

Basis Sets

Hartree-fock calculations for polyatomic molecules rely on expressing the molecular orbitals as a Linear Combination of Atomic Orbitals (LCAO).

The most natural choice for these atomic basis functions are the ones previously obtained for simple atomic systems (e.g., H):

Slater-type orbitals (STO's):

$$\frac{(2J/a_0)^{\frac{n+1}{2}}}{[(2n)!]^{1/2}} r^{n-1} e^{-Jr/a_0} Y_l^m(\theta, \varphi)$$

A set of such functions with n, l, m being integers, but J taking all possible positive values form a complete ^{basis} set that can describe any function.

(This is similar to Fourier decomposition of any function in the form $\sin(kx/L), \cos(kx/L)$)
 J is called the orbital exponent.

In principle, an infinite number of basis functions is required, but for most molecular systems, the electronic wavefunction can be described accurately using 1-5 STO's per atom.

How does the computational cost (CPU time, memory) scale with the number of basis functions?

The interaction term $\sum_{i=1}^N \sum_{j=i+1}^N \frac{e^2}{r_{ij}}$ for electron-

electron contributions can be expressed as an integral:

$$\iint_{\text{volume}} |\varphi_1|^2 |\varphi_2|^2 \frac{1}{r_{12}} d\vec{r}_1 d\vec{r}_2 \quad (1) \quad \varphi_1, \varphi_2 \text{ are complex functions}$$

volume

Expanding in terms of basis functions $\chi_1, \chi_2, \dots, \chi_m$

$$\varphi_1 = c_1 \chi_1(1) + c_2 \chi_2(1) + \dots + c_m \chi_m(1) \quad \left. \begin{array}{l} \\ \end{array} \right\} \textcircled{1} \Rightarrow$$

$$\varphi_2 = d_1 \chi_1(2) + d_2 \chi_2(2) + \dots + d_m \chi_m(2)$$

$$\textcircled{1} \text{ becomes } \left. \begin{array}{l} \\ \\ \end{array} \right\} \sum_i \sum_j \sum_k \sum_l c_i^* c_j d_k^* d_l \iiint \chi_i^*(1) \chi_j(2) \chi_k^*(2) \chi_l(2) \frac{1}{r_{12}} d\vec{r}_1 d\vec{r}_2 d\vec{r}_3 d\vec{r}_4$$

In the general case, i, j, k and l functions can be on different nuclei.

The number of integrals is $\propto \boxed{m^4}$

For 20-100 basis functions \rightarrow 20,000 to 13,000,000 integrals

The integrals for STO functions on different atomic nuclei are extremely hard to compute.

To enable these calculations, Boys (1950) proposed approximating STO's with Gaussian-type orbitals:

$$\text{GTO's} \rightarrow x^l y^m z^n e^{-\beta r^2/a_0^2} \quad \begin{array}{l} l, m, n \text{ integers} \\ \text{(if } \beta=0 \rightarrow s \text{ orbitals)} \\ \text{one} = 1 \rightarrow 2p \text{ orbitals} \end{array}$$

Why does this help?

The product of two Gaussians is another Gaussian centered along the line joining the two centers:

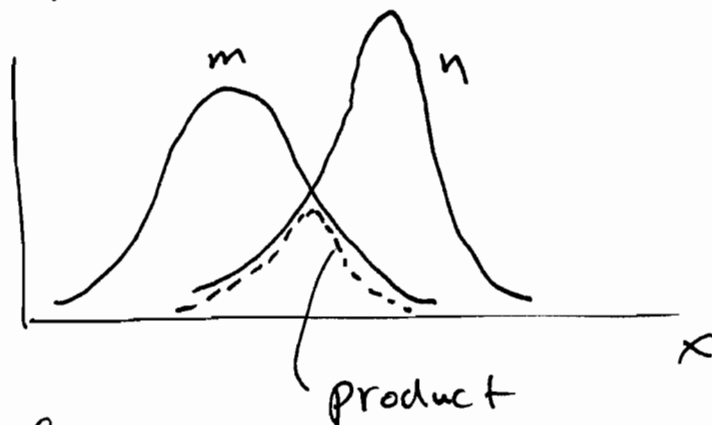
$$\exp(-\alpha_m r_m^2) \exp(-\alpha_n r_n^2) = \exp\left(-\frac{\alpha_m \alpha_n}{\alpha_m + \alpha_n} r_{mn}^2\right) \exp(-\alpha r_c^2)$$

r_{mn} is distance between m and n ; $\alpha = \alpha_m + \alpha_n$

r_c has coordinates

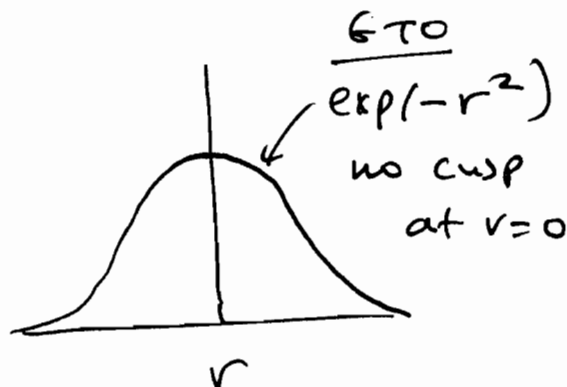
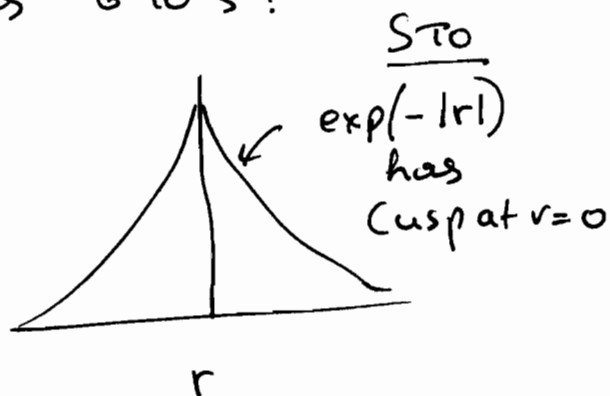
$$x_c = \frac{\alpha_m x_m + \alpha_n x_n}{\alpha_m + \alpha_n}$$

$$y_c = \dots \quad z_c = \dots$$



Thus, for the two-electron integrals, ^{two} Gaussians centered on different atoms can be replaced by a single Gaussian, making the calculations feasible.

Of course, STO's do not have the same shape as GTO's:



Solution: approximate shape of STO with sum of several Gaussians of different exponents α ; The more Gaussians, the better the approximation.

National Grid 100 SHEETS EYE FASE 5 SQUARE

Fitting of STO's by multiple Gaussians:

Each Gaussian has 2 parameters:

Coefficient + exponent (d, α)

Coefficients + exponents found so as to maximize overlap between STO and sum of GTO's.

E.g., for the 1s Slater-type orbital, need to maximize

$$\left. \frac{1}{\sqrt{\pi}} \left(\frac{2\alpha}{\pi} \right)^{3/4} \int e^{-r} e^{-\alpha r^2} d\vec{r} \right\} \Rightarrow$$

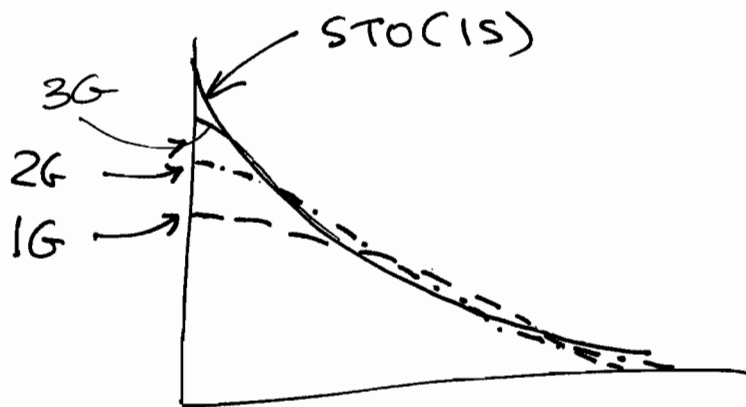
for 1 Gaussian $\alpha = 0.270$

for 1-Gaussian

For 2-Gaussians

$\alpha_1 = 0.151$ $d_1 = 0.678$ } linear
 $\alpha_2 = 0.851$ $d_2 = 0.430$ } combination

etc.



Another issue: extent of Gaussian (range)

Uncontracted (primitive) Gaussians \rightarrow both exponent + coefficient can vary, expensive

Contracted Gaussians \rightarrow only coefficient can vary, cheap.

Minimal Basis Sets: One STO, approximated by 3G, 4G ... (3, 4 ... Gaussians) for each inner-shell or valence-shell AO on each atom.

E.g., for C_2H_2 $\{ 1s, 2s, 2p_x, 2p_y, 2p_z \}$ 5 AO's for each C
 $1s$ } 2 AO's for the hydrogens

Fast, but inaccurate: • two few basis functions for end-of-row atoms (O, F ...)
 • inability to describe expanded states
 • non-spherical aspects limited

Better (more expensive) basis Sets: use more than one function for each orbital, especially for the valence-shell (outer) electrons that contribute most to the chemical properties.

E.g., 6-31G: 6 Gaussians per STO for core, 3 + 1 diffuse Gaussians for valence

* indicates polarization functions (d for 1st + 2nd row elements)

** also p for H's, useful when H acts as "bridging" atom, H-bonds
 +, ++: more diffuse functions for low- Z -nuclei electrons

