T2. CNDO to AM1: The Semiempirical Molecular Orbital Models

The discussion in sections **T2.1** – **T2.3** applies also to *ab initio* molecular orbital calculations.

T2.1 Slater Determinants

Consider the general case of the orbital approximation:

$$\Psi(1,2\ldots 2N) = \psi_1(1)\psi_2(2)\ldots\psi_{2N}(2N)$$

In the section (**T1**) on Hückel theory, we chose a restricted version in which $\psi_2 = \psi_1$, $\psi_4 = \psi_3$ … $\psi_{2N} = \psi_{2N-1}$. However, it has the problem of violating the Pauli Principle. It turned out that it didn't matter because Hückel theory doesn't know anything about spin, but for more sophisticated models that do include electron–electron interactions, we need to get this right.

The Pauli Principle derives from a requirement of statistical physics that the wavefunction for any system of Fermions (of which electrons are an example) must change sign when any two particles change places. We can see that this doesn't work if we take the simple example of a two-MO wavefunction:

$$\Psi(1,2) = \psi_1(1)\psi_2(2)$$

Interchanging electrons 1 and 2 :

$$\Psi'(1,2) = \psi_1(2)\psi_2(1) \neq -\Psi(1,2)$$

A simple solution to the problem would be to choose the expression for $\Psi(1,2)$ slightly differently. If we wrote:

$$\Psi(1,2) = \psi_1(1)\psi_2(2) - \psi_1(2)\psi_2(1)$$

Now interchanging electrons 1 and 2 gives:

$$\Psi'(1,2) = \psi_1(2)\psi_2(1) - \psi_1(1)\psi_2(2)$$
$$= -\Psi(1,2)$$

The successful wavefunction can be written as a determinant:

$$\Psi(1,2) = \begin{vmatrix} \psi_1(1) & \psi_2(1) \\ \psi_1(2) & \psi_2(2) \end{vmatrix}$$

Orbitals →

Electrons ↓

This can be generalized to our $2N$-electron wavefunction:

$$\Psi(1,2\ldots 2N) = \begin{vmatrix} \psi_1(1) & \psi_2(1) & \ldots & \psi_{2N}(1) \\ \psi_1(2) & \psi_2(2) & \ldots & \psi_{2N}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(2N) & \psi_2(2N) & \ldots & \psi_{2N}(2N) \end{vmatrix}$$

creating what is called a <u>Slater determinant</u>. Notice that we can easily show the proper behavior of the wavefunction: interchange of any pair of electrons is equivalent to interchanging two rows of the determinant, which changes its sign.   Note also that the more common form of the Pauli Principle is automatically incorporated into this formulation of the wavefunction. If we set any two MOs equal to each (e.g. $\psi_1 = \psi_2$), then two columns of the determinant will become identical and its value will consequently become zero.   In other words, the total wavefunction vanishes if we try to put two electrons in identical molecular orbitals.

We can choose our molecular orbitals to be orthonormal, i.e.:

$$\int \psi_i^* \psi_j d\tau = \delta_{ij} \quad \text{(where } \delta_{ij} \text{ is the Kronecker delta}$$

$$= 0 \text{ if } i \neq j \text{ but } 1 \text{ if } i = j)$$

but then if we also want the full Ψ to be normalized, we will need a normalization factor. This turns out to be $((2N)!)^{-1/2}$.

T2.2 Restricted Hartree Fock (RHF) Calculations

Most of the molecules on which we will want to do calculations have equal numbers of alpha- and beta-spin electrons in the lowest available energy levels. For such systems it is reasonable to use a restricted version of the Slater determinant in which pairs of electrons are assigned to MOs that have identical spatial parts but different spin parts. Thus, if we split our molecular orbitals ψ_i into spatial parts φ_i and spin parts α or β, we can write our restricted Slater determinant as:

$$\Psi(1,2,\ldots 2N) = \frac{1}{\sqrt{(2N)!}} \begin{vmatrix} \varphi_1(1)\alpha(1) & \varphi_1(1)\beta(1) & \varphi_2(1)\alpha(1) & \ldots & \varphi_N(1)\beta(1) \\ \varphi_1(2)\alpha(2) & \varphi_1(2)\beta(2) & \varphi_2(2)\alpha(2) & \ldots & \varphi_N(2)\beta(2) \\ \varphi_1(3)\alpha(3) & \varphi_1(3)\beta(3) & \varphi_2(3)\alpha(3) & \ldots & \varphi_N(3)\beta(3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varphi_1(2N)\alpha(2N) & \varphi_1(2N)\beta(2N) & \varphi_2(2N)\alpha(2N) & \ldots & \varphi_N(2N)\beta(2N) \end{vmatrix}$$

In order to find the energy levels associated with such a wavefunction, we need to calculate its expectation value, $\int \Psi^* H \Psi d\tau$. This requires that we be explicit about what H is:

$$H = \frac{-h^2}{8\pi^2 m}\sum_p \nabla_p^2 - \sum_A \sum_p \frac{e^2 Z_A}{r_{Ap}} + \sum_{A \,<\, B} \frac{e^2 Z_A Z_B}{r_{AB}} + \sum_{p \,<\, q} \frac{e^2}{r_{pq}}$$

where:
- KE of electrons: $\frac{-h^2}{8\pi^2 m}\sum_p \nabla_p^2$
- PE due to electron-nucleus attraction: $\sum_A \sum_p \frac{e^2 Z_A}{r_{Ap}}$
- PE due to nucleus-nucleus repulsion: $\sum_{A \,<\, B} \frac{e^2 Z_A Z_B}{r_{AB}}$
- PE due to electron-electron repulsion: $\sum_{p \,<\, q} \frac{e^2}{r_{pq}}$

If we use the Born-Oppenheimer approximation, i.e. treat the nuclei as static on the timescale of motion of the electrons, then the nuclei will have zero kinetic energy, and the nuclear-nuclear repulsion term will be some constant that is evaluated for a particular nuclear geometry.

Many of the physical constants in the expression for H can be made to disappear by working in <u>atomic units:</u>

The unit of mass is the rest mass of the electron, $m = 9.10953 \times 10^{-28}$ g.
The unit of charge is the absolute value of the charge on the electron, $e = 4.803242 \times 10^{-10}$ esu.
The unit of distance is the Bohr radius, $a_0 = h^2/(4\pi^2 m e^2) = 0.529177$ Å.
The unit of energy is the hartree $= e^2/a_0 = 27.2116$ eV $= 627.5095$ kcal/mol.

In atomic units, the Hamiltonian becomes:

$$H = \sum_p H_{core}(p) + \sum_{p \ < \ q} \sum \frac{1}{r_{pq}} + \sum_{A \ < \ B} \sum \frac{Z_A Z_B}{r_{AB}}$$

$$H_{core}(p) = -\frac{1}{2}\nabla_p^2 - \sum_A \frac{Z_A}{r_{Ap}}$$

If we use the usual LCAO approximation,

$$\psi_i = \sum_\mu c_{\mu i} \phi_\mu$$

where Greek subscripts are used for atomic orbitals and Roman ones for molecular orbitals, then evaluation of $\int \Psi^* H \Psi d\tau$ leads to the Hartree-Fock-Roothaan equations:

$$(\mathbf{F} - \mathbf{ES})\mathbf{C} = \mathbf{0}$$

$$F_{\mu\nu} = H_{\mu\nu} + \sum_\lambda \sum_\sigma P_{\lambda\sigma}[(\mu\nu \mid \lambda\sigma) - \frac{1}{2}(\mu\lambda \mid \nu\sigma)]$$

$$H_{\mu\nu} = \int \phi_\mu(1) H_{core}(1) \phi_\nu(1) d\tau_1$$

$$P_{\mu\nu} = 2\sum_i c_{\mu i} c_{\nu i}$$

$$(\mu\nu \mid \lambda\sigma) = \iint \phi_\mu(1)\phi_\nu(1)\frac{1}{r_{12}}\phi_\lambda(2)\phi_\sigma(2)d\tau_1 d\tau_2$$

$$(\mu\lambda \mid \nu\sigma) = \iint \phi_\mu(1)\phi_\lambda(2)\frac{1}{r_{12}}\phi_\nu(1)\phi_\sigma(2)d\tau_1 d\tau_2$$

At first sight, solution of the Hartree-Fock-Roothaan equations looks only a bit more complicated than solving the Hückel secular equations. However, the problem is that the Fock operator, **F** contains the density matrix, **P**, whose elements are the LCAO coefficients that we are trying to find. In other words, in order to evaluate the equations we need to know their solutions! The way out of this conundrum is to use an iterative procedure.

1. Calculate $S_{\mu\nu}$.
2. Calculate $H_{\mu\nu}$.
3. Diagonalize **H** and use the eigenvectors (effectively the extended Hückel wavefunctions) as the first estimates of $c_{\mu i}$.
4. Calculate $P_{\mu\nu}$ from $c_{\mu i}$.
5. Calculate $(\mu\nu|\lambda\sigma)$ and $(\mu\lambda|\nu\sigma)$. **This is by far the most time-consuming step**.
6. From 2–5 calculate $F_{\mu\nu}$.
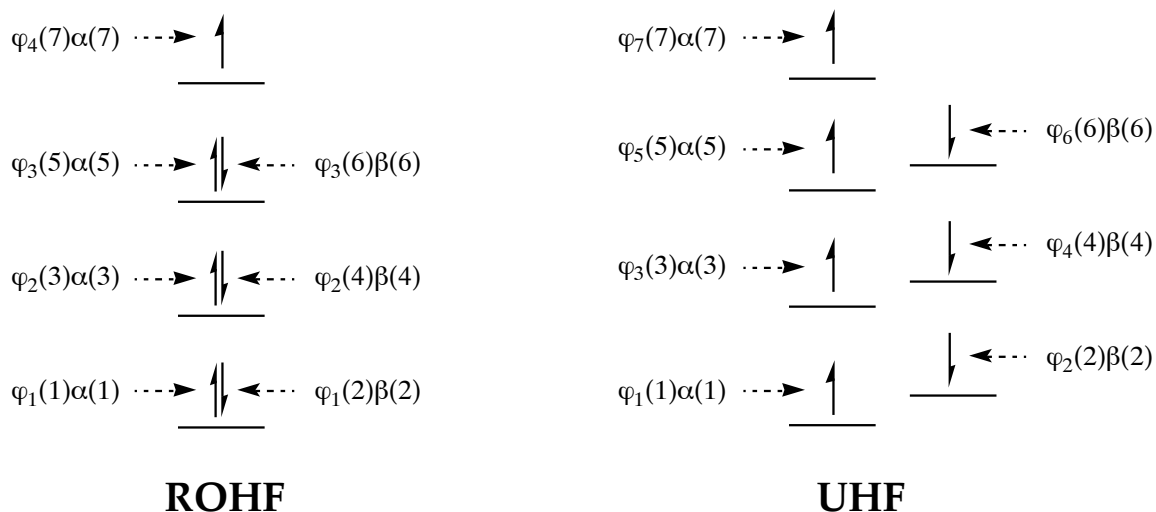7. From 1 and 6 calculate ε_i and new $c_{\mu i}$.
8. Return to step 4.

This cycle is repeated until $P_{\mu\nu}$ and/or ε_i have converged within some selected tolerance. One is then said to have achieved a self-consistent field (SCF).

T2.3 Unrestricted Hartree Fock (UHF) Calculations
The adequacy of the RHF approximation for closed-shell molecules is demonstrated by the fact that if you start with an unrestricted determinant you will invariably find the RHF wavefunction to be the lowest energy solution. This happens because the α- and β-spin electrons have symmetrical environments: in a 2N-electron molecule, every α electron is interacting with N-1 other α electrons and N β electrons. Every β electron is interacting with N-1 other β electrons and N α electrons. Thus the best MOs for the α electrons are the same as the best MOs for the β electrons.

In open-shell molecules (those having an unequal numbers of α- and β-spin electrons) this symmetry no longer holds. One can then do one of two things: either ignore the asymmetry and still use a RHF-type determinant but simply partially occupy one or more of the orbitals (this is then a ROHF or restricted, open-shell Hartree-Fock wavefunction) or, alternatively, use an unrestricted Hartree-Fock (UHF) wavefunction in which every electron is in a MO of unique spatial character.

The situation is illustrated below for a radical:

$\varphi_4(7)\alpha(7)$ $\cdots\blacktriangleright$ ↑

$\varphi_3(5)\alpha(5)$ $\cdots\blacktriangleright$ ↑↓ $\blacktriangleleft\cdots$ $\varphi_3(6)\beta(6)$

$\varphi_2(3)\alpha(3)$ $\cdots\blacktriangleright$ ↑↓ $\blacktriangleleft\cdots$ $\varphi_2(4)\beta(4)$

$\varphi_1(1)\alpha(1)$ $\cdots\blacktriangleright$ ↑↓ $\blacktriangleleft\cdots$ $\varphi_1(2)\beta(2)$

ROHF

$\varphi_7(7)\alpha(7)$ $\cdots\blacktriangleright$ ↑

$\varphi_5(5)\alpha(5)$ $\cdots\blacktriangleright$ ↑   ↓ $\blacktriangleleft\cdots$ $\varphi_6(6)\beta(6)$

$\varphi_3(3)\alpha(3)$ $\cdots\blacktriangleright$ ↑   ↓ $\blacktriangleleft\cdots$ $\varphi_4(4)\beta(4)$

$\varphi_1(1)\alpha(1)$ $\cdots\blacktriangleright$ ↑   ↓ $\blacktriangleleft\cdots$ $\varphi_2(2)\beta(2)$

UHF

The larger number of adjustable parameters for the UHF wavefunction (7 φ_i for the UHF wavefunction instead of just 4 for the ROHF, in our example) means that it is guaranteed to have an energy that is $\leq$ the energy of the ROHF solution. By the Variation Principle, the UHF solution should consequently be preferred. However, UHF wavefunctions are, in general, not eigenfunctions of the $\widehat{\mathbf{S}}^2$ operator, which proper wavefunctions should be. RHF and ROHF wavefunctions are always eigenfunctions of $\widehat{\mathbf{S}}^2$.

$$\widehat{\mathbf{S}}^2\Psi_{R(O)HF} = s(s+1)\Psi_{R(O)HF}$$

$$\widehat{\mathbf{S}}^2\Psi_{UHF} \neq s(s+1)\Psi_{UHF}$$

The scalar quantity s is just the total spin of the unpaired electrons in the system (i.e. $1/2$ for a radical 1 for a triplet state, etc.). Thus $s(s+1)$ should be 0 for a singlet-state molecule, 0.75 for a radical, and 2 for a triplet. Typically, UHF wavefunctions have $<\widehat{\mathbf{S}}^2>$ values greater than they should be because they do not correspond to pure spin states. Doublet (radical) wavefunctions will typically be contaminated with components of quartets, sextets etc., and triplet wavefunctions will typically be contaminated with components of quintets, septets, etc. The extent of the contamination can be estimated from the magnitude of $<\widehat{\mathbf{S}}^2>$. As a rough rule of thumb, if it is more than about 10% higher than the ideal value, the wavefunction and all of the properties calculated from it are likely to be unreliable.

T2.4 The Pople Semiempirical Approximations: CNDO & INDO
Pople, J.A.; Segal, G.A. *J. Chem. Phys.* **1966**, *44*, 3289.

T2.4.1 Basis Sets
All of the semiempirical methods use valence basis sets, i.e. the atomic orbitals of the LCAO approximation consist of a 1s orbital on each hydrogen and a 2s and three 2p orbitals on each element of the second row of the periodic table.

Actually, these atomic orbitals are further approximated. Because electron-electron repulsion integrals are hard to calculate with the analytical hydrogen-like atomic orbitals, they are replaced by Slater-type orbitals (STO) which lack all radial nodes.

T2.4.2 Electron Repulsion Integrals
The point of all semiempirical MO methods is to reduce the size of the full Hartree-Fock-Roothaan calculation by discarding many of the time-consuming electron-electron repulsion integrals, and by setting others equal to various constants, whose values are optimized to get the best match of the overall calculations to some chosen properties of a set of molecules.

The most severe of these kinds of approximation is the Complete Neglect of Differential Overlap (CNDO). It sets:

$$(\mu\nu \mid \lambda\sigma) = (\mu\mu \mid \lambda\lambda)\delta_{\mu\nu}\delta_{\lambda\sigma}$$
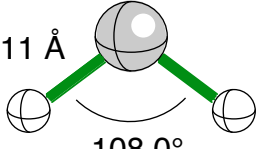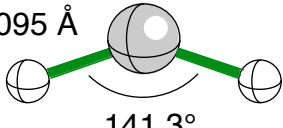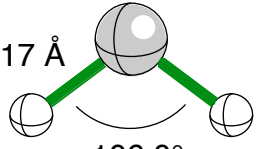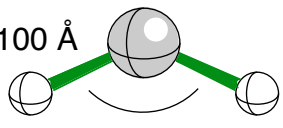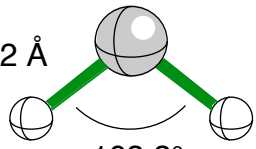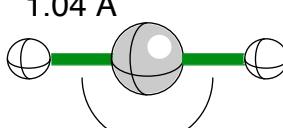
and then further approximates how the remaining integrals $(\mu\mu \mid \lambda\lambda)$ are evaluated. It also uses some empirical parameters for certain elements of the **H** matrix.

A slightly less severe approach is the Intermediate Neglect of Differential Overlap (INDO), which retains a few of the integrals discarded in CNDO – specifically so-called one-center exchange integrals of the type $(\mu\nu \mid \mu\nu)$, in which atomic orbitals ϕ_μ and ϕ_ν are on the same atom. Retaining these integrals leads to a method that does a much better job of getting relative energies of singlet and triplet states correct (*vide infra*).

A crucial philosophical point is that in CNDO and INDO, the empirical parameters are adjusted to give the best possible fit of the semiempirical calculations to the exact calculation using the same basis set. This choice is the principal difference between the Pople and Dewar semiempirical methods, and it has several subtle consequences.

T2.4.3 Results of CNDO and INDO Calculations for CH_2

In 1961 Gerhard Herzberg reported high-resolution UV spectra of CH_2 in both its singlet and triplet states. From the vibrational and rotational fine structure he was able to deduce geometries for the two states (*Proc. Roy. Soc. (London)* **1961**, *A262*, 291). Pople subsequently carried out CNDO and INDO calculations to compare with Herzberg's data. The results are summarized below:

Method	Singlet	Triplet
CNDO	1.111 Å 108.0° E = -8.1462121 hartree	1.095 Å 141.3° E = -8.0889311 hartree
INDO	1.117 Å 106.0° E = -7.8574419 hartree	1.100 Å 131.8° E = -7.8814363 hartree
Experiment	1.12 Å 103.2°	1.04 Å 180°

It looked as if both theoretical models did reasonably well on the singlet, but were badly wrong for the triplet. However, after Pople published his results, Herzberg went back to look again at his spectra and discovered that he had made a mistake. The real structure of the triplet was that the C-H bond distance was 1.08Å and the HCH angle was 134°.

This was really the first time that pure calculation had caused people to reevaluate their experimental results. As much of a triumph for theory as this may have been, it's worth noting that the energy gap between the singlet and the triplet is wildly different for the CNDO (E_S–E_T = –36 kcal/mol) and INDO (E_S–E_T = +15 kcal/mol) methods, and neither agrees with the experimental value (measured much later) of +8.5 kcal/mol.

T2.5 The Dewar Semiempirical Approximations: MINDO, MNDO, & AM1.

Starting in 1969 Michael Dewar introduced a series of differently parameterized versions of INDO called MINDO/1, /2, and /3. These Modified INDO methods had some relatively minor technical differences from Pople's INDO, but they also had one major philosophical difference. Instead of optimizing his adjustable parameters to fit the results of exact calculations, the optimizations of the MINDO methods were done to fit *experimental* heats of formation and geometries of a set of molecules. Dewar's argument for this choice was that the "exact" calculations against which CNDO and INDO were parameterized were actually far from accurate because:

1) Any method that uses the orbital approximation in the form of a single RHF Slater determinant has neglected <u>electron correlation</u>. This is the physical phenomenon by which the motions of electrons in atoms and molecules become correlated with each other in order to minimize electron-electron repulsion.
2) The basis set used to describe the MOs is too small. Although valence atomic orbitals may be appropriate for describing isolated atoms, they are not sufficient to describe the behavior of electrons in molecules because of the lack of spherical symmetry around each atom.

Dewar suggested that one might be able to compensate for the inadequate basis set and lack of explicit electron correlation by parameterizing against experimental data whose values presumably were sensitive to these things. There are a number of consequences of taking this approach. One potential concern is that the errors of Dewar's semiempirical method might be bigger than they appeared because the "calibration set" of molecules consisted of only stable "normal" species. Fitting the adjustable parameters to these might not give one a method that described reactive intermediates or transition structures well at all. In fact, though, MINDO/3 performed very well in the methylene test – it gave $r_{C-H} = 1.122$ Å and an angle of 100.2° for the singlet state and $r_{C-H} = 1.078$ Å and an angle of 134.1° for the triplet state. The triplet was calculated to be more stable than the singlet by 8.7 kcal/mol.

Despite this apparent great success of the MINDO approach, there were things that it did not do well. It made strained rings too stable, and it did poorly in describing highly polar molecules. These deficiencies were significantly reduced by adding back in some more of the electron-electron repulsion integrals that had been thrown out in CNDO, INDO, and MINDO. In the NDDO (Neglect of Diatomic Differential Overlap) approximation, integrals of the type $(\mu\nu|\lambda\sigma)$ are discarded only if ϕ_μ and ϕ_ν or ϕ_λ and ϕ_σ are on different atoms. The NDDO approximation formed the basis of Dewar's MNDO method.

Finally, Dewar added one last fix-up to MNDO to create AM1 (Austin Model 1; *J. Am. Chem. Soc.* **1985**, *107*, 3902). This involved the addition of a semiempirical equation (not based on quantum mechanics at all) to improve the description of hydrogen bonding. A later reparameterization of AM1 by J.J.P. Stewart (a one-time collaborator but subsequent competitor of Dewar's) produced the PM3 method (*J. Comput. Chem.* **1989**, *10*, 209). AM1 and PM3 are the currently most frequently used semiempirical MO methods.