

Introduction to Semiempirical Methods

C. David Sherrill

School of Chemistry and Biochemistry

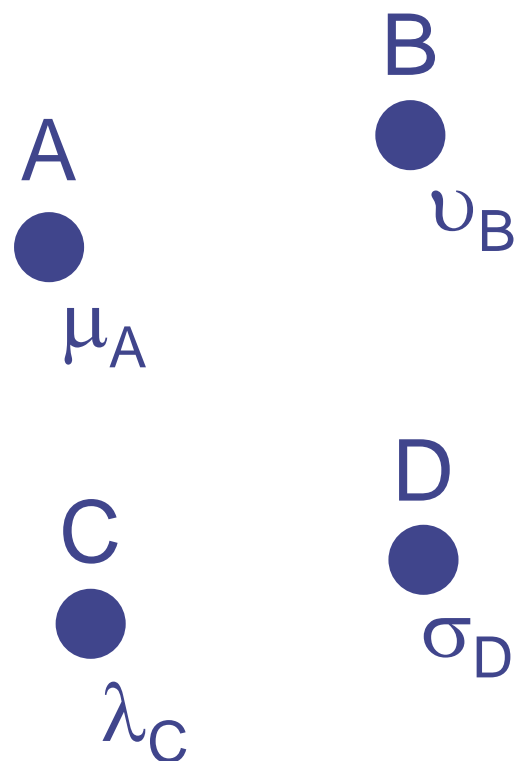
Georgia Institute of Technology

Semiempirical Methods

- ◆ Approximate version of Hartree-Fock
- ◆ Some two-electron and sometimes one-electron integrals are neglected to speed up the computation
- ◆ Some empirical parameters are inserted to make up for the neglected integrals

Eliminating Integrals

- ◆ Hartree-Fock formally scales as $O(N^4)$ because of the two-electron integrals (although this approaches $O(N^2)$ with Schwarz screening, and ultimately $O(N)$ for large molecules with multipole methods)
- ◆ Semiempirical methods reduce cost by eliminating many integrals



Atomic orbital μ_A is centered on atom A, etc.

Approximations in Semiempirical Methods

- ◆ Only valence electrons considered: core is treated by reducing nuclear charge or by adding special core functions
- ◆ Only a minimum basis is used (one AO basis function per real AO in atom): usually Slater functions are used!
- ◆ Various approximations for one- and two-electron integrals

Zero Differential Overlap (ZDO)

- ◆ ZDO is a common approximation to many semi-empirical methods
- ◆ Set the product $\mu_A(r)\nu_B(r)=0$ for μ, ν on different atoms ($A \neq B$)
- ◆ For $(\mu\nu|\lambda\sigma)$ this means all 3 & 4 center 2-electron integrals (most numerous ones) are neglected
- ◆ Get away with this by introducing parameters into (some of the) remaining integrals

Neglect of Diatomic Differential Overlap (NDDO)

- ◆ Semiempirical method based on ZDO
- ◆ Nuclear charge reduced by # of core electrons
- ◆ For two-electron integrals,
$$(\mu_A \nu_B | \lambda_C \sigma_D) = \delta_{AB} \delta_{CD} (\mu_A \nu_B | \lambda_C \sigma_D)$$

(i.e., ZDO)
- ◆ For one-electron integrals, neglect those involving 3 centers (see next page)

Neglect of 3-center 1-electron Integrals in NDDO

$$\hat{h} = -\frac{1}{2}\nabla^2 - \sum_A \frac{Z'_A}{|\vec{R}_A - \vec{r}|} = -\frac{1}{2}\nabla^2 - \sum_A V_A$$

$$\langle \mu_A | \hat{h} | \nu_B \rangle = \langle \mu_A | -\frac{1}{2}\nabla^2 - V_A | \nu_A \rangle - \sum_{B \neq A} \langle \mu_A | V_B | \nu_A \rangle$$

$$\langle \mu_A | \hat{h} | \nu_B \rangle = \langle \mu_A | -\frac{1}{2}\nabla^2 - V_A - V_B | \nu_B \rangle$$

$$\langle \mu_A | V_C | \nu_B \rangle = 0$$

Intermediate Neglect of Differential Overlap (INDO)

- ◆ Similar to NDDO, but also neglect two-center integrals in which the AO's on an atom are different; only keep $(\mu_A \mu_A | \nu_B \nu_B)$ 2-center ints and $(\mu_A \nu_A | \lambda_A \sigma_A)$ 1-center ints
- ◆ Same approximations for 1-electron integrals as NDDO

Complete Neglect of Differential Overlap (CNDO)

- ◆ Most drastic of the NDDO, INDO, CNDO series
- ◆ Only keep two-electron integrals of the type
 $(\mu_A \mu_A | \nu_B \nu_B)$ 2-center ints and
 $(\mu_A \mu_A | \nu_A \nu_A)$ 1-center ints

Parameterization of Semiempirical Methods

For integrals that are kept, some of them are computed exactly, and others are computed using parameters from experiment (semi-empirical)

Modified Intermediate Neglect of Differential Overlap (MINDO)

$$\begin{aligned}\langle \mu_A | \hat{h} | \nu_B \rangle &= \langle \mu_A | -\frac{1}{2} \nabla^2 - V_A | \nu_A \rangle - \sum_{B \neq A} \langle \mu_A | V_B | \nu_A \rangle \\ &= \langle \mu_A | \nu_B \rangle \beta_{AB} (I_\mu + I_\nu)\end{aligned}$$

- ❖ I_μ is the ionization potential of an electron in atomic orbital μ
- ❖ β_{AB} is a fitted "diatomic" parameter

Modified Intermediate Neglect of Differential Overlap (MINDO)

MINDO has been largely superseded by NDDO models based on *atomic parameters*, such as MNDO, AM1, PM3, which differ only in their treatment of core and how parameters are assigned

$$\begin{aligned}\langle \mu_A | \hat{h} | \nu_B \rangle &= \langle \mu_A | -\frac{1}{2} \nabla^2 - V_A | \nu_A \rangle - \sum_{B \neq A} \langle \mu_A | V_B | \nu_A \rangle \\ &= \langle \mu_A | \nu_B \rangle \frac{1}{2} (\beta_\mu + \beta_\nu)\end{aligned}$$

Modified Neglect of Diatomic Overlap (MNDO)

- ◆ Atomic parameterization, modified NDDO type model
- ◆ Largely superseded by AM1, PM3, etc.
- ◆ Some limitations:
 - Steric crowding exaggerated
 - H-bonds not reliable
 - Hypervalent molecules too unstable
 - Bond breaking/forming TS's too high in energy

Austin Model 1 (AM1)

- ◆ Changed core functions relative to MNDO, reparameterized
- ◆ Improved over MNDO:
 - Alkyl groups too stable by ~ 2 kcal/mol per CH_2
 - Peroxide compounds not described well

M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **107**, 3902-3909 (1985)

MNDO Parametric Method 3 (PM3)

- ◆ Reoptimized parameters automatically (previously done by hand) and changed core terms again
- ◆ Quite a few atoms available
- ◆ Performance:
 - H-bonds too short by $\sim 0.1 \text{ \AA}$
 - N atom charges often unreasonable

Common Problems for MNDO, AM1, PM3

- ◆ Rotations about partial double bonds give barriers that are too low
- ◆ Parameters for metals based on only a few data points
- ◆ Weak interactions unreliable (like Hartree-Fock, semiempirical methods do not describe electron correlation effects that give rise to London dispersion forces, unless some special correction is added like a “-D” dispersion correction)

MNDO/d

- ◆ Adds d functions to MNDO
- ◆ For metals, these are essential
- ◆ For 2nd row atoms, these are helpful in describing polarization
- ◆ Can offer significant improvements over MNDO, AM1, PM3

Parameterized Model 6 (PM6)

- ◆ NDDO Method with improved parameters and improved core-core interaction term (diatomic parameters like original MINDO)
- ◆ 70 elements parameterized
- ◆ Corrects some problems with AM1, PM3
- ◆ Includes d functions for metals like MNDO/d

Semiempirical Total Energies

- ◆ Like the Hartree-Fock model they're derived from, semiempirical methods yield total electronic energies (relative to infinitely separated nuclei and valence electrons)
- ◆ Unlike HF, we don't need to add zero-point vibrational energy (ZPVE) when computing enthalpy differences; it's included *implicitly* in the parameters

Average ΔH_f error in kcal/mol

Compounds	MNDO	AM1	PM3
All normal valent (607 compounds)	24.3	14.8	11.2
Hypervalent (106 compounds)	104.5	62.3	17.3
All (713 compounds)	46.2	27.6	11.6

Data from Frank Jensen, *Introduction to Computational Chemistry* (Wiley, New York, 1999)

Average ΔH_f error in kcal/mol

Compounds	MNDO	AM1	PM3	MNDO/d
Al, Si, P, S, Cl, Br, I, Zn, Hg (488 compounds)	29.2	15.3	10.0	4.9
Si, P, S, Cl, Br, I (404 compounds)	31.4	16.1	9.5	5.1

Data from Frank Jensen, *Introduction to Computational Chemistry* (Wiley, New York, 1999)

Average ΔH_f error in kcal/mol

Compounds	AM1	PM3	PM6
H, C, N, O (1157 compounds)	9.4	5.7	4.6
Main group (3188 compounds)	22.3	17.8	6.2

Compounds	PM6	B3LYP/ 6-31G*	HF/ 6-31G*
1373 compounds	4.4	5.2	7.4

J. J. P. Stewart, *J. Mol. Model* **13**, 1173-1123 (2007)

Average errors in bond lengths (Å)

Bonds to	AM1	PM3	PM6
H, C, N, O (413 compounds)	0.031	0.021	0.025
H, C, N, O, F, P, S, Cl, Br, I (712 compounds)	0.046	0.037	0.031
Main group (2636 compounds)	0.131	0.104	0.085

J. J. P. Stewart, *J. Mol. Model* **13**, 1173-1123 (2007)

Typical semiempirical accuracy for other properties

Property	Accuracy
Bond angles	$\pm 4^\circ$
Ionization potentials	$\pm 0.5-1.0$ eV
Dipole moments	± 0.4 D
Relative energies (ΔH_f)	$\pm 5-10$ kcal/mol (not as systematic as ab initio energy errors)

Improvements to PM6 (etc.) for Noncovalent Interactions: PM6-DH, PM6-DH2, PM6-D3H4X

- ◆ Add term for dispersion: damped $C_{6,AB}/R_{AB}^6$ terms with same form as in popular $-D$ dispersion corrections to DFT and analogous to London dispersion terms in force fields
- ◆ Add term for hydrogen bonding

PM6-DH: J. Rezac, J. Fanfrlik, D. Salahub, and P. Hobza, *J. Chem. Theory Comput.* **5**, 1749-1760 (2009)

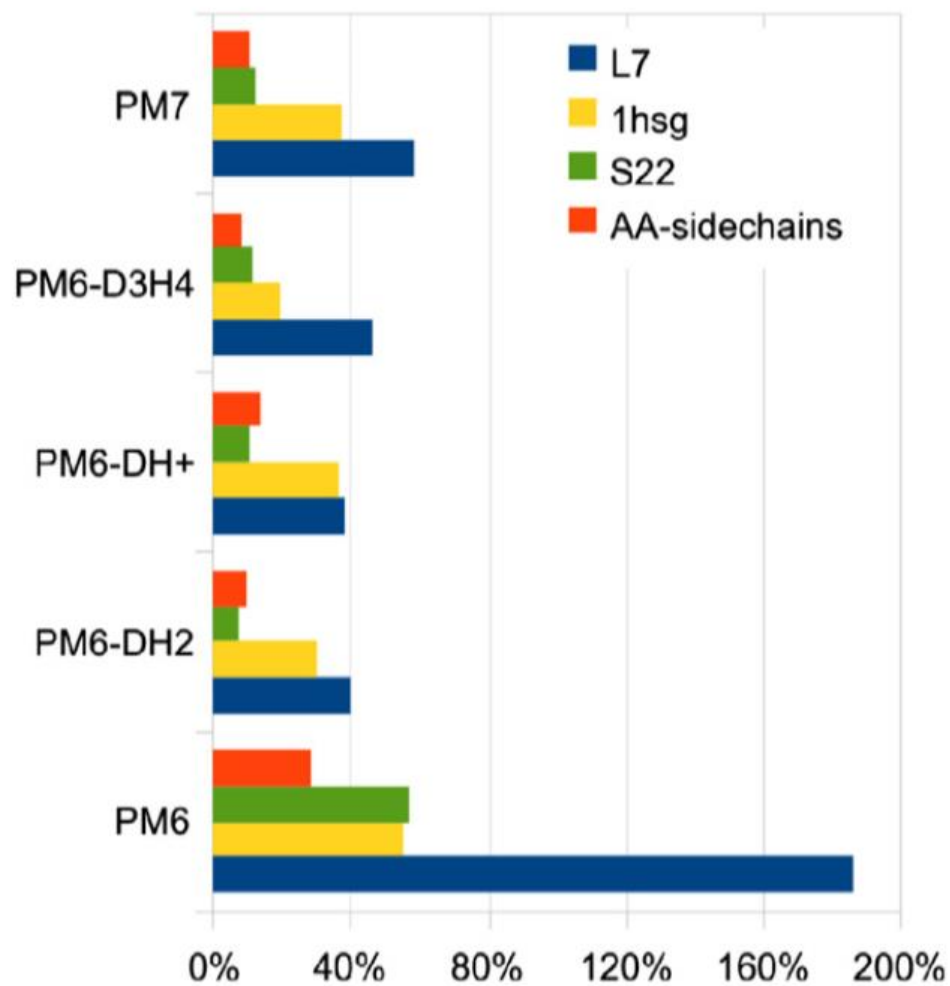
PM6-DH2: M. Korth, M. Pitonak, J. Rezac, and P. Hobza, *J. Chem. Theory Comput.* **6**, 344 (2010)

D3H4X correction: J. Rezac and P. Hobza, *J. Chem. Theory Comput.* **8**, 141 (2012)

Parameterized Model 7 (PM7)

- ◆ Further improvements upon PM6
- ◆ Added explicit terms to describe non-covalent interactions, based on ideas in the -DH2, -DH+, and -D3H4 corrections to PM6 by Hobza and co-workers
- ◆ Aimed at giving better results for molecules different than those in the training set
- ◆ Corrects two minor errors in the NDDO formalism
- ◆ MUE in bond lengths decreased by about 5% and in ΔH_f by about 10% relative to PM6 (for organic solids, improvement of 60% in ΔH_f)

Modern Semiempirical Methods for Non-Covalent Interactions



RMS %error in intermolecular interaction energies, from J. Hostas, J. Rezac, and P. Hobza, *Chem. Phys. Lett.* **568**, 161 (2013)

Extended Hückel Theory

- ◆ Parameterizes the Fock matrix, not the integrals
- ◆ $F_{\mu\mu} = -I_{\mu}$
- ◆ $F_{\mu\nu} = -k S_{\mu\nu} (I_{\mu} + I_{\nu}) / 2$ (k=1.75)
- ◆ Very approximate method, but can be useful for getting trends or generating guess MO's for Hartree-Fock

Summary of Semiempirical Methods

- ◆ Neglect of all 3 and 4 center integrals makes formal scaling go from $O(N^4)$ to $O(N^2)$ for construction of Fock matrix
- ◆ Diagonalization of F scales as $O(N^3)$ and therefore becomes formal rate-determining step (there are tricks to avoid explicit diagonalization)
- ◆ Share advantages/disadvantages of force fields: perform well for systems where experimental data available, not reliable otherwise
- ◆ Somewhat more robust than force fields because at least based on QM