentioned above, MCSCF methods are mainly used for generating a qualitatively correct wave function, i.e. recovering the "static" part of the correlation.

The major problem with MCSCF methods is selecting the necessary configurations to include for the property of interest. One of the most popular approaches is the *Complete Active Space Self-consistent Field* (CASSCF) method (also called *Full Optimized Reaction Space*, FORS). Here the selection of configurations is done by partitioning the MOs into *active* and *inactive* spaces. The active MOs will typically be some of the highest occupied and some of the lowest unoccupied MOs from a RHF calculation. The inactive MOs either have 2 or 0 electrons, i.e. always doubly occupied or empty. Within the active MOs a full CI is performed and all the proper symmetry adapted configurations are included in the MCSCF optimization. Which MOs to include in the active space must be decided manually, by considering the problem at hand and the computational expense. If several points on the potential energy surface are desired, the MCSCF active space should include all those orbitals which change significantly, or for which the electron correlation is expected to change. A common notation is $[n, m]$ -CASSCF, indicating that n electrons are distributed in all possible ways in m orbitals.

As for any full CI expansion, the CASSCF becomes unmanageably large even for quite small active spaces. A variation of the CASSCF procedure is the *Restricted Active Space Self-consistent Field* (RASSCF) method.⁶ Here the active MOs are further divided into three sections, RAS1, RAS2 and RAS3, each having restrictions on the occupation numbers (excitations) allowed. The RAS1 space consists of MOs which are doubly occupied in the HF reference determinant, the RAS2 has both occupied and unoccupied orbitals, while the RAS3 space consists of MOs which are empty in the HF determinant. Configurations in the RAS2 space are generated by a full CI, analogous to

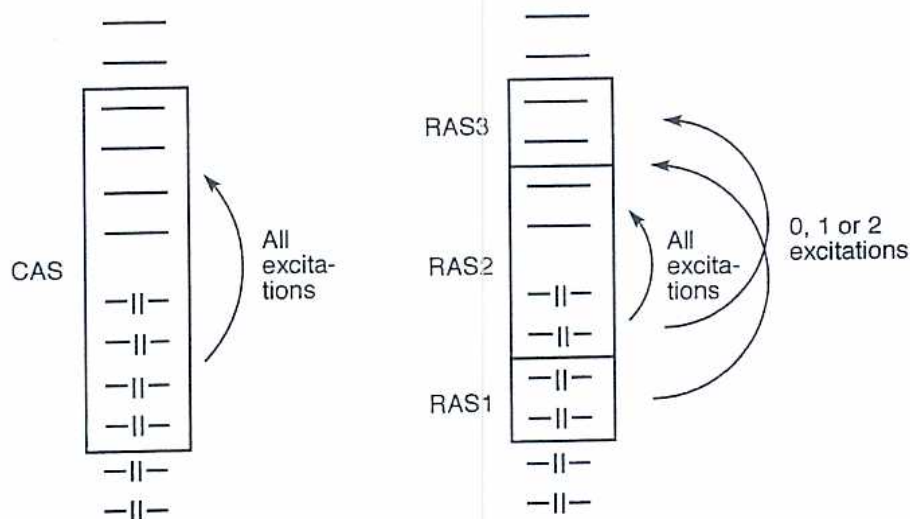


Figure 4.10 Illustration of the CAS and RAS orbital partitions

Table 4.3 Number of configurations generated in a $[n, n]$ -CASSCF wave function

n	Number of CSFs
2	3
4	20
6	175
8	1 764
10	19 404
12	226 512
14	2 760 615

the CASSCF approach, and additional configurations are generated by allowing for example a maximum of two electrons to be excited from the RAS1 and a maximum of two electrons to be excited to the RAS3 space. In essence, a typical RASSCF procedure thus generates configurations by a combination of a full CI in a small number of MOs (RAS2) and a CISD in a somewhat larger MO space (RAS1 and RAS3).

The full CI expansion within the active space severely restricts the number of orbitals and electrons that can be treated by CASSCF methods. Table 4.3 shows how many singlet CSFs are generated for an $[n, n]$ -CASSCF wave function (eq. (4.13)), without any reductions arising from symmetry.

The factorial increase in the number of CSFs effectively limits the active space for CASSCF wave functions to less than 10–12 electrons/orbitals. Selecting the “important” orbitals to correlate therefore becomes very important. The goal of MCSCF methods is usually not to recover a large fraction of the total correlation energy, but rather to recover all the changes that occur in the correlation energy for the given process. Selecting the active space for an MCSCF calculation requires some insight into the problem. There are a few rules of thumb that may be of help in selecting a proper set of orbitals for the active space.

- (1) For each occupied orbital, there will typically be one corresponding virtual orbital. This leads naturally to $[n, m]$ -CASSCF wave functions where n and m are identical or nearly so.
- (2) The orbital energies from an RHF calculation may be used for selecting the important orbitals. The highest occupied and lowest unoccupied are usually the most important orbitals to include in the active space. This can be partly justified by the formula for the second-order perturbation energy correction (Section 4.8.1), the smaller the orbital energy difference, the larger the contribution to the correlation energy. Using RHF orbital energies for selecting the active space may be problematic in two situations. The first is when extended basis sets are used, where there will be many virtual orbitals with low energies, and the exact order is more or less accidental. Furthermore, RHF virtual orbitals basically describe electron attachment (via Koopmans’ theorem, Section 3.4), and are therefore not particularly well suited for describing electron correlation. An inspection of the form of the orbitals may reveal which to choose, they should be the ones which resemble the

occupied orbitals in terms of basis function contribution. The second problem is more fundamental. If the real wave function has significant multi-configurational character, then the RHF may be qualitatively wrong, and selecting the active orbitals based on a qualitatively wrong function may lead to erroneous results. The problem is that we wish to include the important orbitals for describing the multi-determinant nature, but these are not known until the final wave function is known.

- (3) A way of overcoming this problem, is to use the concept of natural orbitals. The natural orbitals are those which diagonalize the density matrix, and the eigenvalues are the occupation numbers. Orbitals with occupation numbers significantly different from 0 or 2 (for a closed-shell system) are usually those which are the most important to include in the active space. An RHF wave function will have occupation numbers of exactly 0 or 2, and some electron correlation must be included to obtain orbitals with non-integer occupation numbers. This may for example be done by running a preliminary MP2 or CISD calculation prior to the MCSCF. Alternatively, a UHF (when different from the RHF) type wave function may also be used. The total UHF density, which is the sum of the α and β density matrices, will also provide fractional occupation numbers since UHF includes some electron correlation. The procedure may still fail. If the underlying RHF wave function is poor, the MP2 correction may also give poor results, and selecting the active MCSCF orbitals based on such MP2 occupation numbers may again lead to erroneous results. In practice, however, selecting active orbitals based on for example MP2 occupation numbers appears to be quite efficient, and better than using RHF orbital energies.

In a CASSCF type wave function the CI coefficients do not have the same significance as for a single reference CI based on HF orbitals. In a full CI (as in the active space of the CASSCF) the orbitals may be rotated among themselves without affecting the total wave function. A rotation of the orbitals, however, influences the magnitude of the coefficients in front of each CSF. While the HF coefficient in a single reference CISD gives some indication of the "multi-reference" nature of the wave function, this is not the case for a CASSCF wave function, where the corresponding CI coefficient is arbitrary.

It should be noted that CASSCF methods inherently tend to give an unbalanced description, since all the electron correlation recovered is in the active space, but none in the inactive space, or between the active and inactive electrons.⁷ This is not a problem if all the valence electrons are included in the active space, but this is only possible for small systems. If only part of the valence electrons are included in the active space, the CASSCF methods tend to overestimate the importance of "biradical" structures. Consider for example acetylene where the hydrogens have been bent 60° away from linearity (this may be considered a model for *ortho*-benzyne). The in-plane " π -orbital" now acquires significant biradical character. The true structure may be described as a linear combination of the three configurations shown in Figure 4.11.

The structure on the left is biradical, while the two others are ionic, corresponding to both electrons being at the same carbon. The simplest CASSCF wave function which qualitatively can describe this system has two electrons in two orbitals, giving the three configurations shown above. The dynamical correlation between the two active electrons will tend to keep them as far apart as possible, i.e. favouring the biradical structure. Now

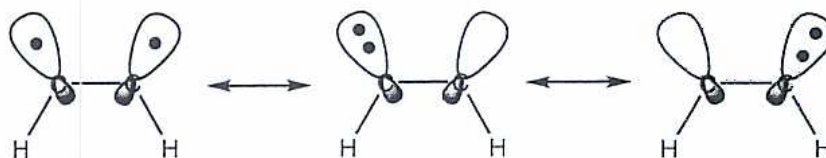


Figure 4.11 Important configurations for a bent acetylene model

Table 4.4 Natural orbital occupation numbers for the distorted acetylene model in Figure 4.11. Only the occupation numbers for the six “central” orbitals are shown

	n_5	n_6	n_7	n_8	n_9	n_{10}
RHF	2.00	2.00	2.00	0.00	0.00	0.00
UHF	2.00	1.72	1.30	0.70	0.28	0.01
[2,2]-CASSCF	2.00	2.00	1.62	0.38	0.00	0.00
[4,4]-CASSCF	2.00	1.85	1.67	0.33	0.14	0.00
[10,10]-CASSCF	1.97	1.87	1.71	0.30	0.13	0.02

consider a full valence CASSCF wave function with 10 electrons in 10 orbitals. This will analogously tend to separate the two electrons in each bond with one being at each end. The correlation of the electrons in the C–H bonds, for example, will place more electron density on the carbon atoms. This in turn will favour the ionic structures in Figure 4.11 and disfavour the biradical, i.e. the dynamical correlation of the other electrons may take advantage of the empty orbital in the ionic structures but not in the biradical. These general considerations may be quantified by considering the natural orbital occupancies for increasingly larger CASSCF wave functions, as shown in Table 4.4 for the 6-31G(d,p) basis. The [4,4]-CASSCF also includes the two out-of-plane π -orbitals in the active space, while the [10,10]-CASSCF generates a full-valence CI wave function. The unbalanced description for the [2,2]-CASSCF is reminiscent of the spin contamination problem for UHF wave functions, although the effect is much less pronounced. Nevertheless, the overestimation may be severe enough to alter the qualitative shape of energy surfaces, for example turning transition structures into minima, as illustrated in Figure 4.9. MCSCF methods are therefore not “black box” methods like for example HF and MP (Section 4.8.1); selecting a proper number of configurations, and the correct orbitals, to give a balanced description of the problem at hand requires some experimentation and insight.

4.7 Multi-reference Configuration Interaction

The CI methods described so far consider only CSFs generated by exciting electrons from a single determinant. This corresponds to having a HF type wave function as the reference. However, a MCSCF wave function may also be chosen as the reference. In that case a CISD involves excitations of one or two electrons out of all the determinants which enter the MCSCF, defining the *Multi-reference Configuration Interaction* (MRCI) method. Compared to the single reference CISD, the number of configurations is

increased by a factor roughly equal to the number of configurations included in the MCSCF. Large-scale MRCI wave functions (many configurations in the MCSCF) can generate very accurate wave functions, but are also computationally very intensive. Since MRCI methods truncate the CI expansion, they are not size extensive.

Even truncating the (MR)CI expansion at the singles and doubles level frequently generates more configurations than can be handled readily. A further truncation is sometimes performed by selecting only those configurations which have an "interaction" with the reference configuration(s) above a selected threshold, where the "interaction" is evaluated by second-order perturbation theory (Section 4.8). Such *state selected* CI (or MCSCF) methods all involve a preset cutoff below which configurations are neglected. This may cause problems for comparing energies of different geometries, since the potential energy surface may become discontinuous, i.e. at some point the importance of a given configuration may drop below the threshold, and the contribution suddenly disappear.

4.8 Many-body Perturbation Theory

The idea in perturbation methods is that the problem at hand only differs slightly from a problem which has already been solved (exactly or approximately). The solution to the given problem should therefore in some sense be close to the solution of the already known system. This is described mathematically by defining a Hamilton operator which consists of two parts, a reference (\mathbf{H}_0) and a perturbation (\mathbf{H}'). The premise of perturbation methods is that the \mathbf{H}' operator in some sense is "small" compared to \mathbf{H}_0 . In quantum mechanics, perturbational methods can be used for adding corrections to solutions which employ an independent particle approximation, and the theoretical framework is then called *Many-Body Perturbation Theory* (MBPT).

Let us assume that the Schrödinger equation for the reference Hamilton operator is solved.

$$\begin{aligned}\mathbf{H} &= \mathbf{H}_0 + \lambda\mathbf{H}' \\ \mathbf{H}_0\Phi_i &= E_i\Phi_i, \quad i = 0, 1, 2, \dots, \infty\end{aligned}\quad (4.27)$$

The solutions for the unperturbed Hamilton operator from a complete set (since \mathbf{H}_0 is hermitian) which can be chosen to be orthonormal, and λ is a (variable) parameter determining the strength of the perturbation. At present we will only consider cases where the perturbation is time-independent, and the reference wave function is non-degenerate. To keep the notation simple, we will furthermore only consider the lowest energy state. The perturbed Schrödinger equation is

$$\mathbf{H}\Psi = W\Psi \quad (4.28)$$

If $\lambda = 0$, then $\mathbf{H} = \mathbf{H}_0$, $\Psi = \Phi_0$ and $W = E_0$. As the perturbation is increased from zero to a finite value, the new energy and wave function must also change continuously, and they can be written as a Taylor expansion in powers of the perturbation parameter λ .

$$\begin{aligned}W &= \lambda^0 W_0 + \lambda^1 W_1 + \lambda^2 W_2 + \lambda^3 W_3 + \dots \\ \Psi &= \lambda^0 \Psi_0 + \lambda^1 \Psi_1 + \lambda^2 \Psi_2 + \lambda^3 \Psi_3 + \dots\end{aligned}\quad (4.29)$$

For $\lambda = 0$, it is seen that $\Psi_0 = \Phi_0$ and $W_0 = E_0$, these are the *unperturbed*, or *zero-*

order wave function and energy. The $\Psi_1, \Psi_2 \dots$ and $W_1, W_2 \dots$ are the *first-, second- etc. order corrections*. The λ parameter will eventually be set equal to 1, and the n th order energy or wave function become a sum of all terms up to order n . It is convenient to chose the perturbed wave function to be *intermediately normalized*, i.e. the overlap with the unperturbed wave function should be 1. This has the consequence that all correction terms are orthogonal to the reference wave function.

$$\begin{aligned} \langle \Psi | \Phi_0 \rangle &= 1 \\ \langle \Psi_0 + \lambda \Psi_1 + \lambda^2 \Psi_2 + \dots | \Phi_0 \rangle &= 1 \\ \langle \Psi_0 | \Phi_0 \rangle + \lambda \langle \Psi_1 | \Phi_0 \rangle + \lambda^2 \langle \Psi_2 | \Phi_0 \rangle + \dots &= 1 \\ \langle \Psi_{i \neq 0} | \Phi_0 \rangle &= 0 \end{aligned} \quad (4.30)$$

Once all the correction terms have been calculated, it is trivial to normalize the total wave function.

With the expansions (4.29) the Schrödinger equation (4.28) becomes

$$\begin{aligned} (\mathbf{H}_0 + \lambda \mathbf{H}')(\lambda^0 \Psi_0 + \lambda^1 \Psi_1 + \lambda^2 \Psi_2 + \dots) &= \\ (\lambda^0 W_0 + \lambda^1 W_1 + \lambda^2 W_2 + \dots)(\lambda^0 \Psi_0 + \lambda^1 \Psi_1 + \lambda^2 \Psi_2 + \dots) \end{aligned} \quad (4.31)$$

Since this holds for any value of λ , we can collect terms with the same power of λ to give

$$\begin{aligned} \lambda^0 : \mathbf{H}_0 \Psi_0 &= W_0 \Psi_0 \\ \lambda^1 : \mathbf{H}_0 \Psi_1 + \mathbf{H}' \Psi_0 &= W_0 \Psi_1 + W_1 \Psi_0 \\ \lambda^2 : \mathbf{H}_0 \Psi_2 + \mathbf{H}' \Psi_1 &= W_0 \Psi_2 + W_1 \Psi_1 + W_2 \Psi_0 \\ \lambda^n : \mathbf{H}_0 \Psi_n + \mathbf{H}' \Psi_{n-1} &= \sum_{i=0}^n W_i \Psi_{n-i} \end{aligned} \quad (4.32)$$

These are zero-, first-, second-, n th-order perturbation equations. The zero-order equation is just the Schrödinger equation for the unperturbed problem. The first-order equation contains two unknowns, the first-order correction to the energy, W_1 , and the first-order correction to the wave function, Ψ_1 . The n th-order energy correction can be calculated by multiplying from the left by Φ_0 and integrating, and using the “turnover rule” $\langle \Phi_0 | \mathbf{H}_0 | \Psi_i \rangle = \langle \Psi_i | \mathbf{H}_0 | \Phi_0 \rangle^*$.

$$\begin{aligned} \langle \Phi_0 | \mathbf{H}_0 | \Psi_n \rangle + \langle \Phi_0 | \mathbf{H}' | \Psi_{n-1} \rangle &= \sum_{i=0}^{n-1} W_i \langle \Phi_0 | \Psi_{n-i} \rangle + W_n \langle \Phi_0 | \Psi_0 \rangle \\ E_0 \langle \Phi_0 | \Psi_n \rangle + \langle \Phi_0 | \mathbf{H}' | \Psi_{n-1} \rangle &= W_n \langle \Phi_0 | \Psi_0 \rangle \\ W_n &= \langle \Phi_0 | \mathbf{H}' | \Psi_{n-1} \rangle \end{aligned} \quad (4.33)$$

From this it would appear that the $(n-1)$ th-order wave function is required for calculating the n th-order energy. However, by using the turnover rule and the n th and lower-order perturbation equations (4.32), it can be shown that knowledge of the n th-order wave function actually allows a calculation of the $(2n+1)$ th-order energy.

$$W_{2n+1} = \langle \Psi_n | \mathbf{H}' | \Psi_n \rangle - \sum_{k,l=1}^n W_{2n+1-k-1} \langle \Psi_k | \Psi_l \rangle \quad (4.34)$$

Up to this point we are still dealing with undetermined quantities, energy and wave function corrections at each order. The first-order equation is one equation with two unknowns. Since the solutions to the unperturbed Schrödinger equation generates a complete set of functions, the unknown first-order correction to the wave function can be expanded in these functions. This is known as *Rayleigh–Schrödinger* perturbation theory, and the λ^1 equation in (4.32) becomes

$$\begin{aligned}\Psi_1 &= \sum_i c_i \Phi_i \\ (\mathbf{H}_0 - W_0) \left(\sum_i c_i \Phi_i \right) + (\mathbf{H}' - W_1) \Phi_0 &= 0\end{aligned}\quad (4.35)$$

Multiplying from the left by Φ_0^* and integrating yields

$$\begin{aligned}\sum_i c_i \langle \Phi_0 | \mathbf{H}_0 | \Phi_i \rangle - W_0 \sum_i c_i \langle \Phi_0 | \Phi_i \rangle + \langle \Phi_0 | \mathbf{H}' | \Phi_0 \rangle - W_1 \langle \Phi_0 | \Phi_0 \rangle &= 0 \\ \sum_i c_i E_i \langle \Phi_0 | \Phi_i \rangle - c_0 E_0 + \langle \Phi_0 | \mathbf{H}' | \Phi_0 \rangle - W_1 &= 0 \\ c_0 E_0 - c_0 E_0 + \langle \Phi_0 | \mathbf{H}' | \Phi_0 \rangle - W_1 &= 0 \\ W_1 &= \langle \Phi_0 | \mathbf{H}' | \Phi_0 \rangle\end{aligned}\quad (4.36)$$

since the Φ_i s are orthonormal (this also follows directly from eq. (4.34)). The latter equation shows that the first-order correction to the energy is an average of the perturbation operator over the unperturbed wave function.

The first-order correction to the wave function can be obtained by multiplying (4.32) from the left by a function other than Φ_0 (Φ_j) and integrating to give

$$\begin{aligned}\sum_i c_i \langle \Phi_j | \mathbf{H}_0 | \Phi_i \rangle - W_0 \sum_i c_i \langle \Phi_j | \Phi_i \rangle + \langle \Phi_j | \mathbf{H}' | \Phi_0 \rangle - W_1 \langle \Phi_j | \Phi_0 \rangle &= 0 \\ \sum_i c_i E_i \langle \Phi_j | \Phi_i \rangle - c_j E_0 + \langle \Phi_j | \mathbf{H}' | \Phi_0 \rangle &= 0 \\ c_j E_j - c_j E_0 + \langle \Phi_j | \mathbf{H}' | \Phi_0 \rangle &= 0 \\ c_j &= \frac{\langle \Phi_j | \mathbf{H}' | \Phi_0 \rangle}{E_0 - E_j}\end{aligned}\quad (4.37)$$

The expansion coefficients determine the first-order correction to the perturbed wave function (eq. (4.35)), and they can be calculated for the known unperturbed wave functions and energies. The coefficient in front of Φ_0 for Ψ_1 cannot be determined from the above formula, but the assumption of intermediate normalization (eq. (4.30)) makes $c_0 = 0$.

Starting from the second-order perturbation equation (4.32), analogous formulas can be generated for the second-order corrections. Using intermediate normalization

($c_0 = d_0 = 0$), the second-order energy correction is

$$\begin{aligned} \Psi_2 &= \sum_i d_i \Phi_i \\ (\mathbf{H}_0 - W_0) \left(\sum_i d_i \Phi_i \right) + (\mathbf{H}' - W_1) \left(\sum_i c_i \Phi_i \right) - W_2 \Phi_0 &= 0 \\ \sum_i d_i \langle \Phi_0 | \mathbf{H}_0 | \Phi_i \rangle - W_0 \sum_i d_i \langle \Phi_0 | \Phi_i \rangle & \\ + \sum_i c_i \langle \Phi_0 | \mathbf{H}' | \Phi_i \rangle - W_1 \sum_i c_i \langle \Phi_0 | \Phi_i \rangle - W_2 \langle \Phi_0 | \Phi_0 \rangle &= 0 \\ \sum_i d_i E_i \langle \Phi_0 | \Phi_i \rangle - d_0 E_0 + \sum_i c_i \langle \Phi_0 | \mathbf{H}' | \Phi_i \rangle & \\ - c_0 W_1 - W_2 &= 0 \\ d_0 E_0 - d_0 E_0 + \sum_i c_i \langle \Phi_0 | \mathbf{H}' | \Phi_i \rangle - W_2 &= 0 \\ W_2 = \sum_i c_i \langle \Phi_0 | \mathbf{H}' | \Phi_i \rangle = \sum_{i \neq 0} \frac{\langle \Phi_0 | \mathbf{H}' | \Phi_i \rangle \langle \Phi_i | \mathbf{H}' | \Phi_0 \rangle}{E_0 - E_i} & \end{aligned} \quad (4.38)$$

The last equation shows that the second-order energy correction may be written in terms of the first order wave function (c_i) and matrix elements over unperturbed states. The second-order wave function correction is

$$\begin{aligned} \sum_i d_i \langle \Phi_j | \mathbf{H}_0 | \Phi_i \rangle - W_0 \sum_i d_i \langle \Phi_j | \Phi_i \rangle & \\ + \sum_i c_i \langle \Phi_j | \mathbf{H}' | \Phi_i \rangle - W_1 \sum_i c_i \langle \Phi_j | \Phi_i \rangle - W_2 \langle \Phi_j | \Phi_0 \rangle &= 0 \\ \sum_i d_i E_i \langle \Phi_j | \Phi_i \rangle - d_j E_0 + \sum_i c_i \langle \Phi_j | \mathbf{H}' | \Phi_i \rangle - c_j W_1 &= 0 \\ d_j E_j - d_j E_0 + \sum_i c_i \langle \Phi_j | \mathbf{H}' | \Phi_i \rangle - c_j \langle \Phi_0 | \mathbf{H}' | \Phi_0 \rangle &= 0 \\ d_j = \sum_{i \neq 0} \frac{\langle \Phi_j | \mathbf{H}' | \Phi_i \rangle \langle \Phi_i | \mathbf{H}' | \Phi_0 \rangle}{(E_0 - E_j)(E_0 - E_i)} - \frac{\langle \Phi_j | \mathbf{H}' | \Phi_0 \rangle \langle \Phi_0 | \mathbf{H}' | \Phi_0 \rangle}{(E_0 - E_j)^2} & \end{aligned} \quad (4.39)$$

The formulas for higher-order corrections become increasingly complex. The main point, however, is that all corrections can be expressed in terms of matrix elements of the perturbation operator over the unperturbed wave functions, and the unperturbed energies.

4.8.1 Møller–Plesset Perturbation Theory

So far the theory has been completely general. In order to apply perturbation theory to the calculation of correlation energy, the unperturbed Hamilton operator must be selected. The most common choice is to take this as a sum over Fock operators, leading to Møller–Plesset (MP) perturbation theory.⁸ The sum of Fock operators counts the (average) electron–electron repulsion twice (eq. (3.43)), and the perturbation becomes

the exact V_{ee} operator minus twice the $\langle V_{ee} \rangle$ operator. The operator associated with this difference is often referred to as the *fluctuation potential*. This choice is not really consistent with the basic assumption that the perturbation should be small compared to H_0 . However, it does fulfill the other requirement that solutions to the unperturbed Schrödinger equation should be known. Furthermore, this is the only choice which leads to a size extensive method, which is a desirable feature.

$$\begin{aligned} H_0 &= \sum_{i=1}^N F_i = \sum_{i=1}^N \left(h_i + \sum_{j=1}^N (J_{ij} - K_{ij}) \right) = \sum_{i=1}^N h_i + 2\langle V_{ee} \rangle \\ &= \sum_{i=1}^N h_i + \sum_{i=1}^N \sum_{j=1}^N \langle g_{ij} \rangle \end{aligned} \quad (4.40)$$

$$\begin{aligned} H' &= H - H_0 = V_{ee} - \sum_{i=1}^N \sum_{j=1}^N (J_{ij} - K_{ij}) = V_{ee} - 2\langle V_{ee} \rangle \\ &= \sum_{i=1}^N \sum_{j>i}^N g_{ij} - \sum_{i=1}^N \sum_{j=1}^N \langle g_{ij} \rangle \end{aligned}$$

The zero-order wave function is the HF determinant, and the zero-order energy is just a sum of MO energies. The first-order energy correction is the average of the perturbation operator over the zero-order wave function (eq. (4.36)).

$$W_1 = \langle \Phi_0 | H' | \Phi_0 \rangle = \langle V_{ee} \rangle - 2\langle V_{ee} \rangle = -\langle V_{ee} \rangle \quad (4.41)$$

This yields a correction for the overcounting of the electron–electron repulsion at zero-order. Comparing eq. (4.40) with the expression for the total energy in eq. (3.32), it is seen that the first-order energy (sum of W_0 and W_1) is exactly the HF energy. Using the notation $E(MPn)$ to indicate the correction at order n , and MPn to indicate the total energy up to order n , we have

$$MP0 = E(MP0) = \sum_{i=1}^N \varepsilon_i \quad (4.42)$$

$$MP1 = MP0 + E(MP1) = E(\text{HF})$$

Electron correlation energy thus starts at order 2 with this choice of H_0 .

In developing perturbation theory it was assumed that the solutions to the unperturbed problem formed a complete set. This is general means that there must be an infinite number of functions, which is impossible in actual calculations. The lowest energy solution to the unperturbed problem is the HF wave function, additional higher energy solutions are excited Slater determinants, analogously to the CI method. When a finite basis set is employed it is only possible to generate a finite number of excited determinants. The expansion of the many-electron wave function is therefore truncated.

Let us look at the expression for the second-order energy correction, eq. (4.38). This involves matrix elements of the perturbation operator between the HF reference and all possible excited states. Since the perturbation is a two-electron operator, all matrix elements involving triple, quadruple etc. excitations are zero. When canonical HF

orbitals are used, matrix elements with singly excited states are also zero, which follows from.

$$\begin{aligned}
 \langle \Phi_0 | \mathbf{H}' | \Phi_i^a \rangle &= \langle \Phi_0 | \mathbf{H} - \sum_{j=1}^N \mathbf{F}_j | \Phi_i^a \rangle \\
 &= \langle \Phi_0 | \mathbf{H} | \Phi_i^a \rangle - \sum_{j=1}^N \langle \Phi_0 | \mathbf{F}_j | \Phi_i^a \rangle \\
 &= \langle \Phi_0 | \mathbf{H} | \Phi_i^a \rangle - \varepsilon_a \langle \Phi_0 | \Phi_i^a \rangle
 \end{aligned}
 \tag{4.43}$$

The first bracket is zero due to Brillouin's theorem (Section 4.2.1), and the second set of brackets is zero due to the orbitals being eigenfunctions of the Fock operators and being orthogonal to each other. The second-order correction to the energy, which is the first contribution to the correlation energy, therefore only involves a sum over doubly excited determinants. These can be generated by promoting two electrons from occupied orbitals i and j to virtual orbitals a and b . The summation must be restricted so that each excited state is only counted once.

$$W_2 = \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} \frac{\langle \Phi_0 | \mathbf{H}' | \Phi_{ij}^{ab} \rangle \langle \Phi_{ij}^{ab} | \mathbf{H}' | \Phi_0 \rangle}{E_0 - E_{ij}^{ab}}
 \tag{4.44}$$

The matrix elements between the HF and a doubly excited state are given by two-electron integrals over MOs (eq. (4.7)). The difference in total energy between two Slater determinants becomes a difference in MO energies (essentially Koopmans' theorem), and the explicit formula for the second-order Møller–Plesset correction is

$$E(\text{MP2}) = \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} \frac{[\langle \phi_i \phi_j | \phi_a \phi_b \rangle - \langle \phi_i \phi_j | \phi_b \phi_a \rangle]^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}
 \tag{4.45}$$

Once the two-electron integrals over MOs are available, the second-order energy correction can be calculated as a sum over such integrals. There are of the order of M^4 integrals, thus the calculation of the energy (only) increases as M^4 with the system size. However, the transformation of the integrals from the AO to the MO basis grows as M^5 (Section 4.2.1). MP2 is an M^5 method, but fairly inexpensive as not all two-electron integrals over MOs are required. Only those corresponding to the combination of two occupied and two virtual MOs are needed. In practical calculations this means that the MP2 energy for systems with ~ 100 – 150 basis functions can be calculated at a cost similar to or less than what is required for calculating the HF energy. MP2 typically accounts for ~ 80 – 90% of the correlation energy, and it is the most economical method for including electron correlation.

The formula for the first-order correction to the wave function (eq. (4.37)) similarly only contains contributions from doubly excited determinants. Since knowledge of the first-order wave function allows calculation of the energy up to third order ($2n + 1 = 3$, eq. (4.34)), it is immediately clear that the third-order energy also only contains contributions from doubly excited determinants. Qualitatively speaking, the MP2 contribution describes the correlation between pairs of electrons while MP3 describes the interaction between pairs. The formula for calculating this contribution is somewhat

more complex than for second order, and involves a computational effort which formally increases as M^6 . The third-order energy typically accounts for $\sim 90-95\%$ of the correlation energy.

The formula for the second-order correction to the wave function (eq. (4.39)) contains products of the type $\langle \Phi_j | \mathbf{H}' | \Phi_i \rangle \langle \Phi_i | \mathbf{H}' | \Phi_0 \rangle$. The Φ_0 is the HF determinant and the last bracket can only be non-zero if Φ_i is a doubly excited determinant. This means that the first bracket only can be non-zero if Φ_j is either a singly, doubly, triply or quadruply excited determinant (\mathbf{H}' is a two-electron operator). The second-order wave function allows calculation of the fourth- and fifth-order energies, these terms therefore have contributions from determinants which are singly, doubly, triply or quadruply excited. The computational cost of the fourth-order energy without the contribution from the triply excited determinants, MP4(SDQ), increases as M^6 , while the triples contribution increases as M^7 . MP4 is still a computationally feasible model for many molecular systems, requiring a time similar to CISD. In typical calculations the T contribution to MP4 will take roughly the same amount of time as the SDQ contributions, but the triples are often the most important at fourth order. The full fourth-order energy typically accounts for $\sim 95-98\%$ of the correlation energy.

The fifth-order correction to the energy also involve S, D, T and Q contributions, and the sixth-order term introduces quintuple and sextuple excitations. The working formulas for the MP5 and MP6 contributions are now so complex that actual calculations are only possible for small systems. The computational effort for MP5 increases as M^8 and for MP6 as M^9 . There is very little experience with the performance of MP_n beyond MP4.

As shown in Table 4.2, the most important contribution to the energy in a CI procedure comes from doubly excited determinants. This is also shown by the perturbation expansion, the second- and third-order energy corrections only involve doubles. At fourth order the singles, triples and quadruples enter the expansion for the first time. This is again consistent with Table 4.2, which shows that these types of excitation are of similar importance.

CI methods determine the energy by a variational procedure, and the energy is consequently an upper bound to the exact energy. There is no such guarantee for perturbation methods; it is possible that the energy will be lower than the exact energy. This is rarely a problem. Limitations in the basis set often mean that the error in total energy is several a.u. (thousands of kcal/mol) anyway. Furthermore, the interest is normally not in total energies, but in energy differences. Having a variational upper bound for two energies does not give any bound for the difference between these two numbers. The main interest is therefore in the error remaining relatively constant for different systems. The lack of size extensivity of CI methods is disadvantageous in this respect. The MP perturbation method is size extensive, but other forms of MBPT are not. It is now generally recognized that size extensivity is an important property, and the MP form of MBPT is used almost exclusively. Combined with the low cost relative to CI methods this often makes MP calculations a good method for including electron correlation.

The main limitation of perturbation methods is the assumption that the zero-order wave function is a reasonable approximation to the real wave function, i.e. the perturbation operator is sufficiently "small". The poorer the HF wave function describes

the system, the larger are the correction terms, and more terms must be included to achieve a given level of accuracy. If the reference state is a poor description of the system, the convergence may be so slow or erratic that perturbation methods cannot be used. Actually it is difficult to prove that the perturbation expansion is convergent, although many systems show a behaviour which suggests that it is the case. This may to some extent be deceptive, as it has been demonstrated that the convergence properties depend on the size of the basis set,⁹ and the majority of studies have employed small or medium sized basis sets. A convergent series in a DZP type basis for example may become divergent or oscillating in a larger basis, especially if diffuse functions are present.

In the ideal case the HF, MP2, MP3 and MP4 results show a monotonic convergence towards a limiting value, with the corrections being of the same sign and numerically smaller as the order of perturbation increases. Unfortunately, this is not the typical behaviour.¹⁰ Even in systems where the reference is well described by a single determinant, oscillations in a given property as a function of perturbation order are often observed. This is not completely understood, but may at least partly be due to the fact that the choice of the unperturbed Hamilton operator does not make the perturbation particularly small. An analysis by Cremer and He¹¹ indicates that a smooth convergence (of the total energy) is only expected for systems containing well-separated electron pairs, and that oscillations occur when this is not the case. These encompass systems containing lone pairs and/or multiple bonds, covering the large majority of molecules. Another study for the neon atom¹² indicates that convergence of the MP series requires that the energy spectrum is well-separated (i.e. the HF reference is significantly below the first excited state), not only for the unperturbed ($\mathbf{H} = \mathbf{H}_0$) and fully perturbed ($\mathbf{H} = \mathbf{H}_0 + \mathbf{H}'$) systems, but also for the “negative” perturbed case ($\mathbf{H} = \mathbf{H}_0 - \mathbf{H}'$). This condition was found not to hold when a basis set containing diffuse functions was used, and consequently the MP series was divergent. At low order of perturbation theory the divergence may not be obvious, but instead shows up as an oscillation in the total energy.

In practice only low orders of perturbation theory can be carried out, and it is often observed that the HF and MP2 results differ considerably, the MP3 result moves back towards the HF and the MP4 moves away again. For “well-behaved” systems the correct answer is normally somewhere between the MP3 and MP4 results. MP2 typically overshoots the correlation effect, but often gives a better answer than MP3, at least if medium sized basis sets are used. Just as the first term involving doubles (MP2) tends to overestimate the correlation effect, it is often observed that MP4 overestimates the effect of the singles and triples contributions, since they enter the series for the first time at fourth order.

When the reference wave function contains substantial multi-reference character, a perturbation expansion based on a single determinant will display poor convergence. If the reference wave function suffers from symmetry breaking (Section 3.8.3), the MP method is almost guaranteed to give absurd results. The ability to calculate terms up to MP4, however, allows an internal evaluation of the performance. If the MP2, MP3 and MP4 results appear to converge, it is possible at least to give a rough estimate of the “MP ∞ ” result (equivalent to full CI). If the MP2, MP3 and MP4 results on the other hand oscillate wildly, then one knows that none of the results can be trusted.

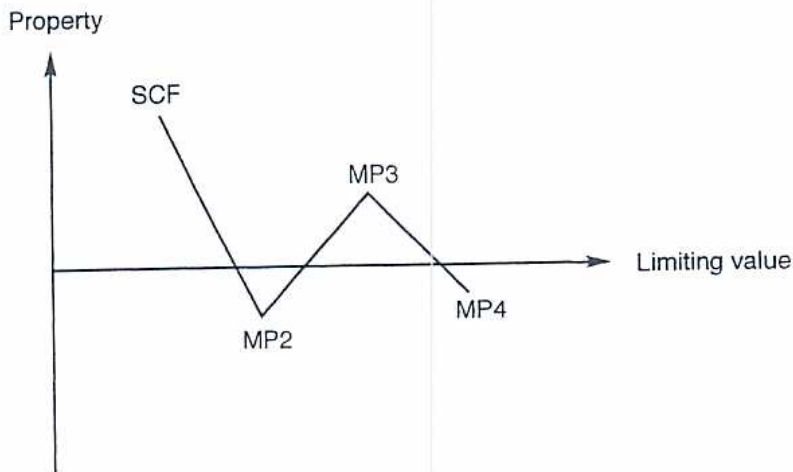


Figure 4.12 Typical oscillating behaviour of results obtained with the MP method

4.8.2 Unrestricted and Projected Møller-Plesset Methods

When the reference is an RHF type wave function the dissociation limit will normally be incorrect. As a bond is stretched, the RHF gives an increasingly poorer description of the wave function, and consequently the perturbation series will eventually break down. The use of a UHF wave function allows a correct dissociation limit in terms of energy, but at the cost of introducing spin contamination (Section 4.4). It is straightforward to derive an MP method based on a UHF reference wave function (UMP), in this case the unperturbed Hamilton operator is a sum of the α and β Fock operators. Addition of electron correlation decreases the spin contamination of the wave function (in the full CI limit the spin contamination is zero) but the improvement is usually small at low orders (2–4) of perturbation theory. As illustrated in Section 4.4, the UHF energy is lower than the RHF owing to the inclusion of some electron correlation (mainly static), but also contains some amount of higher energy spin states. Since MP methods recover a large part of the electron correlation (both static and dynamical), the net effect at the UMP level is an increase in energy due to spin contamination. In the dissociation limit, this has no consequence, as the different spin states have equal energies. In the intermediate region, where the bond is not completely broken, it is usually observed that the RMP n energy is lower than the UMP n energy, although the RHF energy is higher than the UHF (see also Section 11.5.2). The spin contamination in UHF wave functions causes a UMP n expansion to converge much more slowly than an RMP n .¹³ For open-shell systems, where RHF cannot be used, this would suggest that the reference wave function should be of the ROHF type, instead of the UHF. Formulation of ROHF based perturbation methods, however, is somewhat more difficult than for the UHF case. The reason is that for an ROHF wave function it is not possible to choose a set of MOs that makes the matrix of Lagrange multipliers diagonal (eqs. (3.40) and (3.41)). There is therefore not a unique set of canonical MOs to be used in the perturbation expansion, which again has the consequence that several choices of the unperturbed Hamilton operator are possible¹⁴. Different ROMP methods therefore give different energies, and there is no firm theoretical ground for choosing one over the other. In practice, however,

different choices of the unperturbed Hamilton operator lead to similar results, and perturbation calculations based on ROHF type wave functions are now becoming more and more common.

While projection methods for removing spin contamination are not recommended at the HF level, they work quite well at the UMP level. Formulas have been derived for removing all contaminants at the UMP2 level, and also the first few states at the UMP3 and UMP4 levels.¹⁵ The associated acronyms are PUMP and PMP, denoting slightly different methods, although they in practice give similar results. For singlet wave functions with bond lengths only slightly longer than the RHF/UHF instability point, such PUMP methods tend to give results very similar to those based on an RHF wave function. At longer bond lengths the RMP perturbation series eventually breaks down, while the PUMP series approaches the correct dissociation limit. It would therefore appear the PUMP methods should always be preferred. There are, however, also some computational factors to consider. First, UMP methods are by nature a factor of ~ 2 more expensive since there are twice as many MO coefficients. Second, the projection itself also uses CPU time. This is especially true if many of the higher spin states need to be removed, or for projection at the MP4 level. Third, it is difficult to formulate derivatives of projected wave functions, which at present limits PUMP methods to the calculation of energies. A rule of thumb says that for uncomplicated systems the RMP4 treatment gives acceptable accuracy (relative errors of the order of a few kcal/mol) up to bond lengths ~ 1.5 times the equilibrium length. Longer bonds are better treated by PUMP methods, see also Section 11.5.2. Most transition structures have bond lengths shorter than ~ 1.5 times the equilibrium length, and RMP4 often gives quite accurate activation energies.

Just as single reference CI can be extended to MRCI, it is also possible to use perturbation methods with a multi-determinant reference wave function. Formulating MR-MBPT methods, however, is not straightforward. The main problem here is similar to that of ROMP methods, the choice of the unperturbed Hamilton operator. Several different choices are possible, which will give different answers when the theory is carried out only to low order. Nevertheless, there are now several different implementations of MP2 type expansions based on a CASSCF reference, denoted CASMP2 or CASPT2.¹⁶ Experience of their performance is still somewhat limited.

4.9 Coupled Cluster Methods

Perturbation methods add all types of corrections (S, D, T, Q etc.) to the reference wave function to a given order (2, 3, 4 etc.). The idea in *Coupled Cluster* (CC) methods is to include all corrections of a given type to infinite order.¹⁷ The (intermediate normalized) coupled cluster wave function is written as

$$\Psi_{cc} = e^{\mathbf{T}} \Phi_0$$

$$e^{\mathbf{T}} = \mathbf{1} + \mathbf{T} + \frac{1}{2} \mathbf{T}^2 + \frac{1}{6} \mathbf{T}^3 + \dots = \sum_{k=0}^{\infty} \frac{1}{k!} \mathbf{T}^k \quad (4.46)$$

where the cluster operator \mathbf{T} is given by

$$\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3 + \dots + \mathbf{T}_N \quad (4.47)$$

The \mathbf{T}_i operator acting on a HF reference wave function generates all i th excited Slater determinants.

$$\begin{aligned}\mathbf{T}_1\Phi_0 &= \sum_i^{\text{occ}} \sum_a^{\text{vir}} t_i^a \Phi_i^a \\ \mathbf{T}_2\Phi_0 &= \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{vir}} t_{ij}^{ab} \Phi_{ij}^{ab}\end{aligned}\quad (4.48)$$

It is customary to use the term *amplitudes* for the expansion coefficients t , which are equivalent to the a_i coefficients in eq. (4.1).

From eqs. (4.46) and (4.47) the exponential operator may be written as

$$\begin{aligned}e^{\mathbf{T}} &= \mathbf{1} + \mathbf{T}_1 + \left(\mathbf{T}_2 + \frac{1}{2}\mathbf{T}_1^2\right) + \left(\mathbf{T}_3 + \mathbf{T}_2\mathbf{T}_1 + \frac{1}{6}\mathbf{T}_1^3\right) \\ &+ \left(\mathbf{T}_4 + \mathbf{T}_3\mathbf{T}_1 + \frac{1}{2}\mathbf{T}_2^2 + \frac{1}{2}\mathbf{T}_2\mathbf{T}_1^2 + \frac{1}{24}\mathbf{T}_1^4\right) + \dots\end{aligned}\quad (4.49)$$

The first term generates the reference HF and the second all singly excited states. The first parenthesis generates all doubly excited states, which may be considered as *connected* (\mathbf{T}_2) or *disconnected* (\mathbf{T}_1^2). The second parenthesis generates all triply excited states, which again may be either "true" (\mathbf{T}_3) or "product" triples ($\mathbf{T}_2\mathbf{T}_1$, \mathbf{T}_1^3). The quadruply excited states can similarly be viewed as composed of five terms, a true quadruple and four product terms. Physically a connected type such as \mathbf{T}_4 corresponds to four electrons interacting simultaneously, while a disconnected term such as \mathbf{T}_2^2 corresponds to two non-interacting pairs of interacting electrons.

With the coupled cluster wave function (4.46) the Schrödinger equation becomes

$$\mathbf{H}e^{\mathbf{T}}\Phi_0 = Ee^{\mathbf{T}}\Phi_0 \quad (4.50)$$

Multiplying from the left by Φ_0^* and integrating gives

$$\begin{aligned}\langle\Phi_0|\mathbf{H}e^{\mathbf{T}}|\Phi_0\rangle &= E_{\text{cc}}\langle\Phi_0|e^{\mathbf{T}}\Phi_0\rangle \\ \langle\Phi_0|\mathbf{H}e^{\mathbf{T}}|\Phi_0\rangle &= E_{\text{cc}}\langle\Phi_0|(\mathbf{1} + \mathbf{T}_1 + \mathbf{T}_2 + \dots)\Phi_0\rangle \\ E_{\text{cc}} &= \langle\Phi_0|\mathbf{H}e^{\mathbf{T}}|\Phi_0\rangle\end{aligned}\quad (4.51)$$

Expanding out the exponential in eq. (4.46) and using the fact that the Hamilton operator contains only one- and two-electron operators (eq. (3.24)) we get

$$\begin{aligned}E_{\text{cc}} &= \langle\Phi_0|\mathbf{H}|(\mathbf{1} + \mathbf{T}_1 + \mathbf{T}_2 + \frac{1}{2}\mathbf{T}_1^2)\Phi_0\rangle \\ E_{\text{cc}} &= \langle\Phi_0|\mathbf{H}|\Phi_0\rangle + \langle\Phi_0|\mathbf{H}|\mathbf{T}_1\Phi_0\rangle + \langle\Phi_0|\mathbf{H}|\mathbf{T}_2\Phi_0\rangle + \frac{1}{2}\langle\Phi_0|\mathbf{H}|\mathbf{T}_1^2\Phi_0\rangle \\ E_{\text{cc}} &= E_0 + \sum_i^{\text{occ}} \sum_a^{\text{vir}} t_i^a \langle\Phi_0|\mathbf{H}|\Phi_i^a\rangle + \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{vir}} (t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a) \langle\Phi_0|\mathbf{H}|\Phi_{ij}^{ab}\rangle\end{aligned}\quad (4.52)$$

When using HF orbitals for constructing the Slater determinants, the first matrix elements are zero (Brillouins theorem) and the second matrix elements are just

two-electron integrals over MOs (eq. (4.7)).

$$E_{cc} = E_0 + \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} (t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a) (\langle \phi_i \phi_j | \phi_a \phi_b \rangle - \langle \phi_i \phi_j | \phi_b \phi_a \rangle) \quad (4.53)$$

The coupled cluster correlation energy is therefore determined completely by the singles and doubles amplitudes and the two-electron MO integrals.

Only equation for the amplitudes is obtained by multiplying the Schrödinger equation (4.50) from the left by a singly excited determinant $(\Phi_m^e)^*$ and integrating.

$$\begin{aligned} \langle \Phi_m^e | \mathbf{H} e^{\mathbf{T}} | \Phi_0 \rangle &= E_{cc} \langle \Phi_m^e | e^{\mathbf{T}} | \Phi_0 \rangle \\ \langle \Phi_m^e | \mathbf{H} | (1 + \mathbf{T}_1 + \mathbf{T}_2 + \frac{1}{2} \mathbf{T}_1^2 + \mathbf{T}_3 + \mathbf{T}_1 \mathbf{T}_2 + \frac{1}{6} \mathbf{T}_1^3) \Phi_0 \rangle &= E_{cc} \langle \Phi_m^e | \mathbf{T}_1 \Phi_0 \rangle \\ \langle \Phi_m^e | \mathbf{H} | \Phi_0 \rangle + \langle \Phi_m^e | \mathbf{H} | \mathbf{T}_1 \Phi_0 \rangle + \langle \Phi_m^e | \mathbf{H} | \mathbf{T}_2 \Phi_0 \rangle + \frac{1}{2} \langle \Phi_m^e | \mathbf{H} | \mathbf{T}_1^2 \Phi_0 \rangle \\ + \langle \Phi_m^e | \mathbf{H} | \mathbf{T}_3 \Phi_0 \rangle + \langle \Phi_m^e | \mathbf{H} | \mathbf{T}_1 \mathbf{T}_2 \Phi_0 \rangle + \frac{1}{6} \langle \Phi_m^e | \mathbf{H} | \mathbf{T}_1^3 \Phi_0 \rangle &= E_{cc} \langle \Phi_m^e | \mathbf{T}_1 \Phi_0 \rangle \end{aligned} \quad (4.54)$$

Only the indicated terms survive in the expansion when the orthogonality of the Slater determinants and the nature of the Hamilton operator is considered. The first term in the last equation is again zero due to Brillouin's theorem, and the remaining forms a coupled set of equations of all singles, doubles and triples amplitudes (substituting in the expression for the energy, eq. (4.52)). Other equations connecting amplitudes may be obtained by multiplying from the left by a double, triple etc. excited determinant and integrating.

4.9.1 Truncated Coupled Cluster Methods

So far everything is exact. If all cluster operators up to \mathbf{T}_N are included in \mathbf{T} , all possible excited determinants are generated and the coupled cluster wave function is equivalent to full CI. This is as already stated impossible for all but the smallest systems. The cluster operator must therefore be truncated at some excitation level. When the \mathbf{T} operator is truncated, some of the terms in the amplitude equations will become zero, and the amplitudes derived from these approximate equations will no longer be exact. The energy calculated from these approximate singles and doubles amplitudes (eq. (4.53)) will therefore also be approximate. How severe the approximation is depends on how many terms are included in \mathbf{T} . Including only the \mathbf{T}_1 operator does not give any improvement over HF, as matrix elements between the HF and singly excited states are zero. The lowest level of approximation is therefore $\mathbf{T} = \mathbf{T}_2$, referred to as *Coupled Cluster Doubles* (CCD). Compared to the number of doubles, there are relatively few singly excited states. Using $\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2$ gives the CCSD model which is only slightly more demanding than CCD, and yields a more complete model. Both CCD and CCSD involve a computational effort which scales as M^6 in the limit of a large basis set. The next higher level has $\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3$, giving the CCSDT model¹⁸. This involves a computational effort which scales as M^8 , and is more demanding than CISDT. It (and higher-order methods like CCSDTQ) can consequently

only be used for small systems, and CCSD is the only generally applicable coupled cluster method.

Let us look in a bit more detail at the CCSD method. In this case we have (eqs. (4.46) and (4.47)).

$$e^{\mathbf{T}_1+\mathbf{T}_2} = \mathbf{1} + \mathbf{T}_1 + (\mathbf{T}_2 + \frac{1}{2}\mathbf{T}_1^2) + (\mathbf{T}_2\mathbf{T}_1 + \frac{1}{6}\mathbf{T}_1^3) + (\frac{1}{2}\mathbf{T}_2^2 + \frac{1}{2}\mathbf{T}_2\mathbf{T}_1^2 + \frac{1}{24}\mathbf{T}_1^4) + \dots \quad (4.55)$$

The CCSD energy is given by the general CC equation (4.53), and amplitude equations are derived by multiplying (4.50) with a singly excited determinant and integrating (analogously to eq. (4.54)).

$$\begin{aligned} \langle \Phi_m^e | \mathbf{H} | (\mathbf{1} + \mathbf{T}_1 + (\mathbf{T}_2 + \frac{1}{2}\mathbf{T}_1^2) + (\mathbf{T}_2\mathbf{T}_1 + \frac{1}{6}\mathbf{T}_1^3)) \Phi_0 \rangle &= E_{\text{CCSD}} \langle \Phi_m^e | \mathbf{T}_1 \Phi_0 \rangle \\ \langle \Phi_m^e | \mathbf{H} | \Phi_0 \rangle + \langle \Phi_m^e | \mathbf{H} | \mathbf{T}_1 \Phi_0 \rangle + \langle \Phi_m^e | \mathbf{H} | (\mathbf{T}_2 + \frac{1}{2}\mathbf{T}_1^2) \Phi_0 \rangle & \\ + \langle \Phi_m^e | \mathbf{H} | (\mathbf{T}_2\mathbf{T}_1 + \frac{1}{6}\mathbf{T}_1^3) \Phi_0 \rangle &= E_{\text{CCSD}} \sum_{ia} t_i^a \langle \Phi_m^e | \Phi_i^a \rangle \\ \langle \Phi_m^e | \mathbf{H} | \Phi_0 \rangle + \sum_{ia} t_i^a \langle \Phi_m^e | \mathbf{H} | \Phi_i^a \rangle + \sum_{ijab} (t_{ij}^{ab} + t_i^a t_j^b - t_j^b t_i^a) \langle \Phi_m^e | \mathbf{H} | \Phi_{ij}^{ab} \rangle & \\ + \sum_{ijkabc} (t_{ij}^{ab} t_k^c + \dots + t_i^a t_j^b t_k^c + \dots) \langle \Phi_m^e | \mathbf{H} | \Phi_{ijk}^{abc} \rangle &= E_{\text{CCSD}} t_m^e \end{aligned} \quad (4.56)$$

The notation $(t_i^a t_j^b t_k^c + \dots)$ indicates that several other terms involving permutations of the indices are omitted. Multiplying eq. (4.50) with a doubly excited determinant gives

$$\begin{aligned} \langle \Phi_{mn}^{ef} | \mathbf{H} | (\mathbf{1} + \mathbf{T}_1 + (\mathbf{T}_2 + \frac{1}{2}\mathbf{T}_1^2) + (\mathbf{T}_2\mathbf{T}_1 + \frac{1}{6}\mathbf{T}_1^3) & \\ + (\frac{1}{2}\mathbf{T}_2^2 + \frac{1}{2}\mathbf{T}_2\mathbf{T}_1^2 + \frac{1}{24}\mathbf{T}_1^4)) \Phi_0 \rangle &= E_{\text{CCSD}} \langle \Phi_{mn}^{ef} | (\mathbf{T}_2 + \frac{1}{2}\mathbf{T}_1^2) \Phi_0 \rangle \\ \langle \Phi_{mn}^{ef} | \mathbf{H} | \Phi_0 \rangle + \langle \Phi_{mn}^{ef} | \mathbf{H} | \mathbf{T}_1 \Phi_0 \rangle & \\ + \langle \Phi_{mn}^{ef} | \mathbf{H} | (\mathbf{T}_2 + \frac{1}{2}\mathbf{T}_1^2) \Phi_0 \rangle + \langle \Phi_{mn}^{ef} | \mathbf{H} | (\mathbf{T}_2\mathbf{T}_1 + \frac{1}{6}\mathbf{T}_1^3) \Phi_0 \rangle & \\ + \langle \Phi_{mn}^{ef} | \mathbf{H} | (\frac{1}{2}\mathbf{T}_2^2 + \frac{1}{2}\mathbf{T}_2\mathbf{T}_1^2 + \frac{1}{24}\mathbf{T}_1^4) \Phi_0 \rangle &= E_{\text{CCSD}} \sum_{ijab} (t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a) \langle \Phi_{mn}^{ef} | \Phi_{ij}^{ab} \rangle \\ \langle \Phi_{mn}^{ef} | \mathbf{H} | \Phi_0 \rangle + \sum_{ia} t_i^a \langle \Phi_{mn}^{ef} | \mathbf{H} | \Phi_i^a \rangle + \sum_{ijab} (t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a) \langle \Phi_{mn}^{ef} | \mathbf{H} | \Phi_{ij}^{ab} \rangle & \\ + \sum_{ijkabc} (t_{ij}^{ab} t_k^c + \dots + t_i^a t_j^b t_k^c + \dots) \langle \Phi_{mn}^{ef} | \mathbf{H} | \Phi_{ijk}^{abc} \rangle & \\ + \sum_{ijklabcd} (t_{ij}^{ab} t_{kl}^{cd} + \dots + t_{ij}^{ab} t_k^c t_l^d + \dots + t_i^a t_j^b t_k^c t_l^d + \dots) \langle \Phi_{mn}^{ef} | \mathbf{H} | \Phi_{ijkl}^{abcd} \rangle & \\ = E_{\text{CCSD}} (t_{mn}^{ef} + t_m^e t_n^f - t_m^f t_n^e) & \end{aligned} \quad (4.57)$$

The equations (4.56) and (4.57) involve matrix elements between singles and triples and between doubles and quadruples. However, since the Hamiltonian operator only contains

one- and two-electron operators, these are actually identical to matrix elements between the reference and a doubly excited state. Consider for example $\langle \Phi_m^e | \mathbf{H} | \Phi_{ijk}^{abc} \rangle$. Unless m equals either i, j or k , and e equals either a, b , or c , there will be one overlap integral between different MOs which makes the matrix element zero. If for example $m = k$ and $e = c$, then the MO integral over these indices factor out as 1, and the rest is equal to a matrix element $\langle \Phi_0 | \mathbf{H} | \Phi_{ij}^{ab} \rangle$. Similarly, the matrix element $\langle \Phi_0 | \mathbf{H} | \Phi_{ijkl}^{abcd} \rangle$ between a doubly and a quadruply excited determinant is only non-zero if mn matches up with two of the $ijkl$ indices, and ef matches up with $abcd$. Again such non-zero matrix elements are equal to matrix elements between the reference and a doubly excited determinant, eq. (4.7).

All the matrix elements can be evaluated in terms of MO integrals, and when the expression for the energy (4.53) is substituted into (4.56) and (4.57), they form coupled non-linear equations for the singles and doubles amplitudes. The equations contain terms up to quartic in the amplitudes, e.g. $(t_i^a)^4$ (since \mathbf{H} contains one- and two-electron operators), and must be solved by iterative techniques. Once the amplitudes are known, the energy and wave function can be calculated. The important aspect in coupled cluster methods is that excitations of higher order than the truncation of the \mathbf{T} operator enter the amplitude equation. Quadruply excited states, for example, are generated by the \mathbf{T}_2^2 operator in CCSD, and they enter the amplitude equations with a weight given as a product of doubles amplitudes. Quadruply excited states influence the doubles amplitudes, and thereby also the CCSD energy. It is the inclusion of these products of excitations that makes coupled cluster theory size extensive.

In the above the coupled cluster equations have been derived by multiplying the Schrödinger equation with $\langle \Phi_0 |$, $\langle \Phi_m^e |$ and $\langle \Phi_{mn}^{ef} |$. An alternative way of deriving the coupled cluster equations is to multiply with $\langle \Phi_0 | e^{-\mathbf{T}}$, $\langle \Phi_m^e | e^{-\mathbf{T}}$ and $\langle \Phi_{mn}^{ef} | e^{-\mathbf{T}}$. Just as $e^{\mathbf{T}}$ is an excitation operator working on the function to the right, $e^{-\mathbf{T}}$ is a de-excitation operator working on the function to the left. Thus $\langle \Phi_0 | e^{-\mathbf{T}}$ tries to generate de-excitations from the reference, which is impossible, i.e. $\langle \Phi_0 | e^{-\mathbf{T}} = \langle \Phi_0 |$. However $\langle \Phi_m^e | e^{-\mathbf{T}}$ generates in addition to the singly excited state also the reference wave function. Similarly $\langle \Phi_{mn}^{ef} | e^{-\mathbf{T}}$ generates additionally the reference and singly excited states. The main advantage is that the coupled cluster equations are obtained directly without having to substitute in the expression for the energy in the amplitude equations.

4.10 Connections between Coupled Cluster, Configuration Interaction and Perturbation Theory

The general cluster operator is given by

$$e^{\mathbf{T}} = 1 + \mathbf{T}_1 + \left(\mathbf{T}_2 + \frac{1}{2} \mathbf{T}_1^2 \right) + \left(\mathbf{T}_3 + \mathbf{T}_2 \mathbf{T}_1 + \frac{1}{6} \mathbf{T}_1^3 \right) + \left(\mathbf{T}_4 + \mathbf{T}_3 \mathbf{T}_1 + \frac{1}{2} \mathbf{T}_2^2 + \frac{1}{2} \mathbf{T}_2 \mathbf{T}_1^2 + \frac{1}{24} \mathbf{T}_1^4 \right) + \dots \quad (4.58)$$

where terms have been collected according to the excitation they generate. Each of the operators in a given parenthesis generates all the excited determinants of given type. Both \mathbf{T}_2 and \mathbf{T}_1^2 generate all doubly excited determinants, and the terms in (4.58) generate all determinants which are included in a CISDTQ calculation. This cluster

expansion can be viewed as a method of dividing up the contributions from each excitation type. The total contribution from double excitations is the sum of two terms, one which is the square of the singles contributions and the remaining is (by definition) the connected doubles. Similarly the total contribution from triple excitations is a sum of three terms, the cube of the singles contributions, the product of the singles and doubles contributions, and the remaining is the connected triples.

When canonical HF orbitals are used the T_1 effect is small, although not zero since singles enter indirectly via the doubly excited states (note that if non-canonical orbitals are used, the T_1 term can be large). From CI we know that the effect of doubles is the most important, Section 4.2.3. In coupled cluster theory the doubles contribution is divided into T_1^2 and T_2 . If T_1 is small, then T_1^2 must also be small, and the most important term is T_2 . For the triples excitations, T_1^3 must be negligible, and $T_1 T_2$ is small owing to T_1 . The most important contribution is therefore from connected triples T_3 . For the quadruple excitations, all the terms involving T_1 must again be small, and since T_2 is large, we expect the disconnected quadruples T_2^2 to be the dominant term. This again suggests that the connected quadruples term T_4 is small. Higher-order excitations will always contain terms appearing as powers and/or products of T_2 and T_3 , which normally will be dominating. Higher-order connected terms, T_n with $n > 4$, are therefore expected to have small effects. This is consistent with the physical picture that connected T_n operators correspond to n electrons interacting simultaneously. As n becomes large, this is increasingly improbable. It should be noted however, that the higher-order cluster operators (T_4, T_5, \dots) become more and more important as the number of electrons increases.

The principal deficiency of CISD is the lack of the T_2^2 term, which is the main reason for CISD not being size extensive. Furthermore, this term becomes more and more important as the number of electrons increases, and CISD therefore recovers a smaller and smaller percentage of the correlation energy as the system increases. There are various approximate corrections for this lack of size extensivity which can be added to standard CISD. The most widely known of these is the *Davidson correction*, sometimes denoted CISD + Q(Davidson), where the quadruples contribution is approximated as

$$\Delta E_Q = (1 - a_0^2) \Delta E_{\text{CISD}} \quad (4.59)$$

with a_0 being the coefficient in front of the HF reference. If the renormalization of the wave function is also taken into account, the $(1 - a_0^2)$ quantity is divided by a_0^2 , and the corresponding correction is called the *renormalized Davidson correction*. The effect of higher-order excitations is thus estimated from the correlation energy obtained at the CISD level times a factor which measures how important the single determinant reference is at the CISD level. The Davidson correction does not yield zero for two-electron systems, where CISD is equivalent to full CI, and it is likely that it overestimates the higher-order corrections for systems with few electrons. More complicated correction schemes have also been proposed¹⁹ but are rarely used.

Coupled cluster is closely connected with Møller–Plesset perturbation theory, as mentioned at the start of this section. The infinite Taylor expansion of the exponential operator (eq. (4.46)) ensures that the contributions from a given excitation level are included to infinite order. Perturbation theory indicates that doubles are the most important, they are the only contributors to MP2 and MP3. At fourth order, there are contributions from singles, doubles, triples and quadruples. The MP4 quadruples

contribution is actually the disconnected T_2^2 term in the coupled cluster language, and the triples contribution corresponds to T_3 . This is consistent with the above analysis, the most important is T_2 (and products thereof) followed by T_3 . The CCD energy is equivalent to MP ∞ (D) where all disconnected contributions of products of doubles are included. If the perturbation series is reasonably converged at fourth order, we expect that $CCD \sim MP4(DQ)$, and $CCSD \sim MP4(SDQ)$. The MP2, MP3 and MP4(SDQ) results may be obtained in the first iteration for the CCSD amplitudes, allowing a direct test of the convergence of the MP series. This also points out the principal limitation of the CCSD method, the neglect of the connected triples. Including them in the T operator leads to the CCSDT method which, as mentioned above, is too demanding computationally for all but the smallest systems. Alternatively the triples contribution may be evaluated by perturbation theory and added to the CCSD results. Several such hybrid methods exist;²⁰ two of the most common are known by the acronyms $CCSD+T(CCSD)$ and $CCSD(T)$. In both cases the triples contribution is calculated from the formula given by MP4, but using the CCSD amplitudes instead of the perturbation coefficients for the wave function corrections (eqs. (4.37) and (4.39)). For the $CCSD(T)$ method an additional term arising from fifth-order perturbation theory, describing the coupling between singles and triples, is also included. This is computationally inexpensive to calculate, and the $CCSD(T)$ method is preferred over $CCSD+CCSD(T)$. Higher-order hybrid methods such as $CCSD(TQ)$, where the connected quadruples contribution is estimated by fifth-order perturbation theory, are also possible, but they are again so demanding that they can only be used for small systems.²¹

As mentioned, the singles make a fairly small contribution to the correlation energy when canonical HF orbitals are used. *Brueckner* theory²² is a variation of coupled cluster where the orbitals used for constructing the Slater determinants are optimized so that the contribution from singles is exactly zero, i.e. $t_i^a = 0$. The lowest level of Brueckner theory includes only doubles, giving the acronym BD. Although BD in theory should be slightly better than CCSD, since it includes orbital relaxation, it gives in practice essentially identical results (differences between BD and CCSD are of fifth order or higher in term of perturbation theory). This is presumably rooted in the fact that the singles in CCSD introduce orbital relaxation.²³ The computational cost is also very similar for CCSD and BD.²⁴ Similarly $BD(T)$ is essentially equivalent to $CCSD(T)$,²⁵ and $BD(TQ)$ to $CCSD(TQ)$.

Since the singly excited determinants effectively relax the orbitals in a CCSD calculation, non-canonical HF orbitals can also be used in coupled cluster methods. This allows for example the use of open-shell singlet states (which require two Slater determinants) as reference for a coupled cluster calculation.²⁶

Another commonly used method is *Quadratic* CISD (QCISD). It was originally derived from CISD by including enough higher-order terms to make it size extensive.²⁷ It has later been shown that the resulting equations are identical to CCSD where some of the terms have been omitted.²⁸ The omitted terms are computationally inexpensive, and there appears to be no reason for using the less complete QCISD over CCSD (or QCISD(T) in place of CCSD(T)), although in practice they normally give very similar results.²⁹ There are a few other methods which may be considered either as CISD with addition of extra terms to make them approximately size extensive, or as approximate versions of CCSD. Some of the methods falling into this category are *Averaged*

Coupled-Pair Functional (ACPF) and *Coupled Electron Pair Approximation* (CEPA). The simplest form of CEPA, CEPA-0, is also known as *Linear Coupled Cluster Doubles* (LCCD).

Recently two new intermediate coupled cluster methods for calculating molecular properties have been defined, known as CC2 and CC3.³⁰ The single excitations allow the MOs to relax from their HF form, but do not give any direct contribution to the energy due to the Brillouin theorem. For studying properties which measure the response of the energy to a perturbation, the HF orbitals are no longer optimum, and the singles are at least as important as the doubles. The CC2 method is derived from CCSD by including only the doubles contribution arising from the lowest (non-zero) order in perturbation theory, where the perturbation is defined as in MP theory (i.e. as the true electron-electron potential minus twice the average repulsion). The amplitude equations corresponding to multiplication of a doubly excited determinant in the CCSD equations (eq. (4.57)) thereby reduce to an MP2-like expression, and the t_2 amplitudes may be expressed directly in terms of the t_1 amplitudes and MO integrals. The iterative procedure therefore only involves the t_1 amplitudes and MO integrals. The iterative MP2 with the added feature of orbital relaxation arising from the singles. Similarly, CC3 is an approximation to the full CCSDT model, where the triples contribution is approximated by the expression arising from the lowest non-vanishing order in perturbation theory. The triples amplitudes can then be expressed directly in terms of the singles and doubles amplitudes, and MO integrals. Both in terms of computational cost and accuracy, the following progression is expected, although the CC2 and CC3 models are so new that there are few data for comparison.

$$\text{HF} \ll \text{CC2} < \text{CCSD} < \text{CC3} < \text{CCSDT}$$

Analogously to MP methods, coupled cluster theory may also be based on a UHF reference wave function. The resulting UCC methods again suffer from spin contamination of the underlying UHF, but the infinite nature of coupled cluster methods is substantially better at reducing spin contamination relative to UMP.³¹ Projection methods analogous to those of the PUMP case have been considered but are not commonly used. ROHF based coupled cluster methods have also been proposed, but appear to give results very similar to UCC, especially at the CCSD(T) level.³²

Standard coupled cluster theory is based on a single determinant reference wave function. It suffers from the same problem as MP: it works best if the zeroth-order wave function is sufficiently "good". Due to the summation of contributions to infinite order, however, coupled cluster is somewhat more tolerant of a poor reference wave function than MP methods. Since the singly excited determinants allow the MOs to relax in order to describe the multi-reference character in the wave function, the magnitude of the singles amplitude is an indicator of how good the HF single determinant is as a reference. The T_1 -diagnostic defined as the norm of the singles amplitude vector divided by the square root of the number of electrons has been suggested as an internal evaluation of the quality of a CCSD wave function.³³

$$T_1 = \frac{|\mathbf{t}_1|}{\sqrt{N}} \quad (4.60)$$

Specifically, if $T_1 < 0.02$, the CCSD(T) method is expected to give results close the full CI limit for the given basis set. If T_1 is larger than 0.02, it indicates that the reference wave function has significant multi-determinant character, and multi-reference coupled cluster should preferentially be employed. Such methods are being developed³⁴ but have not yet seen any extensive use.

4.11 Methods Involving Interelectronic Distances

The necessity for going beyond the HF approximation is the fact that electrons are further apart than described by the product of their orbital densities, i.e. their motions are correlated. This arises from the electron–electron repulsion operator, which is a sum of terms of the type

$$\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} = \frac{1}{r_{12}} \quad (4.61)$$

Without these terms the Schrödinger equation can be solved exactly, with the solution being a Slater determinant composed of orbitals.

The electron–electron repulsion operator has a singularity for $r_{12} = 0$ which results in the exact wave function having a *cusp* (discontinuous derivative).³⁵

$$\left(\frac{\partial \Psi}{\partial r_{12}} \right)_{r_{12}=0} = \frac{1}{2} \Psi(r_{12} = 0) \quad (4.62)$$

In other words, the exact wave function behaves asymptotically as a constant $+1/2r_{12}$ when r_{12} is small. It would therefore seem natural that the interelectronic distance would be a necessary variable for describing electron correlation. For two-electron systems, extremely accurate wave functions may be generated by taking a trial wave function consisting of an orbital product times an expansion in electron coordinates such as

$$\Psi(\mathbf{r}_1, \mathbf{r}_2) = e^{-\alpha_1 r_1} e^{-\alpha_2 r_2} \sum_{klm} C_{klm} (\mathbf{r}_1 + \mathbf{r}_2)^k (\mathbf{r}_1 - \mathbf{r}_2)^l r_{12}^m \quad (4.63)$$

and variationally optimize the α_i and C_{klm} parameters. Such expansions are known as *Hylleraas* type wave functions. For the hydrogen molecule it is possible to converge the total energy to $\sim 10^{-9}$ a.u., which is more accurate than can be determined experimentally. In fact, the prediction that the experimental dissociation energy for H_2 was wrong, based on calculations, was one of the first hallmarks of quantum chemistry.³⁶ Such wave functions unfortunately become impractical for more than 3–4 electrons.

All electron correlation methods based on expanding the N -electron wave function in terms of Slater determinants built from orbitals (one-electron functions) suffer from an agonizingly slow convergence. Literally millions or billions of determinants are required for obtaining results which in an absolute sense are close to the exact results. This is due to the fact that products of one-electron functions are poor at describing the cusp behaviour of the wave function when two electrons are close together. At the second-order perturbation level (i.e. MP2) it may be shown that the error in the correlation energy behaves asymptotically as $(L + 1/2)^{-4}$, where l is the highest angular

momentum in the basis set. For a general wave function the convergence is $(L + 1/2)^{-4} + (L + 1/2)^{-5} + (L + 1/2)^{-6} + \dots$. This means that the total energy will converge as $(L + 1)^{-3} + (L + 1)^{-4} + (L + 1)^{-5} + \dots$, if the basis set is saturated up to angular momentum L .³⁷ For sufficiently large values of L the convergence is thus $\sim (L + 1)^{-3}$, which is quite slow.

In order to achieve a high accuracy, it would seem desirable to explicitly include terms in the wave functions which are linear in the interelectronic distance. This is the idea in the R12 methods developed by Kutzelnigg and co-workers.³⁸ The first order correction to the HF wave function only involves doubly excited determinants (eqs. (4.35) and (4.37)). In R12 methods additional terms are included which essentially are the HF determinant multiplied with r_{ij} factors.

$$\Psi_{R12} = \Phi_{HF} + \sum_{ijab} a_{ijab} \Phi_{ij}^{ab} + \sum_{ij} b_{ij} r_{ij} \Phi_{HF} \quad (4.64)$$

The exact definition is slightly more complicated, since the wave function has to be properly antisymmetrized and projected onto the actual basis, but for illustration the above form is sufficient. Such R12 wave functions may then be used in connection with the CI, MBPT or CC methods described above. Consider for example a CI calculation with an R12 type wave function. The energy is given as

$$E = \langle \Psi_{R12} | \mathbf{H} | \Psi_{R12} \rangle \quad (4.65)$$

and the a_{ijab} and b_{ij} parameters in (4.64) are optimized variationally. The overwhelming problem is that matrix elements from (4.65) now involve integrals depending on three and four electron coordinates. Consider for example the following terms arising from the r_{ij} operator written out in terms of the one- and two-electron operators (eq. (3.24)).

$$\begin{aligned} \langle \Phi_{HF} | \mathbf{H} | r_{ij} \Phi_{HF} \rangle &= \langle \Phi_{HF} | \mathbf{h} | r_{ij} \Phi_{HF} \rangle + \langle \Phi_{HF} | \mathbf{g} | r_{ij} \Phi_{HF} \rangle \\ \langle r_{ij} \Phi_{HF} | \mathbf{H} | r_{ij} \Phi_{HF} \rangle &= \langle r_{ij} \Phi_{HF} | \mathbf{h} | r_{ij} \Phi_{HF} \rangle + \langle r_{ij} \Phi_{HF} | \mathbf{g} | r_{ij} \Phi_{HF} \rangle \end{aligned} \quad (4.66)$$

the \mathbf{g} -operator leads to integrals over molecular orbitals of the type

$$\begin{aligned} &\left\langle \phi_i(1)\phi_j(2)\phi_k(3) \left| \frac{r_{12}}{r_{13}} \right| \phi_i(1)\phi_j(2)\phi_k(3) \right\rangle \\ &\left\langle \phi_i(1)\phi_j(2)\phi_k(3) \left| \frac{r_{12}r_{23}}{r_{13}} \right| \phi_i(1)\phi_j(2)\phi_k(3) \right\rangle \\ &\left\langle \phi_i(1)\phi_j(2)\phi_k(3)\phi_l(4) \left| \frac{r_{12}r_{13}}{r_{23}} \right| \phi_i(1)\phi_j(2)\phi_k(3)\phi_l(4) \right\rangle \end{aligned} \quad (4.67)$$

Not only are such integrals difficult to calculate, but when the MOs are expanded in a basis set consisting of M AOs, there will be on the order of M^6 three-electron integrals and on the order of M^8 four-electron integrals. Such methods are therefore inherently more expensive than for example the full CCSDT model.

The trick for turning the R12 method into a viable tool is to avoid calculating the three- and four-electron integrals, without jeopardizing the accuracy. In a complete basis, a three-electron integral may be written in terms of products of two-electron

integrals by inserting a "resolution of the identity" between the two operators.

$$\begin{aligned}
 \mathbf{1} &= \sum_p |\phi_p\rangle\langle\phi_p| = \sum_{pqr} |\phi_p\phi_q\phi_r\rangle\langle\phi_p\phi_q\phi_r| \\
 \left\langle \phi_i\phi_j\phi_k \left| \frac{r_{12}}{r_{13}} \right| \phi_i\phi_j\phi_k \right\rangle &= \sum_{pqr} \langle\phi_i\phi_j\phi_k|r_{12}|\phi_p\phi_q\phi_r\rangle \left\langle \phi_p\phi_q\phi_r \left| \frac{1}{r_{13}} \right| \phi_i\phi_j\phi_k \right\rangle \\
 &= \sum_{pqr} (\delta_{kr}\langle\phi_i\phi_j|r_{12}|\phi_p\phi_q\rangle) \left(\delta_{aj} \left\langle \phi_p\phi_r \left| \frac{1}{r_{13}} \right| \phi_i\phi_k \right\rangle \right) \\
 &= \sum_p \langle\phi_i\phi_j|r_{12}|\phi_p\phi_j\rangle \left\langle \phi_p\phi_k \left| \frac{1}{r_{13}} \right| \phi_i\phi_k \right\rangle
 \end{aligned} \tag{4.68}$$

The first reduction occurs since the r_{12} and r_{13}^{-1} operators only involve two-electron coordinates, the second reduction is due to the two delta functions. Three- and four-electron integrals can therefore be written as a sum over products of integrals involving only two electron coordinates. In a finite basis set, the resolution is not exact, and the identities in eq. (4.68) become approximations. The beauty of the R12 methods is that this error can be controlled, and is not significantly larger than the inherent basis set error, once the basis set reaches a certain size. The significance of R12 methods is that the energy error in terms of angular momentum of the basis set now behaves approximately as $(L+1)^{-7}$, a significant improvement over standard methods. The drawback is that R12 methods cannot be used with small basis sets, only in well-polarized basis sets is the resolution obeyed with a reasonable accuracy. In practice this means that a fairly dense basis set including up to at least f -functions must be used. However, if accurate results are desired, a large polarized basis set is required anyway, and R12 methods converge much faster in terms of basis set extension than traditional methods.³⁹ It should be noted that in the limit of a complete basis set the MP2-R12 (for example) will give the same result as a traditional MP2 calculation, i.e. the R12 approach speeds up the basis set convergence, but does not change the fundamental characteristics of the MP2 method. The improved convergence, however, does not come for free: there is a significantly larger number (and different types) of two-electron integrals which must be calculated and handled.

4.12 Direct Methods

Conventional HF methods rely on storing the two-electron integrals over atomic orbitals on disk, and reading them in each SCF iteration, while direct methods generate the integrals as they are needed (Section 3.8.5). This is an easy change in algorithm since the HF energy is expressed directly in terms of AO integrals. Methods involving electron correlation, however, require matrix elements between Slater determinants, which can be expressed in terms of integrals over MOs (eq. (4.7)). Conventional methods for the integral transformation (Section 4.2.1) read the AOs, perform the multiplications with the MO coefficients (eq. (4.11)), and write the MO integrals to disk. These can then be read in and used in the correlation treatment. Although the number of MO integrals typically is somewhat smaller than the number of AO integrals (for example MO integrals involving four virtual orbitals may not be needed), the disk space requirements

are still significant if more than a few hundred basis functions are used. To eliminate the disk space requirements, and remove the relatively inefficient data transfer step for reading/writing to disk, it is desirable also to have direct algorithms for electron correlation method. Direct in this context means that the integrals are calculated as needed and then discarded. The need for integrals over MOs instead of AOs, however, makes the development of direct methods in electron correlation somewhat more complicated than at the HF level.

Consider for example the MP2 energy expression.⁴⁰

$$\text{MP2} = \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} \frac{[\langle \phi_i \phi_j | \phi_a \phi_b \rangle - \langle \phi_i \phi_j | \phi_b \phi_a \rangle]^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \quad (4.69)$$

with the MO integrals given as

$$\langle \phi_i \phi_j | \phi_a \phi_b \rangle = \sum_{\alpha} \sum_{\beta} \sum_{\gamma} \sum_{\delta} c_{\alpha i} c_{\beta j} c_{\gamma a} c_{\delta b} \langle \chi_{\alpha} \chi_{\beta} | \chi_{\gamma} \chi_{\delta} \rangle \quad (4.70)$$

Since each MO integral in principle contains contributions from all the AO integrals, a straightforward calculation of an MO integral each time it is needed will involve a generation of all the AO integrals. In other words, it would be necessary to recalculate the AO integrals $\sim O^2 V^2$ times (O and V being the number of occupied and virtual orbitals), compared to the 15–20 times in an SCF calculation. The MP2 method would therefore change from being an M^5 to an M^8 method, which is clearly an unacceptably large penalty for a direct method.

The M^8 dependence is a consequence of performing the four index transformation with all four indices at once. As shown in Section 4.2.1, it is advantageous to perform the transformation one index at a time.

$$\begin{aligned} \langle \phi_i \phi_j | \phi_a \phi_b \rangle &= \sum_{\delta} c_{\delta b} \langle \phi_i \phi_j | \phi_a \chi_{\delta} \rangle \\ \langle \phi_i \phi_j | \phi_a \chi_{\delta} \rangle &= \sum_{\gamma} c_{\gamma a} \langle \phi_i \phi_j | \chi_{\gamma} \chi_{\delta} \rangle \\ \langle \phi_i \phi_j | \chi_{\gamma} \chi_{\delta} \rangle &= \sum_{\beta} c_{\beta j} \langle \phi_i \chi_{\beta} | \chi_{\gamma} \chi_{\delta} \rangle \\ \langle \phi_i \chi_{\beta} | \chi_{\gamma} \chi_{\delta} \rangle &= \sum_{\alpha} c_{\alpha i} \langle \chi_{\alpha} \chi_{\beta} | \chi_{\gamma} \chi_{\delta} \rangle \end{aligned} \quad (4.71)$$

By choosing the right order of the transformation the scaling can be reduced considerably. In eq. (4.71) the indices corresponding to the occupied orbitals may be transformed before the virtuals. There are of the order of M^4 AO integrals, $\langle \chi_{\alpha} \chi_{\beta} | \chi_{\gamma} \chi_{\delta} \rangle$, but only OM^3 quarter transformed integrals, $\langle \phi_i \chi_{\beta} | \chi_{\gamma} \chi_{\delta} \rangle$. Instead of storing and reading the AO integrals from the SCF step, they can be recalculated in the transformation step, reducing the storage from M^4 to OM^3 . The subsequent quarter transformations require less storage, i.e. the next transformation with an occupied index reduces the number of integrals to $O^2 M^2$, the third to $O^2 VM$, and the last to $O^2 V^2$. Since the MP2 energy can be written as a sum of contributions from each occupied orbital, we can furthermore treat one occupied orbital at a time, i.e. first sum all contributions of $\langle \phi_1 \chi_{\beta} | \chi_{\gamma} \chi_{\delta} \rangle$, then $\langle \phi_2 \chi_{\beta} | \chi_{\gamma} \chi_{\delta} \rangle$ etc. This reduces the necessary

storage to only order M^3 . It may be further reduced to OVM by proper scheduling of the evaluation order of the remaining three indices. The OVM number of integrals is much less than the original M^4 , and will in many cases fit into the memory. The net result is that disk storage is effectively eliminated, or at least greatly reduced. If only one occupied orbital is treated at a time, O integral evaluations are required, however, the more memory that is available, the more occupied orbitals can be treated in a single sweep, decreasing the number of integral evaluations.

The above is an example of how direct algorithms may be formulated for methods involving electron correlation. It illustrates that it is not as straightforward to apply direct methods at the correlated level as at the SCF level. However, the steady increase in CPU performance, and especially the evolution of multiprocessor machines, favours direct (and semi-direct where some intermediate results are stored on disk) algorithms. Recently direct methods have also been implemented at the coupled cluster level.⁴¹

4.13 Localized Orbital Methods

Ab initio calculations involving electron correlation essentially always build on a set of canonical HF orbitals. As illustrated in this chapter, this leads to a computational effort which increases as a rather high power of the system size, i.e. M^5 – M^8 . Considering that the fundamental physical force is only between pairs of particles, this scaling is “unphysical”. One of the reason for the high scaling is that fact the canonical orbitals in general are delocalized over the whole molecule, i.e. essentially all orbitals make a (small) contribution to the wave function for a specific part of the molecule. This suggests that localized orbitals may be a better starting point, since a single or only a few orbitals then would contribute the large majority at a given point, and the remaining contributions could simply be neglected. Such local MP2 and local CC methods have started to appear, but are not yet commonly used.⁴² These methods are somewhat more complicated to formulate as the Fock matrix is only diagonal in the canonical orbitals. Nevertheless, methods based on localized orbitals hold the promise of a near-linear scaling with problem size in the large-scale limit. It is at present not clear exactly how large the systems need to be to reach the “large-scale” limit.

4.14 Summary of Electron Correlation Methods

The only generally applicable methods are: CISD, MP2, MP3, MP4, CCSD and CCSD(T). CISD is variational, but not size extensive, while MP and CC methods are non-variational but size extensive. CISD and MP are in principle non-iterative methods, although the matrix diagonalization involved in CISD usually is so large that it has to be done iteratively. Solution of the coupled cluster equations must be done by an iterative technique since the parameters enter in a non-linear fashion. In terms of the most expensive step in each of the methods they may be classified according to how they formally scale in the large system limit, as shown in Table 4.5.

We have so far been careful to used the wording “formal scaling”. As already discussed, HF is formally an M^4 method but in practice the scaling may be reduced all the way down to M^1 . Similarly, MP2 is formally an M^5 method. However, an MP2 calculation consists of three main parts: the HF calculation, the AO to MO integral

Table 4.5 Limiting scaling in terms of basis set size M for different methods

Scaling	CI methods	MP methods	CC methods
M^5		MP2	CC2 (iterative)
M^6	CISD	MP3, MP4(SDQ)	CCSD (iterative)
M^7		MP4	CCSD(T), CC3 (iterative)
M^8	CISDT	MP5	CCSDT (iterative)
M^9		MP6	
M^{10}	CISDTQ	MP7	CCSDTQ (iterative)

transformation, and the MP2 energy calculation. Only the second part has a formal scaling of M^5 , the other are (formal) M^4 steps. In the large system limit the transformation required for the MP2 procedure will become the most expensive step, however, in practice, where calculations may be restricted to a few hundred basis functions, it is often observed that the MP2 step takes less time than the HF. The formal scaling only indicates what the rate limiting step will be in the large system limit. Whether this limit is actually reached in practical calculations is something different.

The lower value of M^5 scaling for methods involving electron correlation arises from the transformation of the two-electron integrals from the AO to MO basis, but if the transformation is carried out with one of the indices belonging to an occupied MO first, the scaling is actually the number of occupied orbitals (O) times M^4 . If we consider making the system larger by doubling the fundamental unit (for example calculations on a series of increasingly larger water clusters), keeping the basis set per atom constant, O scales linearly with M , and we arrive at the M^5 scaling. This assumption (increasing system size) is the basis for Table 4.5. More often, however, a series of calculations are performed on the same system with increasingly larger basis sets. In this case the number of electrons (occupied orbitals) are constant and the scaling is M^4 . Many of the commonly employed methods for electron correlation (including for example MP2, MP3, MP4, CISD, CCSD and CCSD(T)) scale in fact as M^4 when the number of occupied orbitals is constant.

In terms of accuracy with a medium sized basis set the following order is often observed.

$$\text{HF} \ll \text{MP2} < \text{CISD} < \text{MP4(SDQ)} \sim \text{CCSD} < \text{MP4} < \text{CCSD(T)} \quad (4.72)$$

All of these are single determinant based methods. Multi-reference methods cannot easily be classified as the quality of the results depends heavily on the size of the reference. A two-configurational reference is only a slight improvement over HF, but including all configurations generates a full CI. The ordering above is only valid when the HF reference is a "good" zero-order description of the system. The more multi-reference character in the wave function, the better the "infinite"-order coupled cluster performs relative to perturbation methods.

MP3 has not been included in the above comparison. As already mentioned, MP3 results are often inferior to those at MP2. In fact MP2 often gives surprisingly good results, especially if large basis sets are used.⁴³ Furthermore, it should be kept in mind that the MP perturbation series in many cases may actually be divergent,

although corrections carried out to low order (i.e. 2–4) rarely display excessive oscillations.

HF results are by modern standards more and more approaching model calculations, like semi-empirical methods such as AM1 and PM3. Minimal basis HF calculations often give results which are worse than AM1 or PM3, but at a computational cost of maybe 100 times as much. Medium and large basis set HF calculations usually do not give absolute results which are particularly close to experimental values, but since the errors to a certain degree are systematic (like all vibrational frequencies being overestimated by $\sim 10\%$), they can be used with more or less “empirical” corrections to treat systems for which correlated calculations are not possible. The distinct advantage of *ab initio* methods is their ability to treat all systems at an equal level of accuracy, independently of whether experimental data exist or not. A detailed assessment of the level of accuracy that can be expected at a given level of theory is difficult to establish as it is heavily dependent on the quality of the basis set. Given a sufficiently large basis set, however, the CCSD(T) method is able to meet the goal of an accuracy of ~ 1 kcal/mol for most systems. Even with less complete methods (like MP4) and medium size basis sets such as DZP or TZ2P, it is often possible to get accuracies of the order of a few kcal/mol.

The use of CI methods has been declining in recent years, to the profit of MP and especially CC methods. It is now recognized that size extensivity is important for obtaining accurate results. Excited states, however, are somewhat difficult to treat by perturbation or coupled cluster methods, and CI or MCSCF based methods have been the preferred methods here. More recently propagator or equation of motion (Section 10.9) methods have been developed for coupled cluster wave functions, which allows calculation of excited state properties.

Finally, a few words on the size of system that can be treated. The limiting parameters will again be taken as the number of basis functions, although as noted above, a more detailed breakdown in terms of occupied and virtual MOs can be done. Note also that a given limit in terms of basis functions may translate either into a large molecular system with a small basis per atom, or a small molecular system with a very large basis set on each atom. The ordering in eq. (4.72) suggests three levels of electron correlation: none (HF), MP2 or extended (MP4 or CCSD(T)). HF methods are in general possible with up to ~ 5000 basis functions, MP2 is fairly routine with up to ~ 800 basis functions, while the advanced correlation methods are limited to ~ 300 –400 basis functions. With a DZP basis these values translate into roughly 200, 30 and 10 CH_2 fragments. The limits hold just for calculating the energy at a single geometry. If more advanced features are desired, such as optimizing the geometry or calculating frequencies, the limits drop to roughly half of the above.

With the continuing advances in computer hardware and more efficient algorithms, these limits are gradually being shifted upwards. Owing to the rather steep scaling with system size, however, they will (barring a fundamental breakthrough) give a rough idea of the size of systems which can be handled also in the future. Currently the speed of computer hardware improves by a factor of 2 in a timespan of about 18 months. In other words, a factor of 10 in terms of performance for the same price is gained roughly every five years. Due to a scaling between 4 and 7 for the different methods, however, an increase by a factor of 10 in raw speed only translates into an increase in system size of 1.7 (M^4 scaling) or 1.4 (M^7 scaling). Linear scaling methods

in Hartree–Fock methods of course will benefit fully from increased computational speed.

4.15 Excited States

The development of HF and correlated methods in the previous chapters has focused on the electronic ground state. In some cases it is also of interest to consider electronically excited states. It is useful to distinguish between two cases, depending on whether the excited state has the same or a different symmetry than the lower state(s). The different symmetry case is easy to handle, as the lowest energy state of a given symmetry may be handled completely analogously to the ground state. A HF wave function may be obtained by a proper specification of the occupied orbitals, and the resulting wave function improved by adding electron correlation by for example CI MP or CC methods. The only possible caveat may be that the state is an open shell, which often requires a (small) MCSCF wave function for an adequate zero-order description.

Excited states having lower energy solutions of the same symmetry are somewhat more difficult to treat. It will in general be difficult to generate a HF type wave function for such states, as the variational optimization will collapse to the lowest energy solution of the given symmetry. The lack of a proper HF solution means that perturbation and coupled cluster methods are not well suited to calculating excited states, although excited state properties (for example excitation energies) may be calculated directly using propagator methods (Section 10.9). Propagator methods can be based for example on a coupled cluster wave function. It is, however, relatively easy to generate higher energy states by CI methods; this simply corresponds to using the $(n + 1)$ th eigenvalue from the diagonalization of the CI matrix as a description of the n th excited state (the second root is the first excited state etc.). Such a CI procedure will normally employ a set of HF orbitals from a calculation on the lowest energy state, and the CI procedure is therefore biased against the excited states.

The simplest description of an excited state is the orbital picture where one electron has been moved from an occupied to an unoccupied orbital, i.e. an S-type determinant as illustrated in Figure 4.1. The lowest level of theory for a qualitative description of excited states is therefore a CI including only the singly excited determinants, denoted CIS. CIS gives wave functions of roughly HF quality for excited states, since no orbital optimization is involved. For valence excited states, for example those arising from excitations between π -orbitals in an unsaturated system, this may be a reasonable description. There are, however, normally also quite low-lying states which essentially correspond to a double excitation, and those require at least inclusion of the doubles as well, i.e. CISD.

A more balanced description requires MCSCF based methods where the orbitals are optimized for each particular state, or optimized for a suitable average of the desired states (*state averaged* MCSCF). It should be noted that such excited state MCSCF solutions correspond to saddle points in the parameter space for the wave function, and second-order optimization techniques are therefore almost mandatory. In order to obtain accurate excitation energies it is normally necessary to also include dynamical correlation, for example by using the CASPT2 method.

References

1. A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*, McGraw-Hill, 1982; R. McWeeny, *Methods of Molecular Quantum Mechanics*, Academic Press, 1992; W. J. Hehre, L. Radom, J. A. Pople and P. v. R. Schleyer, *Ab Initio Molecular Orbital Theory*, Wiley, 1986; J. Simons, *J. Phys. Chem.*, **95** (1991), 1017; R. J. Bartlett, J. F. Stanton, *Rev. Comput. Chem.*, **5** (1994), 65.
2. J. Olsen, P. Jørgensen, H. Koch, A. Balkova and R. J. Bartlett, *J. Chem. Phys.*, **104** (1996), 8007.
3. J. Olsen, P. Jørgensen and J. Simons, *Chem. Phys. Lett.*, **169** (1990), 463.
4. E. Davidson, *J. Comput. Phys.*, **17** (1975), 87.
5. B. O. Roos, in *Lecture Notes in Quantum Chemistry*, Ed. B. O. Roos, Springer-Verlag, 1992.
6. J. Olsen, B. O. Roos, P. Jørgensen and H. J. Aa. Jensen, *J. Chem. Phys.*, **89** (1998), 2185.
7. W. T. Borden and E. R. Davidson, *Acc. Chem. Res.*, **29** (1996), 67.
8. C. Møller and M. S. Plesset, *Phys. Rev.*, **46** (1934), 618.
9. J. Olsen, O. Christiansen, H. Koch and P. Jørgensen, *J. Chem. Phys.*, **105** (1996), 5082.
10. T. H. Dunning, Jr. and K. A. Peterson, *J. Chem. Phys.*, **108** (1998), 4761.
11. D. Cremer and Z. He, *J. Phys. Chem.*, **100** (1996), 6173.
12. O. Christiansen, J. Olsen, P. Jørgensen, H. Koch and P.-Å. Malmqvist, *Chem. Phys. Lett.*, **261** (1996), 369.
13. N. C. Handy, P. J. Knowles and K. Somasundram, *Theo. Chem. Acta*, **68** (1985), 87.
14. P. M. Kozłowski and E. R. Davidson, *J. Chem. Phys.*, **100** (1994), 3672.
15. H. B. Schlegel, *J. Phys. Chem.*, **92** (1988), 3075; P. J. Knowles and N. C. Handy, *J. Phys. Chem.*, **92** (1988), 3097.
16. B. O. Roos, K. Andersson, M. P. Fülscher, P.-Å. Malmqvist, L. Serrano-Andres, K. Pierloot and M. Merchán, *Adv. Chem. Phys.*, **93** (1996), 216.
17. R. J. Bartlett, *J. Phys. Chem.*, **93** (1989), 1697.
18. J. D. Watts and R. J. Bartlett, *Int. J. Quantum Chem.*, **S27** (1993), 51.
19. J. M. L. Martin, J. P. Francois and R. Gijbels, *Chem. Phys. Lett.*, **172** (1990), 346.
20. G. E. Scuseria and T. J. Lee, *J. Chem. Phys.*, **93** (1990), 5851.
21. K. Raghavachari, J. A. Pople, E. S. Replogle and M. Head-Gordon, *J. Phys. Chem.*, **94** (1990), 5579; R. J. Bartlett, J. D. Watts, S. A. Kucharski and J. Noga, *Chem. Phys. Lett.*, **165** (1990), 513.
22. K. A. Brueckner, *Phys. Rev.*, **96** (1954), 508; J. F. Stanton, J. Gauss and R. J. Bartlett, *J. Chem. Phys.*, **97** (1992), 5554.
23. E. A. Salter, H. Sekino and R. J. Bartlett, *J. Chem. Phys.*, **87** (1987), 502.
24. C. Hampel, K. A. Peterson and H.-J. Werner, *Chem. Phys. Lett.*, **190**, (1992), 1.
25. T. J. Lee, R. Kobayashi, N. C. Handy and R. D. Amos, *J. Chem. Phys.*, **96** (1992), 8931.
26. A. Balkova and R. J. Bartlett, *Chem. Phys. Lett.*, **193** (1992), 364.
27. J. A. Pople, M. Head-Gordon and K. Raghavachari, *J. Chem. Phys.*, **87** (1987), 5968.
28. G. E. Scuseria and H. F. Schaefer III, *J. Chem. Phys.*, **90** (1989), 3700.
29. T. J. Lee, A. P. Rendall and P. R. Taylor, *J. Phys. Chem.*, **94** (1990), 5463; for an exception see M. Böhme and G. Frenking *Chem. Phys. Lett.*, **224** (1994), 195.
30. O. Christiansen, H. Koch and P. Jørgensen, *Chem. Phys. Lett.*, **243** (1995), 409; O. Christiansen, H. Koch and P. Jørgensen, *J. Chem. Phys.*, **103** (1995), 7429.
31. J. F. Stanton, *J. Chem. Phys.*, **101** (1994), 371.
32. J. P. Watts, J. Gauss and R. J. Bartlett, *J. Chem. Phys.*, **98** (1993), 8718.
33. T. J. Lee and P. R. Taylor, *Int. J. Quant. Chem.*, **S23** (1989), 199.
34. P. G. Szalay and R. J. Bartlett, *J. Chem. Phys.*, **101** (1994), 4936.
35. T. Kato, *Commun. Pure Appl. Math.*, **10** (1957), 151.

REFERENCES

36. W. Kolos and L. Wolniewics, *J. Chem. Phys.*, **49** (1968), 404.
37. W. Kutzelnigg and J. D. Morgan III, *J. Chem. Phys.*, **96** (1992), 4484.
38. W. Kutzelnigg and W. Klopper, *J. Chem. Phys.*, **94** (1991), 1985.
39. W. Klopper, *J. Chem. Phys.*, **102** (1995), 6168.
40. M. Head-Gordon, J. A. Pople and M. J. Frisch, *Chem. Phys. Lett.*, **153** (1988), 503.
41. H. Koch, P. Jørgensen and T. Helgaker, *J. Chem. Phys.*, **104** (1996), 9528.
42. A. K. Wilson and J. Almlöf, *Theo. Chim. Acta*, **95** (1997), 49.
43. T. Helgaker, J. Gauss, P. Jørgensen and J. Olsen, *J. Chem. Phys.*, **106** (1997), 6430.