

Semi-empirical Quantum Chemistry Methods

Contents

- *Hartree-Fock Theory*
- *Minimization*
- *Vibrational Frequencies*

The following section will give a brief overview of the theory underlying MOPAC, with an emphasis on enough knowledge to understand in a practical sense how MOPAC functions. This is of course a wide subject, so, for more details, please explore other resources, such as the MOPAC website at <http://www.openmopac.net> [1], or any of the excellent books on the topic.

1 Hartree-Fock Theory

MOPAC is a semiempirical code for calculating the electronic structure of molecules and solids. It is based on the Hartree-Fock method, which is an approximate method for solving the electronic wavefunction. The basic equation that MOPAC solves is the secular equation

$$C_i|F - e_iS|C_i = 0 \quad (1)$$

where C_i are the coefficients of the orbitals, F , the Fock matrix, e_i , the eigenvalues, and S , the overlap matrix. The Fock matrix depends on the orbitals. So, the above equation is a set of nonlinear equations that must be solved iteratively. The underlying idea is that each electron moves independently in the field of the remaining electrons. In each iteration, this field is adjusted and if the iterations converge to the solution, the field is self-consistent, which leads to the other name for this process: SCF, for self-consistent field.

The energy is given by

$$E = \frac{1}{2}P(H + F)$$

where P is the density matrix and H , the one-electron matrix. The general Fock matrix element is

$$F_{mn}^a = H_{mn} + S_{O}S_P [P_{mn}^{a+b}\langle mn|op\rangle - P_{mn}^a\langle mo|np\rangle] \quad (2)$$

1.1 Approximations in MOPAC

So far, these equations have been quite general. They apply to any Hartree-Fock or SCF program. Here is where MOPAC makes some approximations that are different from other programs. First, the orbitals are expanded in terms of atomic orbitals. MOPAC uses a minimum basis consisting of one atomic orbital for each orbital quantum number. So, for example, second-row elements are described by one S and three p orbitals (p_x , p_y , and p_z).

The second larger approximation is the neglect of differential overlap (NDDO). All terms arising from the overlap of two atomic orbitals that are on different centers are neglected, i.e., set to zero. In addition, all two-electron integrals involving charge clouds arising from the overlap of two atomic orbitals on different centers are ignored. This is a huge simplification. Although it may not be obvious, the two terms $\langle mn|op\rangle$ and $\langle mo|np\rangle$ are integrals over the atomic orbitals up to four centers. While there are on the order of N^2 one and

[1] <http://www.openmopac.net/>

two center integrals, there are N^4 three and four center integrals, where N is the number of atomic orbitals in the basis. For a moderate size molecule, there may be hundreds or thousands of atomic orbitals in the basis set, and N^4 becomes a very large number of integrals to compute.

MOPAC, using this concept of neglect of diatomic differential overlap, throws out this huge number of integrals. A so-called *ab initio* Hartree-Fock code calculates all of these N^4 integrals, using them to calculate the Fock matrix in Equation (2). Since MOPAC ignores all but N^2 of the integrals, it is much faster than the *ab initio* code. However, to make up for what it has thrown away, MOPAC uses parameterizations for the remaining integrals, hence the name *semiempirical*. MOPAC neither calculates everything from first principles, as an *ab initio* code would do, nor does it just use empirical correlations. It is based on the rigorous theory of the *ab initio* methods but parameterizes the integrals, hence the name *semiempirical*, though it could as well be thought of as *semi ab initio*!

1.2 Parameterizations

MOPAC supports a variety of parameterizations:

Parametrization	Description
MNDO	Modified Neglect of Differential Overlap; Dewar & Thiel [2]
AM1	Austin Model 1; Dewar, Zebisch, Healy & Stewart [3]
MNDO-d	MNDO with d orbitals; Thiel & Voityuk [4]
PM3	MNDO, Parametric Method 3; Stewart [5]
PM6	MNDO, Parametric Method 6; Stewart [6]
PM7	MNDO, Parametric Method 7; Stewart [7]
RM1	Recife Model 1; developed by Rocha, Freire, Simas & Stewart [8] in Recife, Brazil

Just for completeness, there are more approximate methods such as CNDO (Complete Neglect of Differential Overlap), which ignore all terms involving two different atomic orbitals, even if they are on one atom.

So now that we have indicated the different methods available in MOPAC, a comment: PM7 is the most recent and generally best parameterization, and as such is the default in MOPAC. PM7 is parameterized to reproduce with good accuracy the thermochemistry of organic and inorganic molecules and solids. PM6 gives slightly better thermochemistry results for organic molecules. RM1 is slightly more appropriate for electronic properties of organic and bio-organic molecules than PM7. Depending on the property and nature of the system, it is best to test which semiempirical method is the most relevant. Unless you have reason to use another parameterization, stay with PM7.

1.3 Restricted and Unrestricted Hartree-Fock

One other point to be made is the handling of spin. For a closed shell system, there is not too much to consider: each molecular orbital is occupied with two electrons, one with spin up, the other down. However,

- [2] Michael J S Dewar and Walter Thiel, "Ground States of Molecules. 38. the MNDO Method. Approximations and Parameters," *Journal of the American Chemical Society* 99, no. 15 (June 1977): 4899-4907.
- [3] Michael J S Dewar, Eve G Zebisch, Eamonn F Healy, and James J P Stewart, "AM1: a New General Purpose Quantum Mechanical Molecular Model," *Journal of the American Chemical Society* 107, no. 13 (June 1985): 3902-3909.
- [4] Walter Thiel and Alexander A Voityuk, "Extension of the MNDO Formalism to D Orbitals: Integral Approximations and Preliminary Numerical Results," *Theoretical Chemistry Accounts: Theory* 81, no. 6 (1992): 391-404. Walter Thiel and Alexander A Voityuk, "Extension of MNDO to D Orbitals: Parameters and Results for the Second-Row Elements and for the Zinc Group," *Journal of Physical Chemistry* 100, no. 2 (January 1996): 616-626.
- [5] James J P Stewart, "Optimization of Parameters for Semiempirical Methods I. Method," *Journal of Computational Chemistry* 10, no. 2 (March 1989): 209-220.
- [6] James J P Stewart, "Optimization of Parameters for Semiempirical Methods v: Modification of NDDO Approximations and Application to 70 Elements," *Journal of Molecular Modeling* 13, no. 12 (September 9, 2007): 1173-1213.
- [7] James J P Stewart, "Optimization of Parameters for Semiempirical Methods vi: More Modifications to the NDDO Approximations and Re-optimization of Parameters," *Journal of Molecular Modeling* 19, no. 1 (January 2013): 1-32.
- [8] Gerd B Rocha, Ricardo O Freire, Alfredo M Simas, and James J P Stewart, "RM1: a Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I," *Journal of Computational Chemistry* 27, no. 10 (2006): 1101-1111.

the situation is more complicated when there are unpaired electrons. Consider, for example, a triplet state, with two unpaired electrons, both with spin up. In the Hartree-Fock equations, we have two choices. We can use the same orbitals for both up and down spins when the occupancy is two electrons, and then add the two orbitals with the unpaired electrons to complete the set of occupied orbitals. Alternatively, we can allow different orbitals for the up and down spins, leading to a pair of secular equations similar to eq. (1)

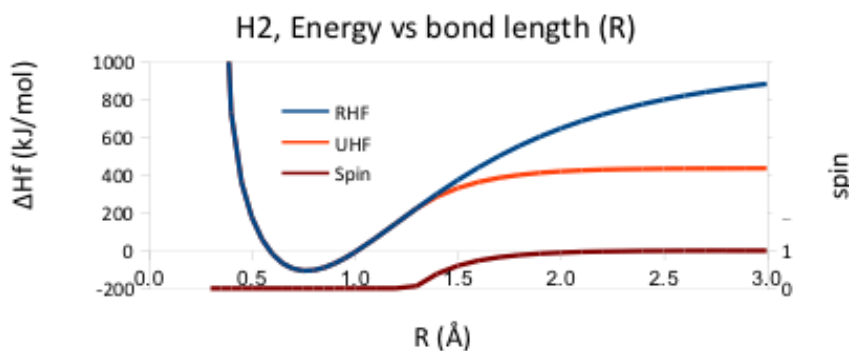
$$\begin{aligned} C_i^\alpha |F\alpha - e_i^\alpha S|C_i^\alpha &= 0 \\ C_i^\beta |F\beta - e_i^\beta S|C_i^\beta &= 0 \end{aligned} \quad (3)$$

where α and β are the two spins. These equations are coupled because the individual electrons move in the field of all the other electrons. This enters the Fock matrices through the density matrix.

The first approach, with the same orbitals for both spins, is called restricted Hartree-Fock (RHF). The second approach, with differing orbitals for different spins, is called unrestricted Hartree-Fock (UHF). UHF is simpler to work with, and has some advantages. However, it does not represent a true spin state. In our example of a triplet, the expectation value of the spin, squared, $\langle S^2 = S(S + 1) \rangle$ should be 2, but in a UHF calculation it will be larger due to contamination by the quintet state. RHF corrects this, guaranteeing that the state is a pure spin state. RHF has the disadvantage that if you stretch a bond to breaking, RHF cannot reasonably describe the unpairing of the electrons, leading to a nonphysical increasing energy for large bond distances. UHF does describe this better, with the dissociated limit usually being correct, though the energy curves at intermediate distances may still be somewhat wrong.

The graph illustrates this point for H_2 . Plotted in blue is the RHF energy vs. the bond length; in orange, UHF. At the bottom of the graph, in red, is $\langle S^2 \rangle$ for the UHF wavefunction plotted against the right axis. The UHF and RHF solutions are identical to well beyond the equilibrium bond length of about 0.75 Å. Past about 1.3 Å the UHF energy diverges from the RHF energy, and the spin differs from zero. At long distances, the UHF solution has flattened out, with a spin of 1 indicating a triplet: one electron on each H atom both with spin up. The bond energy from MOPAC using UHF is about 540 kJ/mol, which is similar to but noticeably larger than the experimental value of 440 kJ/mol.

Clearly, the RHF result is not physical.



2 Minimization

MOPAC provides different methods for minimizing the energy of a structure. The default method for smaller systems is the eigenvalue following (EF) method [9]. When minimizing the energy, the EF method is essentially a simple Newton-Raphson method with some clever schemes to make sure that it does not take a random large step. Since it is working with the Hessian matrix, the EF method is limited to systems that aren't too large - by default no more than 2000 degrees of freedom. The EF method can either start with an approximate Hessian or can calculate it, and can also be set to recalculate the Hessian every so often. Usually, for minimization, it works well enough, never calculating the Hessian that it is not worth the extra cost of calculating even the initial Hessian.

[9] Jon Baker, "An Algorithm for the Location of Transition States," Journal of Computational Chemistry 7, no. 4 (1986): 385-395.

The second major approach to minimization is the Broyden [10]-Fletcher [11]-Goldfarb [12]-Shanno [13] (BFGS) method, which is one of a class of quasi-Newton methods, where an initial approximation to the Hessian is updated as the minimization progresses, eventually leading towards quadratic convergence at the end. Again, the implementation in MOPAC uses clever techniques for the line search and for limiting the step length to make the method quite robust.

As in the EF method, the BFGS method relies on keeping a copy of the Hessian matrix in memory, limiting its applicability to large systems. MOPAC provides another variant, the limited memory BFGS (L-BFGS) method [14]. Rather than storing the complete Hessian matrix, the L-BFGS method stores information from the last several iterations and implicitly creates the Hessian from this more limited information. This is the default method in MOPAC for systems with more than 2000 degrees of freedom.

3 Vibrational Frequencies

Starting with the basics, the vibration of a mass on a spring is given by

$$r = 2\pi \sqrt{\frac{m}{k}}$$

where r is the period of the oscillator, m is the mass and k is the force constant. In chemistry, we tend to express the frequency in terms of wavenumbers, i.e., the number of wavelengths in 1 cm:

$$\bar{\nu} = \frac{1}{2\pi c} \sqrt{\frac{k}{\mu}} \quad (4)$$

where C is the speed of light in the appropriate units.

When we generalize this to a molecule, we need the Hessian, or force-constant matrix:

$$H_{i,j} = \frac{\partial^2 E}{\partial x_i \partial x_j}$$

where E is the energy and x_i and x_j are the (Cartesian) coordinates of the atoms. MOPAC calculates the Hessian by finite differences of the first derivatives, which it can calculate analytically. To calculate the vibrational frequencies, we need the mass-weighted Hessian:

$$H_{ij}^* = \frac{H_{ij}}{\sqrt{M_i M_j}}$$

This is then diagonalized to produce the eigenvalues ϵ_i and the eigenvectors, which are the mass-weighted normal modes. The frequencies are obtained from an equation analogous to Equation (4):

$$\bar{\nu} = \frac{1}{2\pi c} \sqrt{\epsilon_i}$$

At this point, MOPAC has calculated both the frequencies and the normal modes of the molecule.

[10] C G Broyden, "The Convergence of a Class of Double-Rank Minimization Algorithms," *IMA Journal of Applied Mathematics* 6, no. 3 (1970): 222-231.

[11] R Fletcher, "A New Approach to Variable Metric Algorithms," *The Computer Journal* 13, no. 3 (March 1, 1970): 317-322.

[12] Donald Goldfarb, "A Family of Variable-Metric Methods Derived by Variational Means," *Mathematics of Computation* 24, no. 109 (January 1, 1970): 23-23.

[13] DF Shanno, "Conditioning of Quasi-Newton Methods for Function Minimization," *Mathematics of Computation* 24, no. 111 (1970): 647-656.

[14] Hermann Matthies and Gilbert Strang, "The Solution of Nonlinear Finite Element Equations," *International Journal for Numerical Methods in Engineering* 14, no. 11 (1979): 1613-1626. Jorge Nocedal, "Updating Quasi-Newton Matrices with Limited Storage," *Mathematics of Computation* 35, no. 151 (July 1980): 773-782.