Computational Chemistry and Materials Modeling

Lecture Semiempirical approaches

Dmitry Aksenov (lead instructor) Sergey Levchenko (co-instructor) Alexander Kvashnin (co-instructor) Alexander Shapeev (co-instructor) Arseniy Burov (TA)

The electronic structure problem: overview $\left[-\frac{1}{2}\sum_{i}\nabla_{i}^{2}-\sum_{iA}\frac{Z_{A}}{r_{iA}}+\sum_{i>i}\frac{1}{r_{ii}}\right]\psi_{e}(\mathbf{r};\mathbf{R})=E_{e}\psi_{e}(\mathbf{r};\mathbf{R})$

- A system of fermions interacting via Coulomb potential in electrostatic field of nuclei
- •**Problem 1** one-electron problem: use finite basis set (atomic-like orbitals STO/GTO or plane waves)
- Problem 2 many-body problem: use mean field (HF, DFT, TDDFT) and perturbation theories (MP2, CI, CC) in Fock space (basis of Slater determinants)

Method	Explicit e-correlations	Wave-function	Cost (PC)
Ab initio	All	Exact	Large
(MP2, CI, CAS-CI, CC-EOM)	(depends on level of theory)	(for given basis set)	(≥10 electrons)
Density Functional	Dynamic only	Kohn-Sham	Significant
(DFT, TDDFT)		(a single-det. "fit" to e-density)	(≤1000 atoms)
Semiempirical	Coulomb, exchange, static	Hartree-Fock	Low
(AM1, PM7, ZINDO)		(variationally optim. single-det.)	(≤10 000 atoms)
Tight-binding	Νο	One-electron	Approach MM
(Huckel, Frenkel, DFTB)		(total e-energy is inaccurate)	(>10 000 atoms)

Computational Chemistry and Materials Modeling

What is our basis set? Atomic orbitals and the LCAO approach

Basis set = the 'room' for electrons to occupy!

This is a fundamental approximation in ALL electronic structure calculations!

The foundation of every electronic structure code = only one 'type' of the basis set is used in the program

The LCAO concept: construct a guess wave function ϕ as a linear combination of known atomic wave functions φ_i

The choice of the basis set type is just a balance between numerical efficiency and physical nature of the electronic wavefunctions to be described!

Common examples: Slater (exponential), Gaussian, polynomial, plane waves, wavelets, cube functions, ...

$$\phi = \sum_{i=1}^{N} a_i \varphi_i$$

A T

Figure 4.1 (from Cramer) Two different basis sets for representing a C–H σ bonding orbital with the size of the basis functions roughly illustrating their weight in the hybrid MO. Distinguishes where an orbital is centered.

Solving one-electron problem:

N roots: eigenvalues (E) and eigenvectors (a,) defining molecular orbitals (MO)

For one electron system the lowest energy molecular orbital is the ground state and the higher are excited states

$$\phi_j = \sum_{i=1}^N a_{ij} \varphi_i$$

Computational Chemistry and Materials Modeling

Atomic vs molecular orbitals, spin orbital

Atomic orbital (φ) - is a wave function for a single electron in atom.

Molecular orbital (ψ) - in molecule, can be represented as a combination of atomic orbitals $\psi_i(r)$ is spatial orbital

$$\int d\mathbf{r} \,\psi_i^*(\mathbf{r})\psi_j(\mathbf{r}) = \delta_{ij} \quad \text{Orthonormal basis set}$$

$$\chi(\mathbf{x}) = \begin{cases} \psi(\mathbf{r})\alpha(\omega) \\ \text{or} \\ \psi(\mathbf{r})\beta(\omega) \end{cases}$$

 χ is the wave function describing both spatial distribution and spin orbitals α (spin up) and β (spin down) ω is unspecified spin variable

$$\int d\mathbf{x} \, \chi_i^*(\mathbf{x}) \chi_j(\mathbf{x}) = \langle \chi_i \big| \chi_j \rangle = \delta_{ij}$$

Example 1: Slater-type orbitals (STO)

 $\varphi(r,\theta,\phi;\zeta,n,l,m) = \frac{(2\zeta)^{n+1/2}}{[(2n)!]^{1/2}} r^{n-1} e^{-\zeta r} Y_l^m(\theta,\phi)$

Atom-centered polar coordinates: ζ is a Slater exponent (Slater rules, 1930) that depend on the atomic number, *n* is the principal quantum number for the valence orbital, and the spherical harmonic functions $Y_m'(\theta, \varphi)$, depending on the angular momentum quantum numbers I and *m*, (from solution of the Schrödinger equation for the hydrogen atom)

Advantages: chemically intuitive, physically transparent, 'tails' of the wavefunctions are important

Disadvantages: numerically difficult (2e integrals need to be evaluated numerically) Where used: small molecules semiempirical approaches (few 2e integrals), density functional theory (kernels without exact exchange) **Package:** ADF (Amsterdam Density Functional)

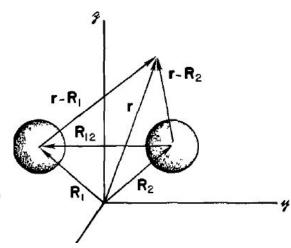
Example: H₂ molecule (from Szabo)

$$\phi(\mathbf{r} - \mathbf{R}) = (2\alpha/\pi)^{3/4} e^{-\alpha|\mathbf{r} - \mathbf{R}|^2}$$
(1s orbital fo

$$S_{12} = \int d\mathbf{r} \ \phi_1^*(\mathbf{r})\phi_2(\mathbf{r})$$
(overlap)

 $\psi_1 = [2(1 + S_{12})]^{-1/2}(\phi_1 + \phi_2)$ $\psi_2 = [2(1 - S_{12})]^{-1/2}(\phi_1 - \phi_2)$ rH)

(bonding and antibonding MOs, the HF solutions for H_2)



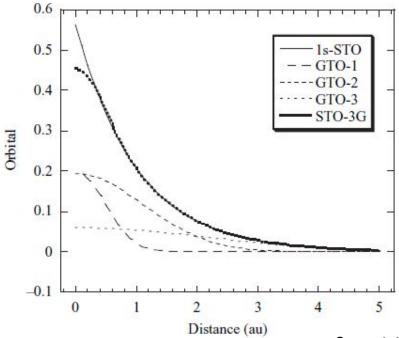
Example 2: Gaussian-type orbitals (GTO)

$$\phi(x, y, z; \alpha, i, j, k) = \left(\frac{2\alpha}{\pi}\right)^{3/4} \left[\frac{(8\alpha)^{i+j+k}i!j!k!}{(2i)!(2j)!(2k)!}\right]^{1/2} x^i y^j z^k e^{-\alpha(x^2+y^2+z^2)}$$

Idea: Fit Slater-type atomic orbitals with a superposition of Gaussian orbitals

$$\varphi(x, y, z; \{\alpha\}, i, j, k) = \sum_{a=1}^{M} c_a \phi(x, y, z; \alpha_a, i, j, k)$$

Generally 3 GTO fit well 1 STO:



Advantages: chemically intuitive, physically transparent for finite size molecules, numerically easy (2e integrals are evaluated analytically) Disadvantages: 'tails' of the wavefunctions are 'too short', no 'cusp' of the wavefunction near nuclei, issues with over-completeness and orthogonalization in extended sets Where used: majority of electronic structure modeling (both wavefunction and DFT methods) in the finite size molecules Package: Gaussian, Turbomole, Q-Chem, GAMESS, Firefly, etc.

Evaluation of integrals is analytic:

$$\phi_s^{GF}(\alpha, \mathbf{r} - \mathbf{R}_{\mathbf{A}}) \ \phi_s^{GF}(\beta, \mathbf{r} - \mathbf{R}_{\mathbf{B}}) = K_{AB}\phi_s^{GF}(p, \mathbf{r} - \mathbf{R}_{\mathbf{P}})$$
$$m_A k_B |n_C l_D\rangle = K_{AB} K_{CD} \int d\mathbf{r}_1 d\mathbf{r}_2 \phi_s^{GF}(p, \mathbf{r}_1 - \mathbf{R}_{\mathbf{P}}) r_{12}^{-1} \phi_s^{GF}(q, \mathbf{r}_2 - \mathbf{R}_{\mathbf{Q}})$$

Computational Chemistry and Materials Modeling

GTO extravaganza

-Contracted basis set: GTO basis function is a FIXED superposition of primitive GFs
-Minimum basis set: the smallest BS able only to host electrons on an atom
-Double zeta (DZ) basis set: doubling of all basis functions (tighter and diffuse exp)
-Split valence basis set: doubling of all basis functions only on valence orbitals
-Triple Zeta (TZ), triple split valence, Quadruple Zeta (QZ), Pentuple Zeta (PZ)....
-Adding polarization (e.g. TZP) and diffuse functions

Basis	Hydrogen		First row elements		Second row elements	
	Contracted	Primitive	Contracted	Primitive	Contracted	Primitive
STO-3G	1s	38	2s1p	6s3p	3s2p	9s6p
3-21G	2s	3s	3s2p	6s3p	4s3p	9s6p
6-31G(d,p)	2s1p	4s	3s2p1d	10s4p	4s3p1d	16s10p
6-311G(2df,2pd)	3s2p1d	5s	4s3p2d1f	11s5p	6s4p2d1fa	13s9p ^a

Table 5.1 (from Jensen): The Pople-style basis sets

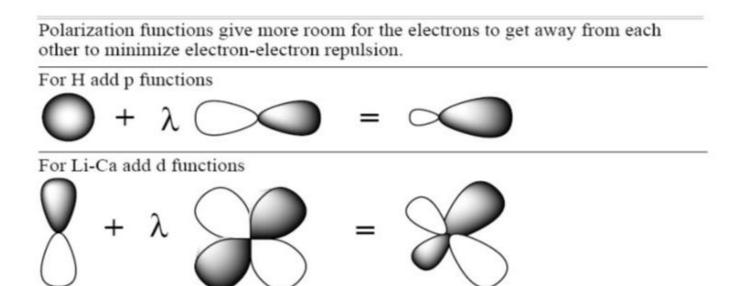
Table 5.2 (from Jensen): The Ahlrichs type basis sets

Basis	Hydrogen		First row elements		Second row elements	
	Contracted	Primitive	Contracted	Primitive	Contracted	Primitive
SVP	2s1p	4s	3s2p1d	7s4p	4s3p1d	10s7p
TZV	3s2p1d	5s	5s3p2d1f	11s6p	5s4p2d1f	14s9p
QZV	4s3p2d1f	7s	7s4p3d2f1g	15s8p	9s6p4d2f1g	20s14p

Polarization and diffuse functions

• The orbitals can distort and adapt better to the molecular environment.

(Example) Double-Zeta Polarization (DZP) or Split-Valence Polarization (SVP) $6-31G(d,p) = 6-31G^{**}, 6-31G(d) = 6-31G^{*}$ (Pople)



Diffuse functions (denoted by +) are "large" s and p orbitals for "diffuse electrons" such as lone pairs, anions, excited states

Lets consider 6-31+G* Popple basis set for carbo

C: $1s^22s^22p_x^22p_y^22p_z^2$

Inner shell (each 6 GF) 1s

Valence shell $2s 2p_x 2p_y 2p_z$ (2 sets, each 3 GF) $2's 2'p_x 2'p_y 2'p_z$

Diffuse functions: + (next shell, each 3 GF)

3s 3p_x 3p_y 3p_z

Polarized functions: * $3d_{xx} 3d_{yy} 3d_{zz}$ (next shell, each 3 GF) $3d_{xy} 3d_{xz} 3d_{yz}$

Total 19 basis functions

Many-electron Wave Functions: Hartree product

Huckel theory example is a *one-electron formalism*, though the interaction is taken into account in some average way due to relying on experimental data the approach is crude.

Let's assume that the Hamiltonian is separable Hartree product:

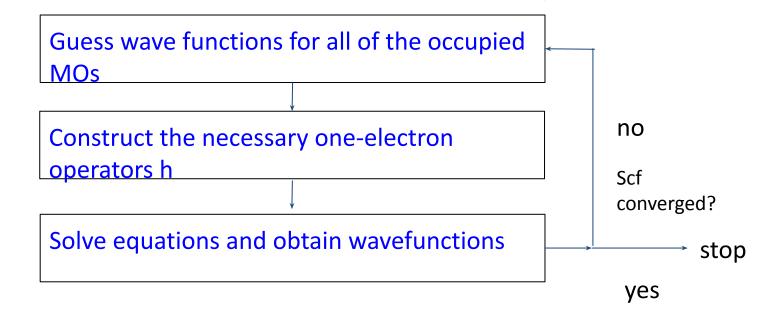
$$H = \sum_{i=1}^{N} h_i \qquad h_i = -\frac{1}{2} \nabla_i^2 - \sum_{k=1}^{M} \frac{Z_k}{r_{ik}} \qquad h_i \psi_i = \varepsilon_i \psi_i$$

$$\Psi_{\rm HP}=\psi_1\psi_2\cdots\psi_N$$

The Hartree Hamiltonian and self-consistent field (SCF) method

$$h_i = -\frac{1}{2}\nabla_i^2 - \sum_{k=1}^M \frac{Z_k}{r_{ik}} + V_i\{j\} \quad V_i\{j\} = \sum_{j \neq i} \int \frac{\rho_j}{r_{ij}} d\mathbf{r} \quad \rho_j = |\psi_j|_{\tau}^2$$
$$h_i \psi_i = \varepsilon_i \psi_i$$

How to use individual orbitals in Hamiltonian before they are known?



Double counting

 h_i includes the repulsion between electron i and electron j, but so too does h_j , sum all of the one-electron eigenvalues for the operators *hi* double-count the electron–electron repulsion.

To correct we should subtract double counting:

$$E = \sum_{i} \varepsilon_{i} - \frac{1}{2} \sum_{i \neq j} \int \int \frac{|\psi_{i}|^{2} |\psi_{j}|^{2}}{r_{ij}} d\mathbf{r}_{i} d\mathbf{r}_{j}$$

Coulomb integral, J_{ij}

Electron spin and antisymmetry

Hartree product for a triplet state (two spins in parallel), electron 1 in state *a* and electron 2 in state b ${}^{3}\Psi_{\rm HP} = \psi_{a}(1)\alpha(1)\psi_{b}(2)\alpha(2)$

Lets apply permutation operator P_{12} :

 $P_{12}[\psi_a(1)\alpha(1)\psi_b(2)\alpha(2)] = \psi_b(1)\alpha(1)\psi_a(2)\alpha(2)$

 $\neq -\psi_a(1)\alpha(1)\psi_b(2)\alpha(2)$

According to Pauli principle the wave functions must change sign whenever the coordinates of two electrons are interchanged, but this condition is not satisfied for Hartree product, but it is satisfied for

$${}^{3}\Psi_{\rm SD} = \frac{1}{\sqrt{2}} [\psi_{a}(1)\alpha(1)\psi_{b}(2)\alpha(2) - \psi_{a}(2)\alpha(2)\psi_{b}(1)\alpha(1)]$$

$${}^{3}\Psi_{\rm SD} = \frac{1}{\sqrt{2}} \begin{vmatrix} \psi_{a}(1)\alpha(1) & \psi_{b}(1)\alpha(1) \\ \psi_{a}(2)\alpha(2) & \psi_{b}(2)\alpha(2) \end{vmatrix}$$
The permittion of the permittio

The permutation operator interchanges two rows, determinant changes sign

Every electron appears in every spin orbital somewhere in the expansion. This is a manifestation of the indistinguishability of quantum particles

Quantum mechanical exchange

$$\int {}^{3}\Psi_{\rm SD} \frac{1}{r_{12}} {}^{3}\Psi_{\rm SD} d\mathbf{r}_{1} d\omega_{1} d\mathbf{r}_{2} d\omega_{2} = \frac{1}{2} \left[\int |\psi_{a}(1)|^{2} \frac{1}{r_{12}} |\psi_{b}(2)|^{2} d\mathbf{r}_{1} d\mathbf{r}_{2} \right. \\ \left. - 2 \int \psi_{a}(1)\psi_{b}(1) \frac{1}{r_{12}}\psi_{b}(2)\psi_{a}(2)d\mathbf{r}_{1} d\mathbf{r}_{2} \right. \\ \left. + \int |\psi_{a}(2)|^{2} \frac{1}{r_{12}} |\psi_{b}(1)|^{2} d\mathbf{r}_{1} d\mathbf{r}_{2} \right] \\ \left. = \frac{1}{2} \left(J_{ab} - 2 \int \psi_{a}(1)\psi_{b}(1) \frac{1}{r_{12}}\psi_{a}(2)\psi_{b}(2)d\mathbf{r}_{1} d\mathbf{r}_{2} + J_{ab} \right) \\ \left. = J_{ab} - K_{ab} \right]$$

 K_{ab} is exchange integral. Coulomb repulsion between the electron clouds in orbitals a and b is reduced by K_{ab}

One-electron Fock operator

$$f_i = -\frac{1}{2}\nabla_i^2 - \sum_k^{\text{nuclei}} \frac{Z_k}{r_{ik}} + V_i^{\text{HF}}\{j\} \text{ where } V_i^{\text{HF}} = 2J_i - K_i$$

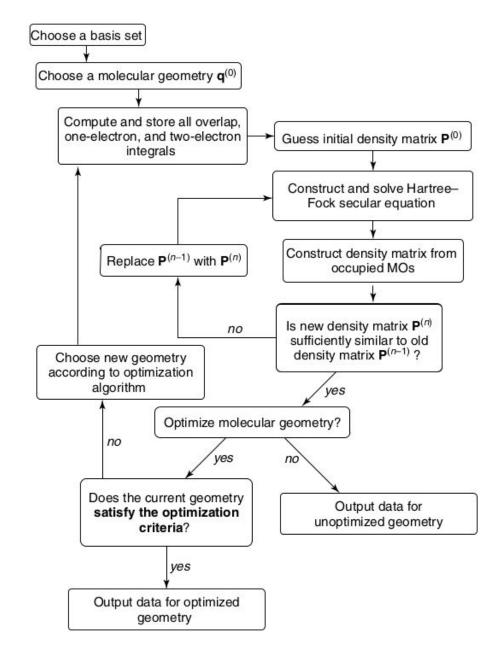
To find MO and roots we solve secular equation, F and S are computed explicitly:

$$\begin{vmatrix} F_{11} - ES_{11} & F_{12} - ES_{12} & \cdots & F_{1N} - ES_{1N} \\ F_{21} - ES_{21} & F_{22} - ES_{22} & \cdots & F_{2N} - ES_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ F_{N1} - ES_{N1} & F_{N2} - ES_{N2} & \cdots & F_{NN} - ES_{NN} \end{vmatrix} = 0$$

$$F_{\mu\nu} = \left\langle \mu \left| -\frac{1}{2} \nabla^2 \right| \nu \right\rangle - \sum_{k}^{\text{nuclei}} Z_k \left\langle \mu \left| \frac{1}{r_k} \right| \nu \right\rangle \quad \begin{array}{l} \text{here Greek letters} \\ \text{define basis} \\ \text{functions} \\ + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[(\mu\nu|\lambda\sigma) - \frac{1}{2} (\mu\lambda|\nu\sigma) \right] \quad P \text{ is density matrix} \end{array}$$

One electron integrals: Two electron integrals, N⁴ (here fi is probability density!): $\langle \mu | g | \nu \rangle = \int \phi_{\mu}(g\phi_{\nu}) d\mathbf{r}$ $(\mu \nu | \lambda \sigma) = \int \int \phi_{\mu}(1)\phi_{\nu}(1) \frac{1}{r_{12}}\phi_{\lambda}(2)\phi_{\sigma}(2)d\mathbf{r}(1)d\mathbf{r}(2)$

Flow chart of the HF SCF procedure



Semiempirical QM: Philosophy

1. Numerical approximations

HF is a challenge, the number of four index integrals is N⁴

Therefore the idea is to make it more tractable despite some loss in accuracy

2. Chemically virtuous approximation

HF is called *ab initio* or *first-principles* approach, since no empirical information is used. However, the mean-field nature not all electron-electron interactions are taken into account. $E_{HF} > H_{exact}$; $E_{cor} = H_{exact} - E_{HF}$. An approximation can introduce correlation improving chemical accuracy

3. Approximations have parametric form, parameters are chosen so as best to reproduce experiment, therefore methods are called semiempirical

4. Since large molecules are of high interest, the semiempirical methods continue to develop

Extended Huckel theory

Slater orbitals are used to compute overlap integrals
 Diagonal H is the same, but the non diagonal is additionally considered

$$\begin{array}{cccccc} H_{11} - ES_{11} & H_{12} - ES_{12} & \dots & H_{1N} - ES_{1N} \\ H_{21} - ES_{21} & H_{22} - ES_{22} & \dots & H_{2N} - ES_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{N1} - ES_{N1} & H_{N2} - ES_{N2} & \dots & H_{NN} - ES_{NN} \end{array} \right| = 0$$

$$\varphi(r,\theta,\phi;\zeta,n,l,m) = \frac{(2\zeta)^{n+1/2}}{[(2n)!]^{1/2}} r^{n-1} e^{-\zeta r} Y_l^m(\theta,\phi)$$

$$H_{\mu\nu} = \frac{1}{2} C_{\mu\nu} (H_{\mu\mu} + H_{\nu\nu}) S_{\mu\nu}$$

Complete neglect of differential overlap (CNDO)

One STO per valence orbital

 $S_{\mu\nu} = \delta_{\mu\nu}$ $(\mu\nu|\lambda\sigma) = \delta_{\mu\nu}\delta_{\lambda\sigma}(\mu\mu|\lambda\lambda)$ only the same basis orbitals on atom survive $(\mu\mu|\lambda\lambda) = \gamma_{AB}$ where A and B $\gamma_{AA} = IP_A - EA_A$ $\gamma_{AB} = \frac{\gamma_{AA} + \gamma_{BB}}{2 + r_{AB}(\gamma_{AA} + \gamma_{BB})}$ are different atoms $\left\langle \mu \left| -\frac{1}{2} \nabla^2 - \sum_{k} \frac{Z_k}{r_k} \right| \mu \right\rangle = -\mathrm{IP}_{\mu} - \sum_{k} (Z_k - \delta_{Z_A Z_k}) \gamma_{Ak}$ $\left\langle \mu \left| -\frac{1}{2} \nabla^2 - \sum_{k} \frac{Z_k}{r_k} \right| \nu \right\rangle = \frac{(\beta_{\rm A} + \beta_{\rm B}) S_{\mu\nu}}{2}$

Intermediate neglect differential overlap

INDO - Intermediate neglect differential overlap, relax constraints on 1c2e integrals,

NDDO and MNDO - (modified) neglect of diatomic differential overlap, relax constraints on 2c2e integrals,

AM1 Austin Model 1: (Dewar 1985)

$$V_N(\mathbf{A}, \mathbf{B}) = V_{AB}^{\text{MNDO}} + \frac{Z_A Z_B}{r_{AB}} \sum_{i=1}^4 \left[a_{A,i} e^{-b_{A,i} (r_{AB} - c_{A,i})^2} + a_{B,i} e^{-b_{B,i} (r_{AB} - c_{B,i})^2} \right]$$

PM3, PM7 - Parameterized Model 3, 7 (Stewart 1989), very similar to AM1 but the algorithm of parametrization is different

Accuracy: heat formation

Table 5.2 Mean unsigned errors (kcal mol^{-1}) in predicted heats of formation from basic NDDOmodels

Elements (number)	Subset (number)	MNDO	AM1	PM3	MNDO/d
Lighter (181)		7.35	5.80	4.71	
	CH (58)	5.81	4.89	3.79	
	CHN (32)	6.24	4.65	5.02	
	CHNO (48)	7.12	6.79	4.04	
	CHNOF (43)	10.50	6.76	6.45	
	Radicals (14)	9.3	8.0	7.4	
Heavier (488)	(187) (187)	29.2	15.3	10.0	4.9
	Al (29)	22.1	10.4	16.4	4.9
	Si (84)	12.0	8.5	6.0	6.3
	P (43)	38.7	14.5	17.1	7.6
	S (99)	48.4	10.3	7.5	5.6
	Cl (85)	39.4	29.1	10.4	3.9
	Br (51)	16.2	15.2	8.1	3.4
	I (42)	25.4	21.7	13.4	4.0
	Hg (37)	13.7	9.0	7.7	2.2
	Normal (421)	11.0	8.0	8.4	4.8
	Hypervalent (67)	143.2	61.3	19.9	5.4
Cations (34)		9.55	7.62	9.46	
Anions (13)		11.36	7.11	8.81	

Accuracy: geometry and charges

- Dewar, Jie, and Yu (1993) evaluated AM1 and PM3 for 344 bond lengths and 146 valence angles in primarily organic molecules composed of H, C, N, O, F, Cl, Br, and I; the average unsigned errors were 0.027 and 0.022 A°, respectively, for the bond lengths, and 2.3 and 2.8°, respectively, for the angles.
- Stewart (1991) performed a similar analysis for a larger set of molecules, some of them including Al, Si, P, and S. For 460 bond lengths, the mean unsigned errors were 0.054, 0.050, and 0.036 A° for MNDO, AM1, and PM3, respectively.
- For a 125-molecule test set including H, C, N, O, F, Al, Si, P, S, Cl, Br, and I functionality, Stewart found mean unsigned errors in dipole moments of 0.45, 0.35, and 0.38 D, respectively, for MNDO, AM1, and PM3 (Stewart 1989)

Individual studies:

• *Reading. Required: Cramer (Ch. 4 and Ch. 5) Additional:*