

Basis Sets

$$\Psi = (N!)^{-1/2} \begin{vmatrix} \chi_i(x_1) & \chi_j(x_1) & \dots & \chi_K(x_1) \\ \chi_i(x_2) & \chi_j(x_2) & \dots & \chi_K(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ \chi_i(x_N) & \chi_j(x_N) & \dots & \chi_K(x_N) \end{vmatrix}$$

$$\chi_i(x_j) = \Phi_i(r_j) \sigma(\omega_j) \quad \text{with:} \quad \Phi_i = \sum_{\mu=1}^K c_{\mu i} \varphi_{\mu} \quad \{\varphi_{\mu}\} - \text{a set of known functions}$$

For UHF wave-functions two sets of coefficients are needed:

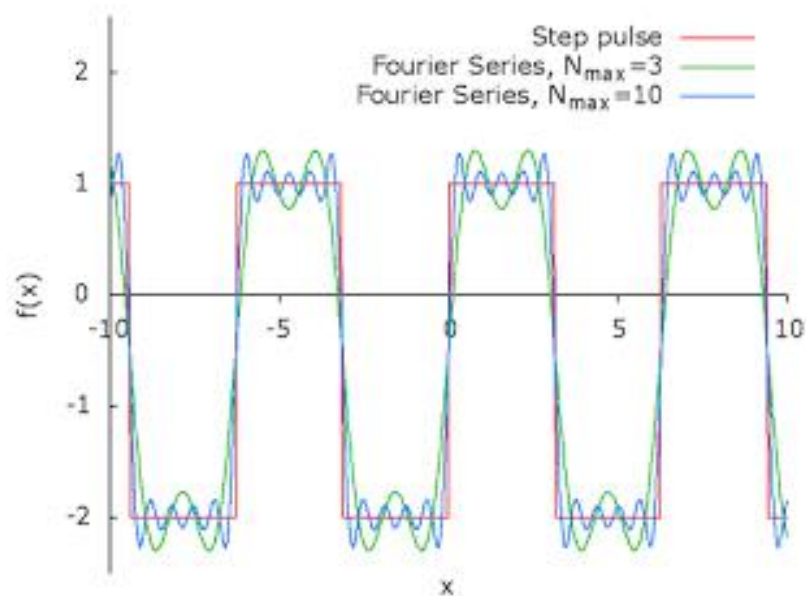
$$\left. \begin{aligned} \Phi_i^{\alpha} &= \sum_{\mu=1}^K c_{\mu i}^{\alpha} \varphi_{\mu} \\ \Phi_i^{\beta} &= \sum_{\mu=1}^K c_{\mu i}^{\beta} \varphi_{\mu} \end{aligned} \right\} \text{the same basis functions are used for } \alpha \text{ and } \beta \text{ orbitals}$$

if $\varphi_{\mu} \equiv \text{AO} \rightarrow \text{LCAO-MO}$

if $\varphi_{\mu} \neq \text{AO} \rightarrow \text{LCBF-MO}$

Basis functions

- mathematical functions chosen to give the maximum flexibility to the molecular orbitals
- must have physical significance
- their coefficients are obtained variationally
- Type: polynomials, Fourier series, spline, etc.

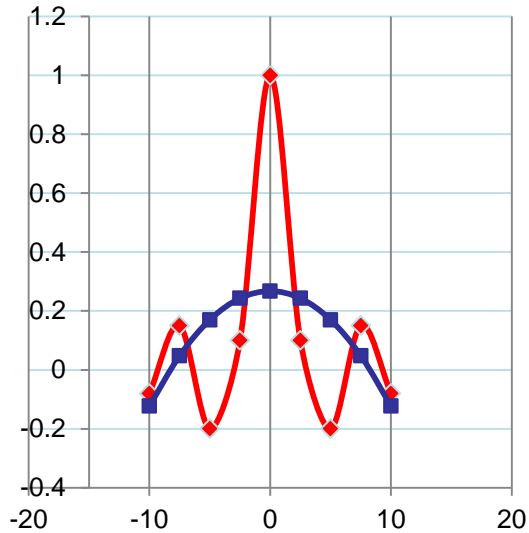


Basis set

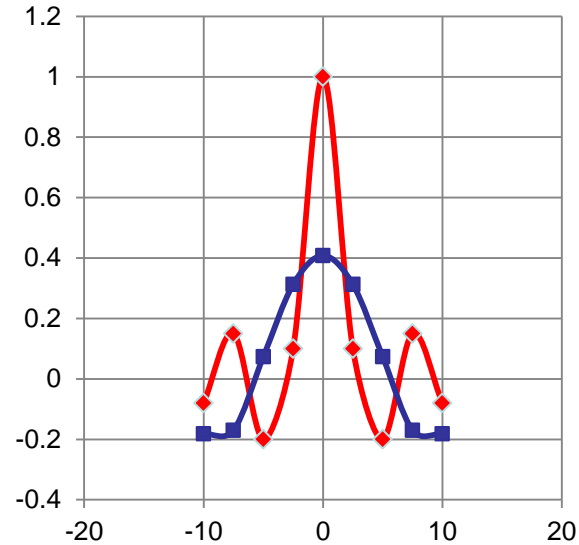
- a set of mathematical functions used to expand the molecular orbitals in order to solve the HFR equations.
- each function is centered (has its origin) at some point in the molecule (usually on the nuclei).
- each function is a function of the x,y,z coordinates of an electron.

Fitting a function with polynomials

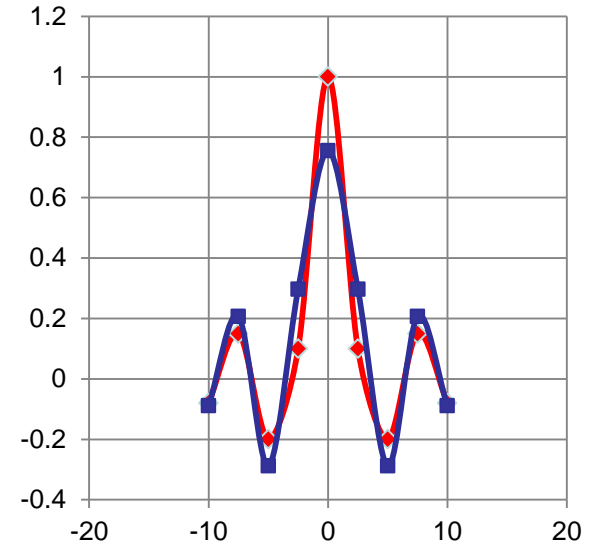
$$y = \sum_{k=0} c_{2k} x^{2k}$$



c0 0.2674
c2 -0.0039



c0 0.408
c2 -0.0159
c4 0.0001



c0 0.755
c2 -0.08594
c4 0.0021
c6 -0.00001325

Slater Type Orbitals (STO)

$$\varphi_i(\xi, n, l, m; r, \theta, \varphi) = Nr^{n-1}e^{-\xi r}Y_{lm}(\theta, \varphi)$$

- introduced by Slater in 1930
- similar to atomic orbitals of the hydrogen atom
- more convenient (from the numerical calculation point of view) than AO, especially when $n-l \geq 2$ (radial part is simply r^2, r^3, \dots and not a polynomial)



John C. Slater (1900-1976)

STO – are labeled like hydrogen atomic orbitals and their normalized form is:

$$\varphi_{1s} = \left(\frac{\xi_1^3}{\pi} \right)^{1/2} \exp(-\xi_1 r)$$

$$\varphi_{2s} = \left(\frac{\xi_2^5}{96\pi} \right)^{1/2} r \exp\left(-\frac{\xi_2 r}{2}\right) \quad \varphi_{2p_x} = \left(\frac{\xi_2^5}{32\pi} \right)^{1/2} x \exp\left(-\frac{\xi_2 r}{2}\right)$$

STO

- provide reasonable representations of atomic orbitals
- however, they are not well suited to numerical (fast) calculations of especially two-electron integrals
- their use in practical molecular orbital calculations has been limited

STO

Advantages:

- Physically, the exponential dependence on distance from the nucleus is very close to the exact hydrogenic orbitals.
- Ensures fairly rapid convergence with increasing number of functions.

Disadvantages:

- Three and four center integrals cannot be performed analytically.
- No radial nodes. These can be introduced by making linear combinations of STOs.

Practical Use:

- Calculations of very high accuracy, atomic and diatomic systems.
- Semi-empirical methods where 3- and 4-center integrals are neglected.

Gaussian Type Orbitals (GTO)

- introduced by Boys (1950)
- powers of x, y, z multiplied by $e^{-\alpha r^2}$
- α is a constant (called exponent) that determines the size (radial extent) of the function

$$g(\alpha, l, m, n; x, y, z) = N e^{-\alpha r^2} x^l y^m z^n$$

or:

$$g(\alpha, l, m, n, f; x, y, z) = N e^{-\alpha f^2 r^2} x^l y^m z^n$$

N - normalization constant

f - scaling factor

scale all exponents in the related gaussians in molecular calculations

l, m, n are not quantum numbers

$L=l+m+n$ - used analogously to the angular momentum quantum number for atoms to mark functions as s-type ($L=0$), p-type ($L=1$), d-type ($L=2$), etc (shells)

The absence of r^{n-1} pre-exponential factor restricts single Gaussian primitives to approximate only 1s, 2p, 3d, 4f, ... orbitals.

However, **combinations of Gaussian primitives** are able to approximate correct nodal properties of atomic orbitals

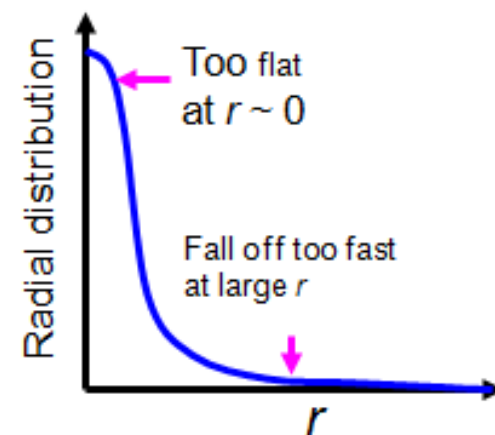
GTO – uncontracted gaussian function (gaussian primitive)

$\sum GTO$ - contracted gaussian function (gaussian contraction)

$$STO = \sum GTO$$

GTOs are inferior to STOs in three ways:

1. GTO's behavior near the nucleus is poorly represented. At the nucleus, the GTO has zero slope; the STO has a cusp. GTOs diminish too rapidly with distance.
2. The 'tail' behavior is poorly represented.
3. Extra d -, f -, g -, etc. functions may lead to linear dependence of the basis set. They are usually dropped when large basis sets are used.



Advantage:

Use a linear combination of GTOs to overcome these deficiencies.

The first ten normalized gaussian primitives are:

$$g_s(\alpha, r) = \left(\frac{2\alpha}{\pi} \right)^{3/4} \exp(-\alpha r^2)$$

$$g_x(\alpha, r) = \left(\frac{128\alpha^5}{\pi^3} \right)^{1/4} x \exp(-\alpha r^2)$$

$$g_y(\alpha, r) = \left(\frac{128\alpha^5}{\pi^3} \right)^{1/4} y \exp(-\alpha r^2)$$

$$g_z(\alpha, r) = \left(\frac{128\alpha^5}{\pi^3} \right)^{1/4} z \exp(-\alpha r^2)$$

$$g_{xx}(\alpha, r) = \left(\frac{2048\alpha^7}{9\pi^3} \right)^{1/4} x^2 \exp(-\alpha r^2)$$

$$g_{yy}(\alpha, r) = \left(\frac{2048\alpha^7}{9\pi^3} \right)^{1/4} y^2 \exp(-\alpha r^2)$$

$$g_{zz}(\alpha, r) = \left(\frac{2048\alpha^7}{9\pi^3} \right)^{1/4} z^2 \exp(-\alpha r^2)$$

$$g_{xy}(\alpha, r) = \left(\frac{2048\alpha^7}{\pi^3} \right)^{1/4} xy \exp(-\alpha r^2)$$

$$g_{xz}(\alpha, r) = \left(\frac{2048\alpha^7}{\pi^3} \right)^{1/4} xz \exp(-\alpha r^2)$$

$$g_{yz}(\alpha, r) = \left(\frac{2048\alpha^7}{\pi^3} \right)^{1/4} yz \exp(-\alpha r^2)$$

There are 6 possible d-type cartesian gaussians while there are only 5 linearly independent and orthogonal d orbitals

The g_s , g_x , g_y and g_z primitives have the angular symmetries of the four corresponding AO.

The 6 d-type gaussian primitives may be combined to obtain a set of 5 d-type functions:

$$\begin{array}{ll} g_{xy} \rightarrow d_{xy} & \frac{1}{2}(2g_{zz} - g_{xx} - g_{yy}) \rightarrow d_{z^2} \\ g_{xz} \rightarrow d_{xz} & \\ g_{yz} \rightarrow d_{yz} & \sqrt{\frac{3}{4}}(g_{xx} - g_{yy}) \rightarrow d_{x^2-y^2} \end{array}$$

The 6-th linear combination gives an s-type function:

$$g_{rr} = 5^{-1/2}(g_{xx} + g_{yy} + g_{zz}) \rightarrow g_s$$

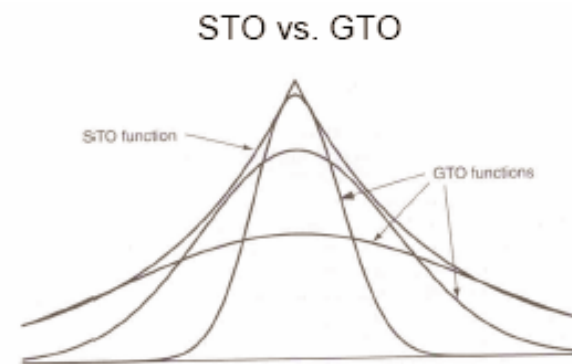
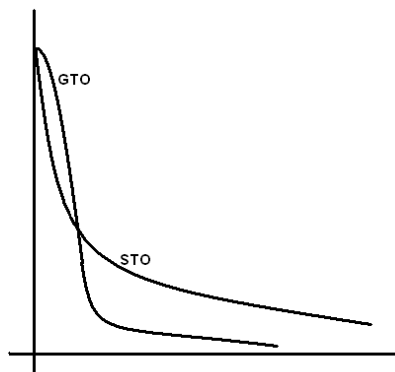
In a similar manner, the 10 f-type gaussian primitives may be combined to obtain a set of 7 f-type functions

GTOs are less satisfactory than STOs in describing the AOs close to the nucleus. The two type functions substantially differ for $r=0$ and also, for very large values of r .

cuspid condition:

for STO: $[d/dr e^{-\xi r}]_r \neq 0$

for GTO: $[d/dr e^{-\alpha r^2}]_{r=0} = 0$



With GTO the two-electron integrals are more easily evaluated. The reason is that the product of two gaussians, each on different centers, is another gaussian centered between the two centers:

$$\Phi_{1s}^{GF}(\alpha, \mathbf{r} - \mathbf{R}_A) \Phi_{1s}^{GF}(\beta, \mathbf{r} - \mathbf{R}_B) = K_{AB} \Phi_{1s}^{GF}(p, \mathbf{r} - \mathbf{R}_P)$$

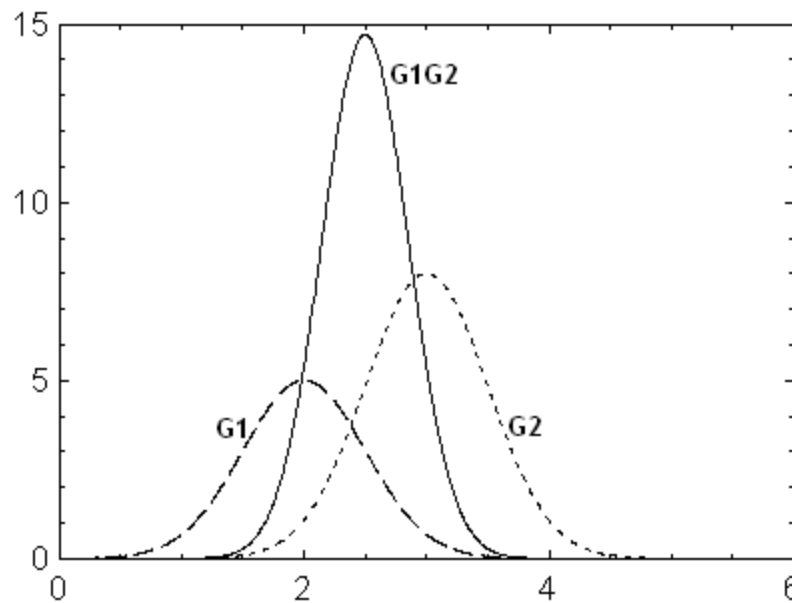
where:

$$K_{AB} = (2\alpha\beta/[(\alpha+\beta)\pi])^{3/4} \exp(-\alpha\beta/(\alpha+\beta)|\mathbf{R}_A - \mathbf{R}_B|^2)$$

The exponent of the new gaussian centered at R_p is: $p = \alpha + \beta$

and the third center P is on line joining the centers A and B (see the Figure below)

$$\mathbf{R}_P = (\alpha\mathbf{R}_A + \beta\mathbf{R}_B) / (\alpha + \beta)$$



The product of two 1s gaussian is a third 1s gaussian

- GTO
- allow a more rapidly and efficiently calculation of the two-electron integrals
 - have different functional behavior with respect to known functional behavior of AOs.

GTOs are obtained from HF calculations on isolated atoms by varying the exponents to obtain the minimum energy.

It is normally to assume that these GTOs are not suited for molecular calculations.

⇒ CGTOs are used for molecular calculations

Since GTOs from different shells are orthogonal, such primitives will not be combined in a given CGTO.

⇒ contractions (CGF or CGTO)

$$\Phi_{\mu}^{CGF}(\mathbf{r} - \mathbf{R}_A) = \sum_{p=1}^L d_{p\mu} \Phi_p^{GF}(\alpha_{p\mu}, \mathbf{r} - \mathbf{R}_A)$$

L – the length of the contraction
 $d_{p\mu}$ – contraction coefficients

How the gaussian primitives are derived?

by fitting the CGF to an STO using a least square method

varying the exponents in quantum calculations on atoms in order to minimize the energy

Example

STO-3G basis set for H₂ molecule

Each BF is approximated by a STO, which in turn, is fitted to a CGF of 3 primitives
hydrogen 1s orbital in STO-3G basis set

For molecular calculations, first we need a BF to describe the H 1s atomic orbital
then: MO(H₂) = LCBF

3 gaussian primitives:

exponent	coefficient
0.222766	0.154329
0.405771	0.535328
0.109818	0.444636

$$\Phi_{1s}^{\text{STO-3G}} = N_1 c_1 e^{-\alpha_1 r^2} + N_2 c_2 e^{-\alpha_2 r^2} + N_3 c_3 e^{-\alpha_3 r^2}$$

If we use a scaling factor:

$$\Phi_{1s}^{\text{STO-3G}} = N'_1 c_1 e^{-\alpha_1 f^2 r^2} + N'_2 c_2 e^{-\alpha_2 f^2 r^2} + N'_3 c_3 e^{-\alpha_3 f^2 r^2}$$

$$\beta_i = \alpha_i f^2$$

$$N'_i = \left(\frac{2\beta_i}{\pi} \right)^{3/4} \quad \left(\frac{2}{\pi} \right)^{3/4} = 0.7127054$$

! Using normalized primitives we do not need a normalization factor for the whole contraction

If the primitives are not normalized, we have to obtain a normalization factor. For this, we use the condition:

$$S = \langle \Phi_{1s}^{\text{STO-3G}} | \Phi_{1s}^{\text{STO-3G}} \rangle = 1$$

$$\begin{aligned} \langle \Phi_{1s}^{\text{STO-3G}} | \Phi_{1s}^{\text{STO-3G}} \rangle = F^2 [& \underbrace{\int N_1^2 c_1^2 e^{-2\beta_1 r^2} d\tau}_{I_1} + \underbrace{\int N_2^2 c_2^2 e^{-2\beta_2 r^2} d\tau}_{I_2} + \underbrace{\int N_3^2 c_3^2 e^{-2\beta_3 r^2} d\tau}_{I_3} + \\ & + 2 \underbrace{\int N_1 N_2 c_1 c_2 e^{-(\beta_1 + \beta_2) r^2} d\tau}_{I_4} + 2 \underbrace{\int N_1 N_3 c_1 c_3 e^{-(\beta_1 + \beta_3) r^2} d\tau}_{I_5} \\ & + 2 \underbrace{\int N_2 N_3 c_2 c_3 e^{-(\beta_2 + \beta_3) r^2} d\tau}_{I_6}] \end{aligned}$$

$$S = F^2 [I_1 + I_2 + I_3 + 2I_4 + 2I_5 + 2I_6]$$

$$I_1 = N_1^2 c_1^2 4\pi \int_0^{\infty} e^{-2\beta_1 r^2} r^2 dr = N_1^2 c_1^2 4\pi \cdot \frac{1}{4} \sqrt{\frac{\pi}{(2\beta_1)^3}} = N_1^2 c_1^2 \cdot \frac{\pi^{3/2}}{(2\beta_1)^{3/2}} = N_1^2 c_1^2 \cdot \left(\frac{\pi}{2\beta_1}\right)^{3/2}$$

But:

$$N_1 = \left(\frac{2\beta_1}{\pi}\right)^{3/4} \Rightarrow N_1^2 = \left(\frac{2\beta_1}{\pi}\right)^{3/2}$$

so that:

$$I_1 = c_1^2$$

Analogously:

$$I_2 = c_2^2$$

$$I_3 = c_3^2$$

$$I_4 = 4\pi N_1 N_2 c_1 c_2 \int_0^{\infty} e^{-(\beta_1 + \beta_2)r^2} r^2 dr = 4\pi N_1 N_2 c_1 c_2 \frac{1}{4} \sqrt{\frac{\pi}{(\beta_1 + \beta_2)^3}} = N_1 N_2 c_1 c_2 \frac{\pi^{3/2}}{(\beta_1 + \beta_2)^{3/2}}$$

$$N_1 N_2 = \left(\frac{2\beta_1}{\pi}\right)^{3/4} \left(\frac{2\beta_2}{\pi}\right)^{3/4} = \left(\frac{2}{\pi}\right)^{3/2} \beta_1^{3/4} \beta_2^{3/4}$$

and thus:

$$I_4 = \frac{2^{3/2} \beta_1^{3/4} \beta_2^{3/4}}{(\beta_1 + \beta_2)^{3/2}} c_1 c_2$$

$$I_5 = \frac{2^{3/2} \beta_1^{3/4} \beta_3^{3/4}}{(\beta_1 + \beta_3)^{3/2}} c_1 c_3$$

$$I_6 = \frac{2^{3/2} \beta_2^{3/4} \beta_3^{3/4}}{(\beta_2 + \beta_3)^{3/2}} c_2 c_3$$

Now,

$$S = F^2 \left[c_1^2 + c_2^2 + c_3^2 + 2^{5/2} \frac{\beta_1^{3/4} \beta_2^{3/4}}{(\beta_1 + \beta_2)^{3/2}} c_1 c_2 + 2^{5/2} \frac{\beta_1^{3/4} \beta_3^{3/4}}{(\beta_1 + \beta_3)^{3/2}} c_1 c_3 + 2^{5/2} \frac{\beta_2^{3/4} \beta_3^{3/4}}{(\beta_2 + \beta_3)^{3/2}} c_2 c_3 \right]$$

Imposing that S=1 we obtain:

$$F = \frac{1}{\left\{ c_1^2 + c_2^2 + c_3^2 + 2^{5/2} \left[\frac{\beta_1^{3/4} \beta_2^{3/4}}{(\beta_1 + \beta_2)^{3/2}} c_1 c_2 + \frac{\beta_1^{3/4} \beta_3^{3/4}}{(\beta_1 + \beta_3)^{3/2}} c_1 c_3 + \frac{\beta_2^{3/4} \beta_3^{3/4}}{(\beta_2 + \beta_3)^{3/2}} c_2 c_3 \right] \right\}^{1/2}}$$

In the general case of a contraction of dimension n, the above expression become:

$$F = \sum_{i=1}^n \sum_{j=1}^n \left(\frac{2}{\beta_i + \beta_j} \right)^{3/2} \beta_i^{3/4} \beta_j^{3/4} c_i c_j$$

Summary

The 1s hydrogen orbital in STO-3G basis set will be:

$$\Phi(r) = N \sum_{i=1}^3 c_i g_{1s}(\beta_i, r)$$

with:

$$g_{1s}(\beta_i, r) = N_i e^{-\beta_i r^2}$$

$$N_i = \left(\frac{2\beta_i}{\pi} \right)^{3/4} \quad \text{- normalization factors for primitives}$$

$$N = \sum_{i=1}^3 \sum_{j=1}^3 \left(\frac{2}{\beta_i + \beta_j} \right)^{3/2} \beta_i^{3/4} \beta_j^{3/4} c_i c_j$$

- normalization factor for the whole contraction (when un-normalized primitives or segmented contractions are used)

$$\Rightarrow \Phi(r) = N \sum_{i=1}^3 c_i N_i e^{-\beta_i r^2}$$

α_i	β_i	c_i	N_i	$c_i N_i$
2.227660	3.425250	0.154329	1.794441	0.276934
0.405771	0.623913	0.535328	0.500326	0.267839
0.109818	0.168856	0.444635	0.187736	0.083474

$$\Rightarrow N=1.0000002$$

Explicitly:

$$\Phi_{1s}(r) = 1 \left[0.276934e^{-3.425250r^2} + 0.267839e^{-0.623913r^2} + 0.083474e^{-0.168856r^2} \right]$$

If the exponents are not scaled:

$$\Phi_{1s}(r) = 1 \left[0.200560e^{-2.227660r^2} + 0.193973e^{-0.405771r^2} + 0.060453e^{-0.109818r^2} \right]$$

Segmented contractions

- usually structured in such a way that the most diffuse primitives ((with the smallest exponent) are left uncontracted (i.e. one primitive per basis function)
- more compact primitives (those with larger exponents) are used to construct one or more contractions which are subsequently renormalized

Notations for segmented contractions

Examples:

() – contains the number of primitives that are given in the order of angular number

(12s,9p,1d) \equiv (12,9,1)

[] – used to specify the number of resulting contractions

[5,4,1] – means that s-shell has 5 contractions, p-shell has 4 contractions and d-shell has only one contraction

To denote how contractions were performed the following notation is used:

(12,9,1) \rightarrow [5,4,1]

or

(12,9,1)/[5,4,1]

or

(12s,9p,1d) \rightarrow [5s,4p,1d]

\rightarrow 12 s-type primitives were contracted to form 5 s-type contractions (BF)

9 p-type primitives were contracted to form 4 p-type contractions (BF)

(actually 12 BF were created because each p-type BF has 3 variants)

1 d-type primitive was used as a BF by its self

(5 d-type BF were created because each d-type BF has 5 variants)

A more complete notation

- explicitly list the number of primitives in each contraction

(63111,4311,1)

means that:

from 12 s-type primitives (6+3+1+1+1) 5 s-type BF were formed:

one consists from 6 primitives

one consists from 3 primitives

three consists from 1 primitive

from 9 p-type primitives (4+3+1+1) 4 (12) p-type BF were obtained

one consists from 4 primitives

one consists from 3 primitives

two consists from 1 primitive

from 1 d-type primitive 1 (5) d-type BF was (were) formed

Equivalent notations

(63111/4311/1)

(633x1,432x1,1)

s(6/3/1/1/1), p(4/3/1/1), d(1)

(6s,3s,1s,1s,1s/4p,3p,1p,1p/1d)

(6,3,1,1,1/4,3,1,1/1)

When specifying the structure of the basis set for the entire molecule, slashes are used to separate information for different atoms. The information is given starting from the heaviest atom.

Example

water molecule

(10s,5p,1d/5s,1p) → [4s,2p,1d/2s,1p]

→ contractions for oxygen atom: (10,5,1)/[4,2,1]

→ contractions for hydrogen atoms (5,1)/[2,1]

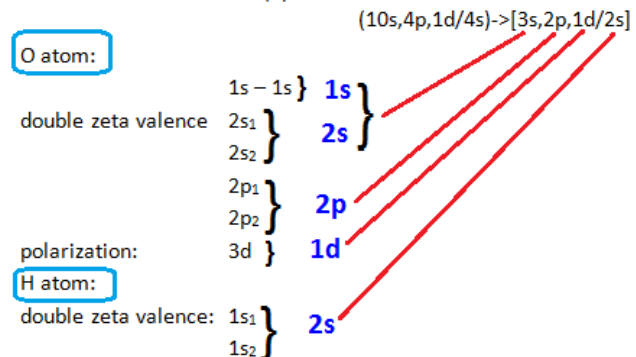
→ **19 BFs for water. Explain!**

further reading

Jan Labanowski <http://www.ccl.net/ccs/documents/basis-sets/basis.html>

Example

The contraction for 6-31G(d) basis set is:



Minimal basis sets

-one basis function for every atomic orbital that is required to describe the free atom

For carbon, the minimal basis set consists of a '1s' orbital, a '2s' orbital and the full set of three '2p' orbitals.

The minimal basis set for the methane molecule consists of 4 '1s' orbitals - one per hydrogen atom, and the set of '1s', '2s' and '2p' as described above for carbon. The total basis set comprises 9 basis functions.

H – 1s orbital

C – 1s, 2s, 2px, 2py, 2pz

→ for CH₄ molecule: 4 x H1s orbitals
C1s, C2s and 3 x C2p orbitals
→ 9BF

STO-nG

STO-3G

- a linear combination of 3 GTOs are fitted to an STO

-for CH₄ molecule → 9BF → 27 primitives

Each basis function is a contraction of three primitive Gaussian.

The exponents and expansion coefficients for the primitives are obtained from a least squares fit to Slater type orbitals (STOs).

STO-3G basis set example

http://www.chem.utas.edu.au/staff/yatesb/honours/modules/mod5/c_sto3g.html

This is an example of the STO-3G basis set for methane in the format produced by the "gfinput" command in the Gaussian computer program. The first atom is carbon. The other four are hydrogens.

Standard basis: STO-3G (5D, 7F) Basis set in the form of general basis input:

```
1 0 //C atom
S 3 1.00
.7161683735D+02 .1543289673D+00
.1304509632D+02 .5353281423D+00
.3530512160D+01 .4446345422D+00
SP 3 1.00
.2941249355D+01 -.9996722919D-01 .1559162750D+00
.6834830964D+00 .3995128261D+00 .6076837186D+00
.2222899159D+00 .7001154689D+00 .3919573931D+00
****

2 0 // H atom
S 3 1.00
.3425250914D+01 .1543289673D+00
.6239137298D+00 .5353281423D+00
.1688554040D+00 .4446345422D+00
****

3 0 // H atom
S 3 1.00
.3425250914D+01 .1543289673D+00
.6239137298D+00 .5353281423D+00
.1688554040D+00 .4446345422D+00
****

4 0 // H atom
S 3 1.00
.3425250914D+01 .1543289673D+00
.6239137298D+00 .5353281423D+00
.1688554040D+00 .4446345422D+00
****

5 0 // H atom
S 3 1.00
.3425250914D+01 .1543289673D+00
.6239137298D+00 .5353281423D+00
.1688554040D+00 .4446345422D+00
****
```

The energy decreases by increasing the number of primitives used.

The limit of an infinite basis set is known as the *Hartree-Fock limit*.

This energy is still greater than the exact energy that follows from the Hamiltonian because of the independent particle approximation.

Split valence basis sets

http://www.chem.utas.edu.au/staff/yatesb/honours/modules/mod5/split_bas.html

Valence orbitals are represented by more than one basis function, (each of which can in turn be composed of a fixed linear combination of primitive Gaussian functions). Depending on the number of basis functions used for the representation of valence orbitals, the basis sets are called valence double, triple, or quadruple-zeta basis sets. Since the different orbitals of the split have different spatial extents, the combination allows the electron density to adjust its spatial extent appropriate to the particular molecular environment.

Split is often made for valence orbitals only, which are chemically important.

3-21G basis set

The valence functions are split into one basis function with two GTOs, and one with only one GTO. (This is the "two one" part of the nomenclature.) The core consists of three primitive GTOs contracted into one basis function, as in the STO-3G basis set.

```
1 0 //C atom
S 3 1.00
.1722560000D+03 .6176690000D-01
.2591090000D+02 .3587940000D+00
.5533350000D+01 .7007130000D+00
SP 2 1.00
.3664980000D+01 -.3958970000D+00 .2364600000D+00
.7705450000D+00 .1215840000D+01 .8606190000D+00
SP 1 1.00
.1958570000D+00 .1000000000D+01 .1000000000D+01
****

2 0 //H atom
S 2 1.00
.5447178000D+01 .1562850000D+00
.8245472400D+00 .9046910000D+00
S 1 1.00
.1831915800D+00 .1000000000D+01
****
```

- ❖ The split-valence (SV) basis set uses one function for orbitals that are not in the valence shell and 2 functions for those in the valence shell.
- ❖ The double-zeta (DZ) basis set uses two basis functions where the minimal basis set had only one function.

Extended basis sets

The most important additions to basis sets are [polarization functions](#) and [diffuse basis functions](#).

Polarization basis functions

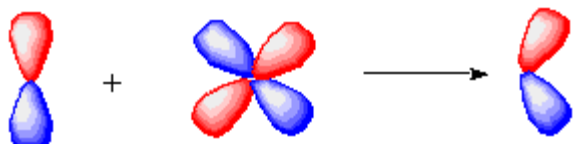
The influence of the neighboring nuclei will distort (polarize) the electron density near a given nucleus. In order to take this effect into account, orbitals that have more flexible shapes in a molecule than the s, p, d, etc., shapes in the free atoms are used.

- ❖ A set of **Gaussian functions one unit higher in angular momentum** than what are present in the ground state of the atom are added as **polarization functions**, again increasing the flexibility of the basis set in the valence region in the molecule.
- ❖ Orbital polarization phenomenon may be introduced well by adding 'polarization functions' to the basis set.

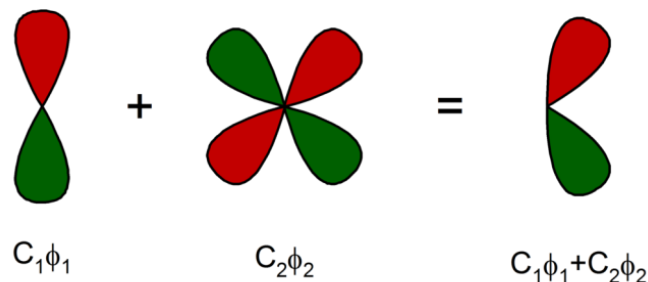
Let's polarize!



An s orbital is polarized by using a p-type orbital



A p orbital is polarized by mixing in a d-type orbital



"D-polarization function" by Rifleman 82 - Own work. Licensed under CC BY-SA 3.0 via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:D-polarization_function.png#mediaviewer/File:D-polarization_function.png

6-31G(d) – “**spectroscopic**” basis set

a set of d orbitals is used as polarization functions on heavy atoms

6-31G(d,p)

a set of d orbitals are used as polarization functions on heavy atoms

and a set of p orbitals are used as polarization functions on hydrogen atoms

Diffuse basis functions

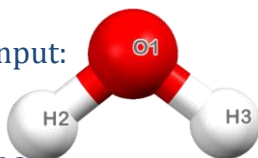
For **excited states and anions** where the electronic density is more spread out over the molecule, some basis functions which themselves are more spread out are needed (i.e. GTOs with small exponents). These additional basis functions are called **diffuse functions**. They are normally added as single GTOs.

6-31+G - adds a set of diffuse sp orbitals to the atoms in the first and second rows (Li - Cl).

6-31++G - adds a set of diffuse sp orbitals to the atoms in the first and second rows (Li- Cl) *and* a set of diffuse s functions to hydrogen.

Diffuse functions can also be added along with polarization functions.

This leads, for example, to the 6-31+G(d), 6-31++G(d), 6-31+G(d,p) and 6-31++G(d,p) basis sets.



Input format for Gaussian 6-31+G(d) basis set for C

1 0					
S	6 1.00				
	.3047524880D+04	.1834737130D-02			
	.4573695180D+03	.1403732280D-01			
	.1039486850D+03	.6884262220D-01			
	.2921015530D+02	.2321844430D+00			
	.9286662960D+01	.4679413480D+00			
	.3163926960D+01	.3623119850D+00			
SP	3 1.00				
	.7868272350D+01	-.1193324200D+00	.6899906660D-01		
	.1881288540D+01	-.1608541520D+00	.3164239610D+00		
	.5442492580D+00	.1143456440D+01	.7443082910D+00		
SP	1 1.00				
	.1687144782D+00	.1000000000D+01	.1000000000D+01		
SP	1 1.00				
	.4380000000D-01	.1000000000D+01	.1000000000D+01		
D	1 1.00				
	.8000000000D+00	.1000000000D+01			

6-31+G(d)

Exponents

Contraction coefficients

Standard basis: 6-31+G(d) (5D, 7F)

AO basis set in the form of general basis input:

1 0

S 6 1.00

0.5484671660D+04 0.1831074430D-02

0.8252349460D+03 0.1395017220D-01

0.1880469580D+03 0.6844507810D-01

0.5296450000D+02 0.2327143360D+00

0.1689757040D+02 0.4701928980D+00

0.5799635340D+01 0.3585208530D+00

SP 3 1.00

0.1553961625D+02 -0.1107775495D+00 0.7087426823D-01

0.3599933586D+01 -0.1480262627D+00 0.3397528391D+00

0.1013761750D+01 0.1130767015D+01 0.7271585773D+00

SP 1 1.00

0.2700058226D+00 0.1000000000D+01 0.1000000000D+01

SP 1 1.00

0.8450000000D-01 0.1000000000D+01 0.1000000000D+01

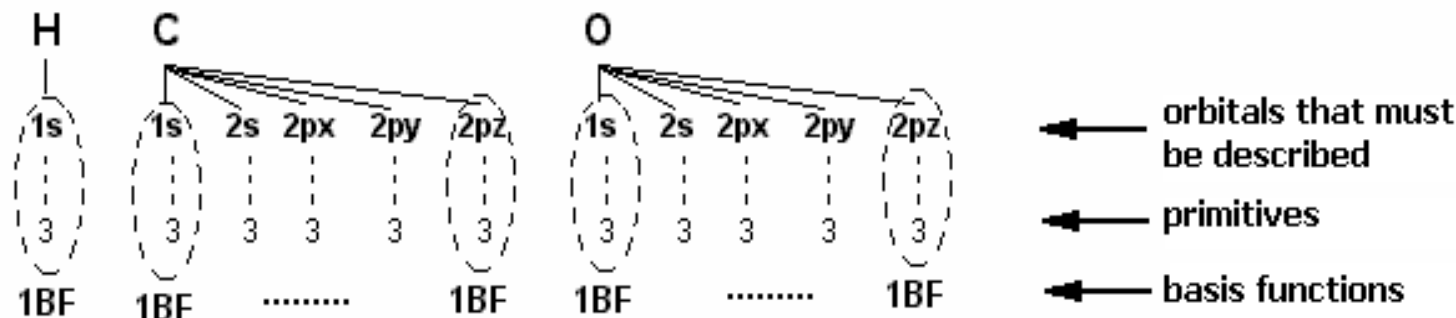
D 1 1.00

0.8000000000D+00 0.1000000000D+01

Basis sets

How many basis functions?

1,2-Benzosemiquinona $C_6H_4O_2$ STO-3G



Number of primitives and basis functions for 1,2-Benzosemiquinone free radical with the STO-3G basis set

Primitives:

atom C: nr.primitives = 15 x nr. atoms = 6 → 90
 atom H: nr.primitives = 3 x nr. atoms = 4 → 12
 atom O: nr.primitives = 15 x nr. atoms = 2 → 30

TOTAL: 132 GTO primitives

Basis functions:

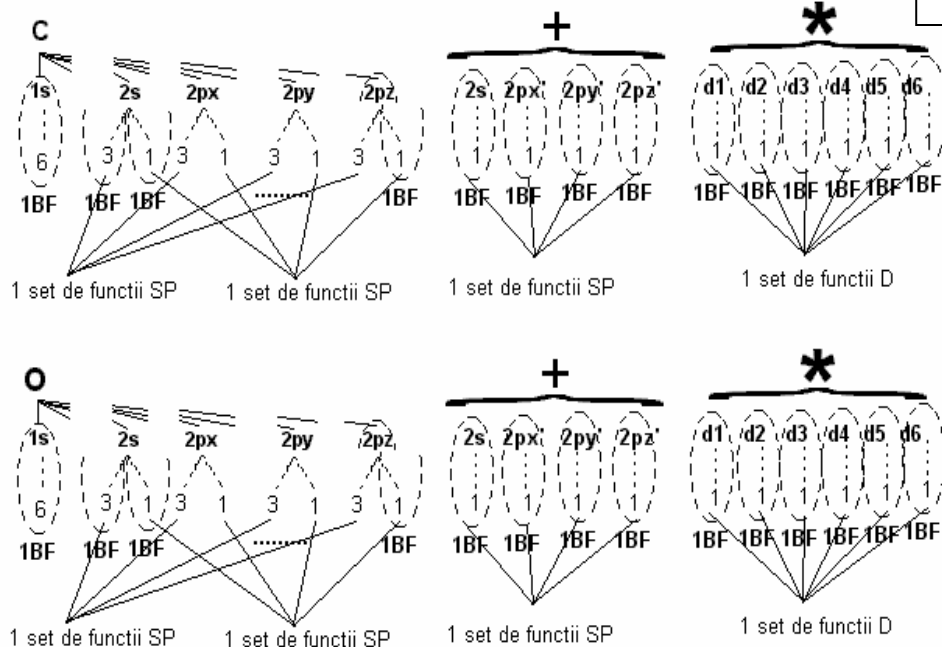
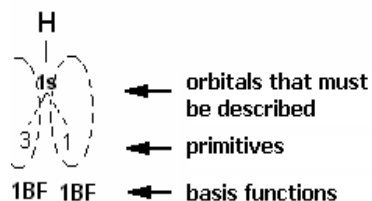
atom C: nr. BF = 5 x nr.atoms = 6 → 30
 atom H: nr. BF = 1 x nr.atoms = 4 → 4
 atom O: nr. BF = 5 x nr.atoms = 2 → 10

TOTAL: 44BF

Basis sets

How many basis functions?

1,2-Benzosemiquinona $C_6H_4O_2$ 6-31+G* (6-31+G(d))



Number of primitives and basis functions for 1,2-Benzosemiquinone free radical with the 6-31+G(d) basis set

Primitives:

atom C: nr.primitives = 32 x nr. atoms = 6 → 192
atom H: nr.primitives = 4 x nr. atoms = 4 → 16
atom O: nr.primitives = 32 x nr. atoms = 2 → 64
TOTAL: 272 GTO

primitives

Basis functions:

atom C: nr. BF = 19 x nr.atoms = 6 → 114
atom H: nr. BF = 2 x nr.atoms = 4 → 8
atom O: nr. BF = 19 x nr.atoms = 2 → 38
TOTAL: 160BF

Pople Style Basis Sets

k-nlm++G** or k-nlm++G(idf,jpd)

- *k* primitive GTOs for core electrons
- n* primitive GTOs for inner valence orbitals
- l* primitive GTOs for medium valence orbitals
- m* primitive GTOs for outer valence orbitals

E.g., 3-21G, 6-31G,
and 6-311G

- + means 1 set of P (SP) diffuse functions added to heavy atoms.
- ++ means 1 set of P (SP) diffuse functions added to heavy atoms and 1 s diffuse function added to H atom.

E.g., 6-31+G,
6-31++G

- * means 1 set of d polarization functions added to heavy atoms.
- ** means 1 set of d polarization functions added to heavy atoms and 1 set of p (sp) polarization functions added to H atom.

E.g., 6-31G* or
6-31G(d)

- *idf* means *i* d sets and 1 f set of polarization functions added to heavy atoms.
- idf,jpd* means *i* d sets and 1 f set polarization functions added to heavy atoms and *j* p sets and 1 d set of polarization functions added to H atom.

E.g., 6-31+G(d,p)

Common Basis Sets

- **Pople's Basis Sets**

- **3-21G**

- 3 primitive GTO for core electrons, 2 for inner and 1 for outer valence orbitals

- Preliminary geometry optimization; Poor for energy

- Common moderate basis set**

- 6-31G

- 6-31G(d) -> “spectroscopic” basis set

- 6-31G(d,p)

- More flexible basis sets**

- 6-31+G(d,p)

- Good for geometry and energy

- 6-311+G(2df,2p)

- Good for geometry and accurate energy

Dunning's Correlation-consistent Basis Sets

The basis sets are designated as either:

- cc-pVXZ $X_{\max}=8$
- aug-cc-pVXZ.

'cc' means "correlation consistent".

'p' means "polarization functions added".

'aug' means "augmented" with (essentially) diffuse functions.

'VXZ' means "valence-X-zeta" where X could be any one of the following

D' for "double", 'T' for "triple", Q for "quadruple", or 5 or 6, *etc.*

- Systematically converge the correlation energy to the basis set limit.
- Work typically with high-level electron-correlated wave function methods.

Plane wave basis sets-In addition to localized basis sets, plane wave basis sets can also be used in quantum chemical simulations. Typically, a finite number of plane wave functions are used, below a specific cutoff energy which is chosen for a certain calculation.

- **used (recommended) for periodical calculations**

Complete basis set extrapolation limit for electronic structure calculations: Energetic and nonenergetic properties of HeBr and HeBr₂ van der Waals dimers

M. P. de Lara-Castells

Quantum Theory Project, Department of Chemistry and Physics, University of Florida, Gainesville, Florida 32611-8535

R. V. Krems

Department of Chemistry, Physical Chemistry, Göteborg University, SE-412 96, Göteborg, Sweden

A. A. Buchachenko,⁸⁰ G. Delgado-Barnó, and P. Villarreal

Instituto de Matemáticas y Física Fundamental (C.S.I.C.), Serrano 123, E-28006-Madrid, Spain

$$f(n) = f_{\text{CBS}} + \alpha \exp^{-\beta n},$$

$$f(n) = f_{\text{CBS}} + \frac{B}{(n + \frac{1}{2})^4},$$

$$f(n) = f_{\text{CBS}} + \frac{B}{(n + \frac{1}{2})^4} + \frac{C}{(n + \frac{1}{2})^6},$$

$$f(n) = f_{\text{CBS}} + \frac{B}{(n + \frac{1}{2})^a},$$

$$E^{\text{cor}} = E_z^{\text{cor}} + A n^{-3}$$

$$f(n) = f_{\text{CBS}} + B \exp^{-(n-1)} + C \exp^{-(n-1)^2}$$

n is the cardinal number of the basis set: $n=2$ for DZ, 3 for TZ, etc., $f(n)$ is the property of interest and f_{CBS} is the asymptotic value, which is taken to approximate the CBS limit.

f_{CBS} , α , B , C – fitting parameters

property of interest

total energy
interaction energy
geometrical
parameters
dipole moment
polarizability
etc.

Basis sets

Complete basis set extrapolation Example

TABLE I. Equilibrium properties derived from the nonrelativistic FCP-corrected HeBr interaction energies. Percentage accuracy given in parentheses for V5Ze values is with respect to V5Z value. Results corresponding to the calculations without frozen-core approximation are given between brackets in the second line for each basis set.

Basis set	V_{Σ}		V_{Π}		$T_e, \text{ cm}^{-1}$
	$R_e, \text{ \AA}$	$D_e, \text{ cm}^{-1}$	$R_e, \text{ \AA}$	$D_e, \text{ cm}^{-1}$	
VDZ	3.779 [3.779]	14.083 [14.061]	4.289 [4.299]	7.749 [7.974]	17.57 [17.58]
VTZ	3.641 [3.652]	20.372 [20.761]	4.066 [4.076]	12.098 [12.159]	20.72 [19.76]
VQZ	3.549 [3.558]	25.156 [25.101]	4.020 [4.025]	14.182 [14.183]	29.44 [28.63]
V5Z	3.535	26.991	4.002	14.979	30.30
V5Ze/mix	3.525 (0.2) [3.526]	27.291 (1.1) [28.058]	4.002 (0.0) [3.995]	14.979 (0.0) [15.419]	32.47 (7) [32.44]
/exp	3.514 (0.6)	28.457 (5.2)	4.013 (0.3)	14.934 (0.3)	35.37 (17)
$/n^{-3}$	3.557 (0.6)	25.169 (6.7)	4.023 (0.5)	14.186 (5.2)	28.58 (6)
$/(n+1/2)^{-4}$	3.535 (0.0)	26.882 (0.4)	4.002 (0.0)	14.880 (0.7)	30.96 (2)
$/(n+1/2)^{-4,-6}$	3.525 (0.2)	27.361 (1.4)	4.002 (0.0)	14.986 (0.0)	32.55 (7)
CBSe ^a /mix	3.514 [3.526]	28.573 [28.059]	3.992 [3.995]	15.451 [15.419]	33.90 [32.44]
CBSe ^b /mix	3.514	28.080	3.970	15.206	32.45
CBSe/ $(n+1/2)^{-4}$	3.514	28.524	3.992	15.459	33.71
DIM-inverted	3.74	26.6	3.99	14.0	15.64

^aUsing the basis sets $D \sim Q$ in the extrapolation procedure.

^bUsing the basis sets $T \sim 5$ in the extrapolation procedure.

M. P. de Lara-Castells

J. Chem. Phys., Vol. 115, No. 22, 8 December 2001

Effective core potentials (ECPs)

Core electrons, which are not chemically very important, require a large number of basis functions for an accurate description of their orbitals. This normally applies to third and higher row elements.

Core (inner) orbitals are in most cases not affected significantly by changes in chemical bonding. Effective Core Potential (ECP) approaches allow treatment of inner shell electrons as if they were some averaged potential rather than actual particles.

This separation suggests that inner electrons can be ignored in a large number of cases.

The use of a pseudo-potential that approximates the potential felt by the valence electrons was first proposed by Fermi in 1934. In 1935 Helman suggested the following potential for the valence electron of potassium:

$$w(r) = -\frac{1}{r} + \frac{2.74}{r}e^{-1.16r}$$

Using pseudo-potentials, the need for core basis functions, which usually require a large number of primitives to describe them is eliminated.

It is quite easy to incorporate relativistic effects into ECP, while all-electron relativistic computations are very expensive. The relativistic effects are very **important in describing heavier atoms**, and luckily ECP's simplify calculations and at the same time make them more accurate with popular non-relativistic *ab initio* packages.

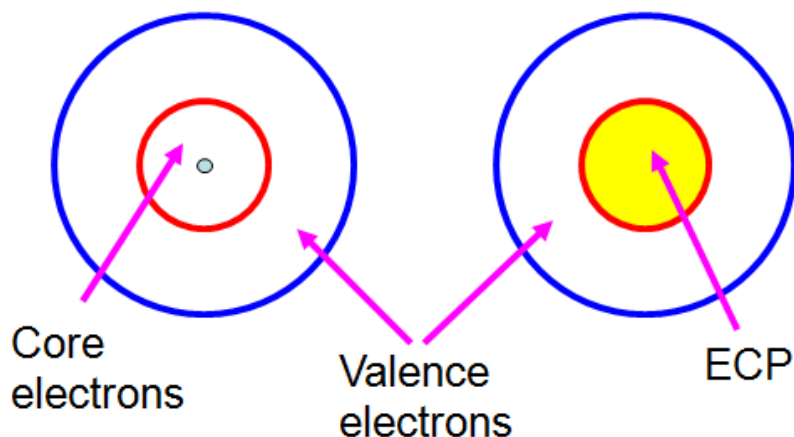
For the rest of electrons (i.e. valence electrons), basis functions must be provided.

These are **special basis sets optimized for the use with specific ECP's**.

ECP potentials are specified as parameters of the following equation:

$$U_{\text{ECP}}(r) = \sum_{i=1}^p d_i r_0^{n_i} e^{-\xi_i r_0^2}$$

where p is the dimension of the expansion d_i are the coefficients for the expansion terms, r_0 is the distance from nucleus and ξ_i represents the exponents for each term.



- Saving computational effort
- Taking care of relativistic effects
- Important for heavy atoms, e.g., transition metal atoms

Examples:

CEP-4G, CEP-31G, CEP-121G, LANL2MB (STO-3G 1st row), LANL2DZ (D95V 1st row), SHC (D95V 1st row), SDD

Example:

Ag with Lanl2DZ basis set and ECP

Ag: 1s², 2s², 2p⁶, 3s², 3p⁶, 4s², 3d¹⁰, 4p⁶, 5s¹, 4d¹⁰

Valence electrons: 4s², 4p⁶, 4d¹⁰, 5s¹

⇒ 19 valence electrons

⇒ 22 valence basis functions

Explain!!

Basis sets

Type of ECPs

CEP-4G, CEP-31G, CEP-121G, LANL2MB, LANL2DZ, SDD

complexPd1.chk

#P Opt B3LYP/gen pseudo=read

complex Pd v1

O 2

C 8.89318310	9.90388210	6.72569337
C 9.52931379	8.77525770	6.27102032
H 9.29586123	7.93893890	6.60431879
C 10.52592748	8.89096200	5.30965653
H 10.95942133	8.13380930	4.98695425
C 10.85850598	10.13123090	4.84438728
H 11.51852449	10.22866610	4.19609286
C 10.20972534	11.23549650	5.34144511

etc.

H 4.15752044	17.83312399	10.48668123
H 5.63848578	17.14049639	11.10318367

N C O H O

6-31G(d)

Pd 0

CEP-121G

Pd 0

CEP-121G



Recommendations for basis set selection

- Always a **compromise** between accuracy and computational cost!
- With the **increase of basis set** size, calculated **energy will converge** (complete basis set (CBS) limit).
- **Special cases** (anion, transition metal, transition state)
- Use **smaller basis sets for preliminary calculations** and for heavy duties (e.g., geometry optimizations), and use **larger basis sets to refine** calculations.
- Use larger basis sets for critical atoms (e.g., atoms directly involved in bond-breaking/forming), and use smaller basis sets for unimportant atoms (e.g., atoms distant away from active site). **(ONIOM method)**
- Use **popular and recommended basis sets**. They have been tested a lot and shown to be good for certain types of calculations.
- Special properties:
 - **IGLO** basis sets for **NMR** spectra
 - **EPR** style basis sets for **EPR** spectra (EPR-II, EPR-III of Barone et al.)

Basis sets

Do you need a basis set?

EMSL Gaussian Basis Set Exchange

<http://www.emsl.pnl.gov/forms/basisform.html>

The screenshot shows the EMSL Basis Set Exchange portal. The browser address bar displays <https://bse.pnl.gov/bse/portal>. The page features the EMSL Office of Science logo and the 'BASIS SET EXCHANGE' title. A search bar on the left lists various basis sets, with 'pcJ-3' selected and circled in red. The main content area shows a periodic table with elements H, C, N, and O circled in red, indicating the search criteria. Below the table, the 'Format' dropdown is set to 'Gaussian94' and the 'Optimized General Contractions' checkbox is checked, both circled in red. A 'Get Basis Set' button is also circled in red. The top right includes fields for 'Username' and 'Password' with 'Login' and 'Become a Contributor' buttons.

"pcJ-3" Basis Set Information

Summary:
Primary Developer:
Last Modified:

N/A
Frank Jensen
Mon, 27 Dec 2010 21:28:22 GMT

Contributor:
Curation Status:

Frank Jensen
published

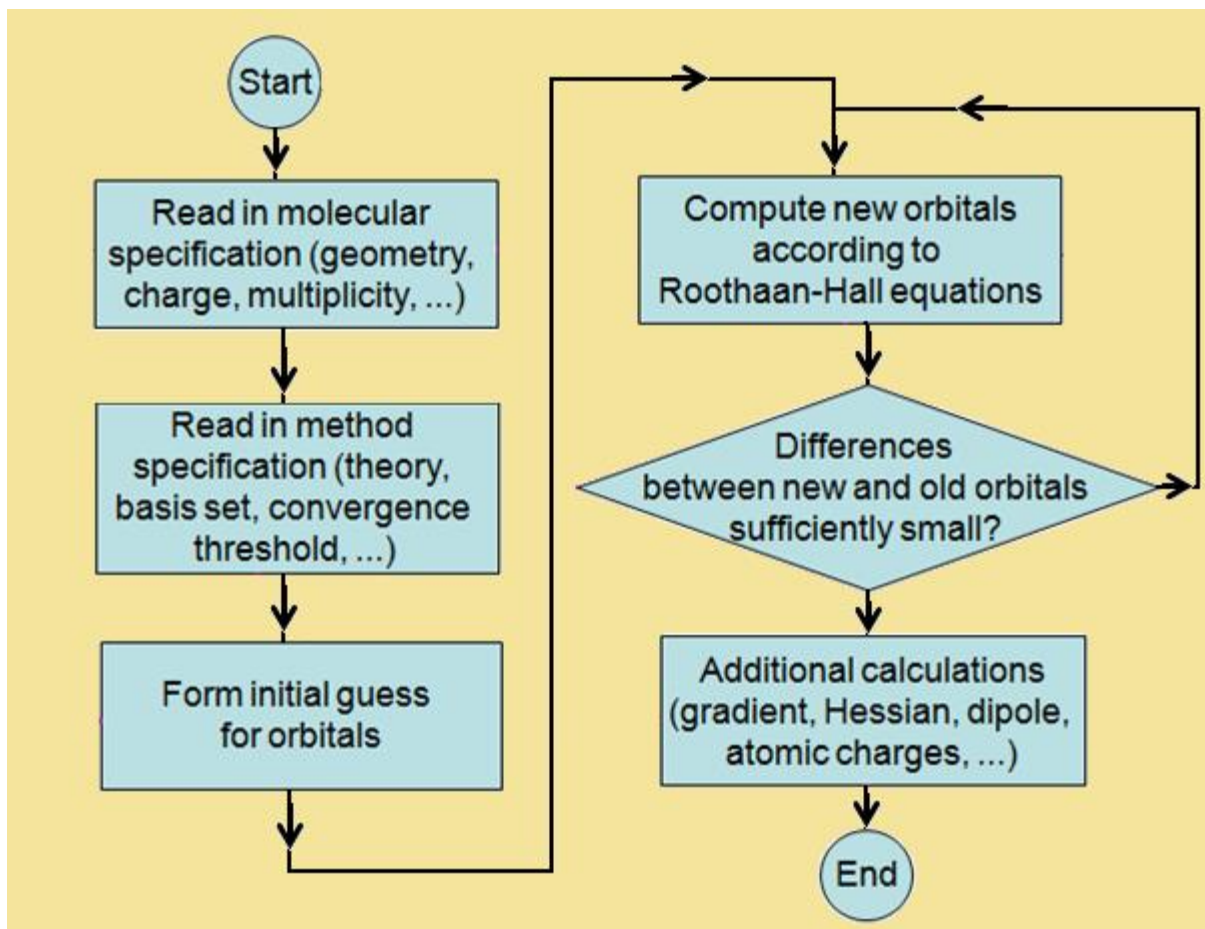
[More information...](#)
[User annotations...](#)

When publishing results obtained from use of the Basis Set Exchange (BSE) software and the EMSL Basis Set Library, please cite:

The Role of Databases in Support of Computational Chemistry Calculations
Feller, D., J. Comp. Chem., 17(13), 1571-1586, 1996.

Basis Set Exchange: A Community Database for Computational Sciences
Schuchardt, K.L., Didier, B.T., Elsethagen, T., Sun, L., Gurumoorhi, V., Chase, J., Li, J., and Windus, T.L.
J. Chem. Inf. Model. 47(3), 1035-1037, 2007. doi:10.1021/ci060510i

SCF Algorithm



Molecular properties as derivatives of the energy

Consider a molecule in an external electric field ϵ .

$$E(\epsilon) = E(\epsilon = 0) + \epsilon \left. \frac{dE}{d\epsilon} \right|_{\epsilon=0} + \frac{1}{2} \epsilon^2 \left. \frac{d^2 E}{d\epsilon^2} \right|_{\epsilon=0} + \dots$$

- Dipole moment (μ) $\mu = - \left. \frac{dE}{d\epsilon} \right|_{\epsilon=0}$
- Polarizability (α) $\alpha = - \left. \frac{d^2 E}{d\epsilon^2} \right|_{\epsilon=0}$
- First hyperpolarizability (β) $\beta = - \left. \frac{d^3 E}{d\epsilon^3} \right|_{\epsilon=0}$

$\frac{dE}{d\epsilon_\alpha}$	dipole moment; in a similar way also multipole moments, electric field gradients, etc.
$\frac{d^2 E}{d\epsilon_\alpha d\epsilon_\beta}$	polarizability
$\frac{d^3 E}{d\epsilon_\alpha d\epsilon_\beta d\epsilon_\gamma}$	(first) hyperpolarizability
$\frac{dE}{dx_i}$	forces on nuclei
$\frac{d^2 E}{dx_i dx_j}$	harmonic force constants; harmonic vibrational frequencies
$\frac{d^3 E}{dx_i dx_j dx_k}$	cubic force constants; anharmonic corrections to distances and rotational constants
$\frac{d^4 E}{dx_i dx_j dx_k dx_l}$	quartic force constants; anharmonic corrections to vibrational frequencies
$\frac{d^2 E}{dx_i d\epsilon_\alpha}$	dipole derivatives; infrared intensities
$\frac{d^3 E}{dx_i d\epsilon_\alpha d\epsilon_\beta}$	polarizability derivatives; Raman intensities

$\frac{d^2 E}{dB_\alpha dB_\beta}$	magnetizability
$\frac{d^2 E}{dI_{\alpha} dB_\beta}$	nuclear magnetic shielding tensor; relative NMR shifts
$\frac{d^2 E}{dI_{\alpha} dI_{\beta}}$	indirect spin-spin coupling constants
$\frac{d^2 E}{dB_\alpha dJ_\beta}$	rotational g-tensor; rotational spectra in magnetic field
$\frac{d^2 E}{dI_{\alpha} dJ_\beta}$	nuclear spin-rotation tensor; fine structure in rotational spectra
$\frac{dE}{dS_\alpha}$	spin density; hyperfine interaction constants
$\frac{d^2 E}{dS_\alpha dS_\beta}$	electronic g-tensor

Example:

$$\frac{dE}{dx} = \sum_{\mu\nu} P_{\mu\nu} \frac{\partial h_{\mu\nu}}{\partial x} + \frac{1}{2} \sum_{\mu\nu\sigma\rho} P_{\mu\nu} P_{\sigma\rho} \left(\frac{\partial \langle \mu\nu | \sigma\rho \rangle}{\partial x} - \frac{1}{2} \frac{\partial \langle \mu\nu | \sigma\rho \rangle}{\partial x} \right) - \sum_{\mu\nu} W_{\mu\nu} \frac{\partial S_{\mu\nu}}{\partial x}$$

See: G. Gauss, *Modern Methods and Algorithms of Quantum Chemistry*, J. Grotendorst (Ed.), John von Neumann Institute for Computing, Jülich, NIC Series, Vol. 1, ISBN 3-00-005618-1, pp. 509-560, 2000.

Basis Set Superposition Error

See C. David Sherrill, <http://vergil.chemistry.gatech.edu/notes/cp.pdf>

Interaction energy in a complex or cluster is computed as the difference between the energy of the complex and the total energy of the (noninteracting) monomers, which form the complex.

$$A+B \rightarrow AB \quad \Delta E = E_{AB} - E_A - E_B \quad (\Delta E < 0)$$

ΔE – often too large for weakly (H-bond, dispersion) bound systems
- more evident for small basis sets

He dimer – interaction energy

method	r(A)	E _{int} (kJ/mol)
RHF/cc-pVDZ	3.211	-0.0038
RHF/cc-pVTZ	3.662	-0.0023
RHF/cc-pVQZ	3.887	-0.0011
RHF/cc-pV5Z	4.131	-0.0005

Data from: H. Zipse, <http://www.cup.uni-muenchen.de/oc/zipse/basis-set-superposition-error-bsse.html>



Basis set superposition error

$$\Delta E = f(\text{\#BFs})$$

- decreases by increasing the the #BFs (for non-correlated methods)

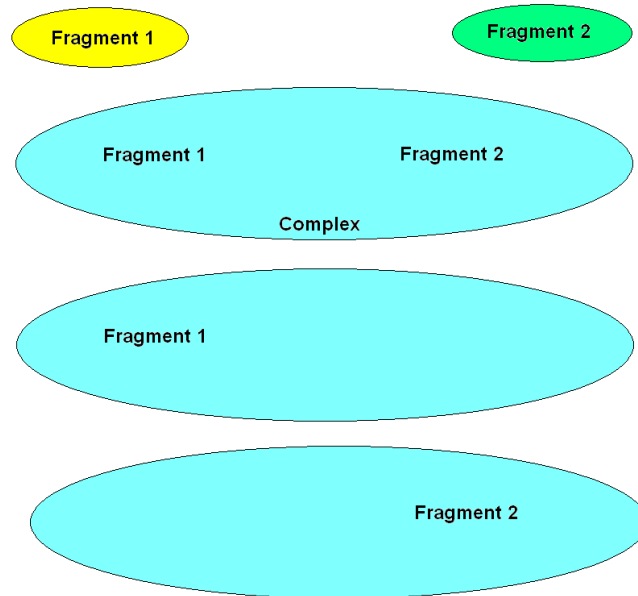
When the two fragments approaches each other

F1 is artificially stabilized because it uses extra BF's centered on F2 and vice-versa

=> inside the complex, **F1 and F2 are better described** than in the case when they are infinitely separated

=> inconsistent treatment of monomers = source of BSSE

In the limit of CBS => BSSE -> 0



Boys and Bernardi (Mol. Phys., 19, 553 (1970)) suggested an elegant method, which they named the *counterpoise* (CP) correction, to cope with this problem. According to this method, **the individual monomers are calculated using the basis set of the complex.** Since the energies of the individual molecules usually are lower when computed within the composite basis of the interacting molecules rather than in the monomer's own basis, it follows that the **CP corrected interaction energies are smaller** (in magnitude) than the uncorrected ones.

Counter Poise method for BSSE correction

(Boys & Bernardi, Mol.Phys., 19, 553 (1970))

The complex AB is formed by two fragments A and B: **A+B -> AB**

Binding (stabilization) energy:

- energy required to form a complex AB from the infinitely separated monomers

$$\Delta E(AB) = E_{AB}(AB) - E_A(A) - E_B(B)$$

As the monomers get closer, their geometries change. Thus, the binding energy will be written as:

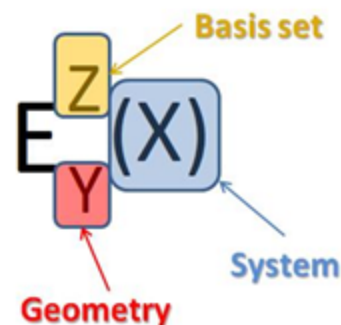
$$\Delta E(AB) = \underbrace{\Delta E_{\text{int}}(AB)}_{\text{interaction energy}} + \underbrace{\Delta E_{\text{rel}}(AB)}_{\text{relaxation(deformation) energy}}$$

Interaction energy

- energy required to form a complex AB formed from two infinitely separated rigid monomers (their geometries do not change as the monomers get closer)
- to account for BSSE, we use for monomers the same basis set as for the complex

$$\Delta E_{\text{int}}(AB) = E_{AB}^{AB}(AB) - E_A^{AB}(A) - E_B^{AB}(B)$$

- **rigid monomers**



Relaxation (deformation) energy

- the energy required to change the geometry of A and B monomers from their equilibrium structures to those they have in the complex
- calculated in the monomers basis sets only

$$\Delta E_{rel}(AB) = E_{AB}^A(A) - E_A^A(A) + E_{AB}^B(B) - E_B^B(B)$$

If the monomers' geometries change as they approach each other and form the complex, the interaction energy will be:

$$\Delta E_{int}(AB) = E_{AB}^{AB}(AB) - E_{AB}^{AB}(A) - E_{BB}^{AB}(B)$$

- deformed monomers

⇒ CP corrected binding energy ($\Delta E = \Delta E_{int} + \Delta E_{rel}$)

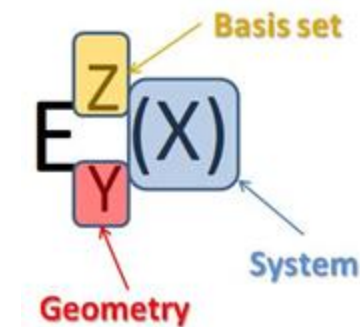
$$\Delta E^{CP} = \underbrace{E_{AB}^{AB}(AB) - E_{AB}^{AB}(A) - E_{AB}^{AB}(B)}_{\text{interaction energy}} + \underbrace{E_{AB}^A(A) - E_A^A(A) + E_{AB}^B(B) - E_B^B(B)}_{\text{relaxation energy}} =$$

$$= E_{AB}^{AB}(AB) - E_A^A(A) - E_B^B(B) + \underbrace{E_{AB}^A(A) - E_{AB}^{AB}(A)}_{\substack{\text{relaxation energy of A} \\ >0}} + \underbrace{E_{AB}^B(B) - E_{AB}^{AB}(B)}_{\substack{\text{relaxation energy of B} \\ >0}}$$

$$\Rightarrow \Delta E^{CP} = \Delta E^{uncorrected} + \delta_{BSSE}$$

$$\delta_{BSSE} > 0$$

$$\delta_{BSSE} = E_{AB}^A(A) - \underbrace{E_{AB}^{AB}(A)}_{\text{ghost atoms on B}} + E_{AB}^B(B) - \underbrace{E_{AB}^{AB}(B)}_{\text{ghost atoms on A}}$$



Note:

7 calculations are needed for obtaining CP corrected binding energy for a heterodimer only 4 in case of a homodimer

Algorithm:

- 1) optimize A fragment $\Rightarrow E_A^A(A)$
- 2) optimize B fragment $\Rightarrow E_B^B(B) \Rightarrow \Delta E^{unc} = \dots$
- 3) optimize the complex AB $\Rightarrow E_{AB}^{AB}(AB)$
- 4) on the optimized AB geometry of the complex, make a single point calculation with counter=2
(don't forget to assign each atom to a fragment)

$$\Rightarrow \delta_{BSSE}$$

Then:

$$\Delta E^{CP \text{ corrected}} = \Delta E^{unc.} + \delta_{BSSE}$$

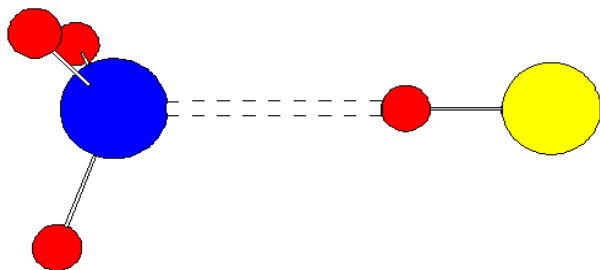
Basis set superposition error:

- ☐ Tends to zero as the fragment's basis set approaches completeness
- ☐ It is a positive value
- ☐ Depends on the geometrical parameters of the complex

Quantum chemical calculations are frequently used to estimate strengths of hydrogen bonds. We can distinguish between intermolecular and intra-molecular hydrogen bonds. The first of these are usually much more straightforward to deal with.

1. Intermolecular Hydrogen Bond energies

In this case it is normal to define the hydrogen bond energy as the energy of the hydrogen bonded complex minus the energies of the constituent molecules/ions. Let us first consider a simple example with high (C_{3v}) symmetry – $H_3N...HF$



Practical aspects

1. Add in the section route:

Counter=n

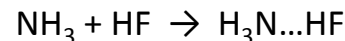
where n - # of fragments

2. Specify the groups (1, 2, ...) to which the atoms belong)

In the geometry specification section each atom's line will be finished by an index specifying the fragment to which it belongs

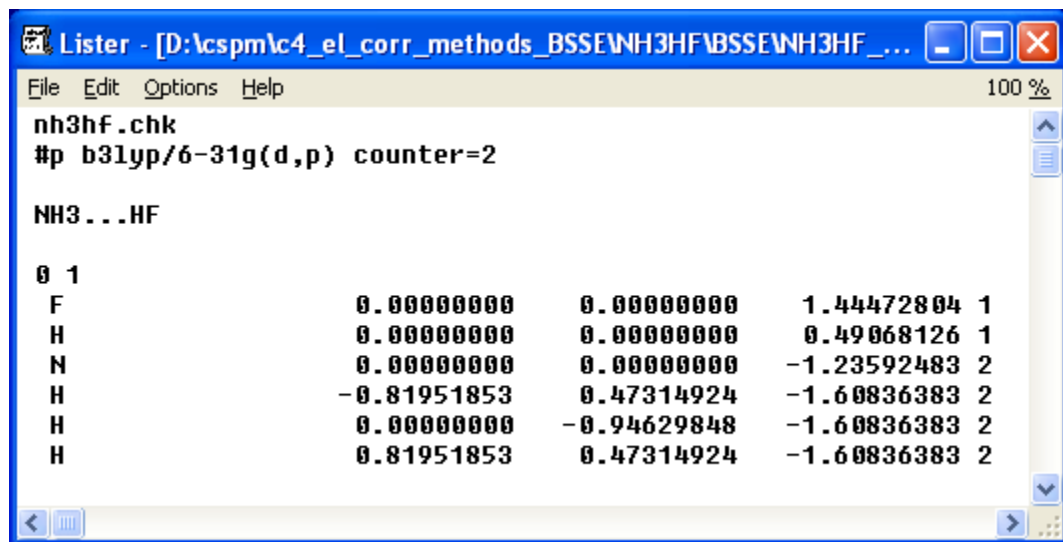
	Electronic energy (a.u.)
NH_3	-56.19554
HF	-100.01169
$H_3N...HF$	-156.22607

E_{HB}



$$E_{HB} = 2625.5 \times (156.22607 - 100.01169 - 56.19554) = 38.8 \text{ kJ/mol}$$

-not corrected value



```
nh3hf.chk
#p b31yp/6-31g(d,p) counter=2

NH3...HF

0 1
F      0.00000000  0.00000000  1.44472804  1
H      0.00000000  0.00000000  0.49068126  1
N      0.00000000  0.00000000 -1.23592483  2
H     -0.81951853  0.47314924 -1.60836383  2
H      0.00000000 -0.94629848 -1.60836383  2
H      0.81951853  0.47314924 -1.60836383  2
```

References:

1. Pedro Salvador Sedano, Implementation and Application of BSSE Schemes to the Theoretical Modeling of Weak Intermolecular Interactions, PhD Thesis, Department of Chemistry and Institute of Computational Chemistry, University of Girona;
http://www.tdx.cesca.es/TESIS_UdG/AVAILABLE/TDX-0228102-130339//02tesis_corrected.pdf
2. M. L. SENENT, S. WILSON, Intramolecular Basis Set Superposition Errors, International Journal of Quantum Chemistry, Vol. 82, 282–292 (2001)
3. A. BENDE, Á. VIBÓK, G. J. HALÁSZ, S. SUHAI, BSSE-Free Description of the Formamide Dimers, International Journal of Quantum Chemistry, Vol. 84, 617–622 (2001)

Basis sets

Example

```
%chk=MP2_D2h_2PTDCI_BSSE.chk
```

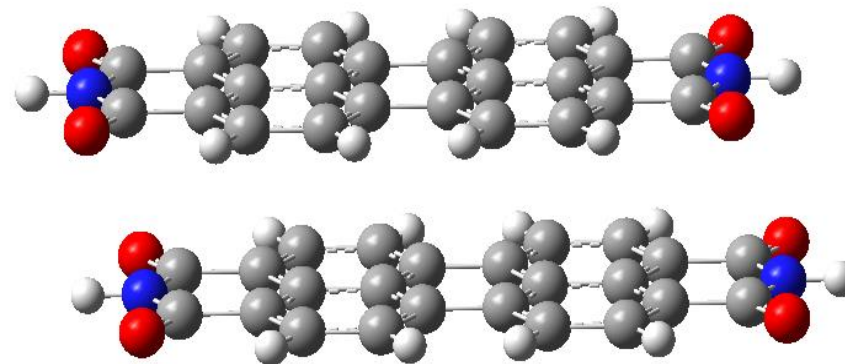
```
%mem=4GB
```

```
%nprocshared=4
```

```
#p mp2/6-31g(d) Counter=2
```

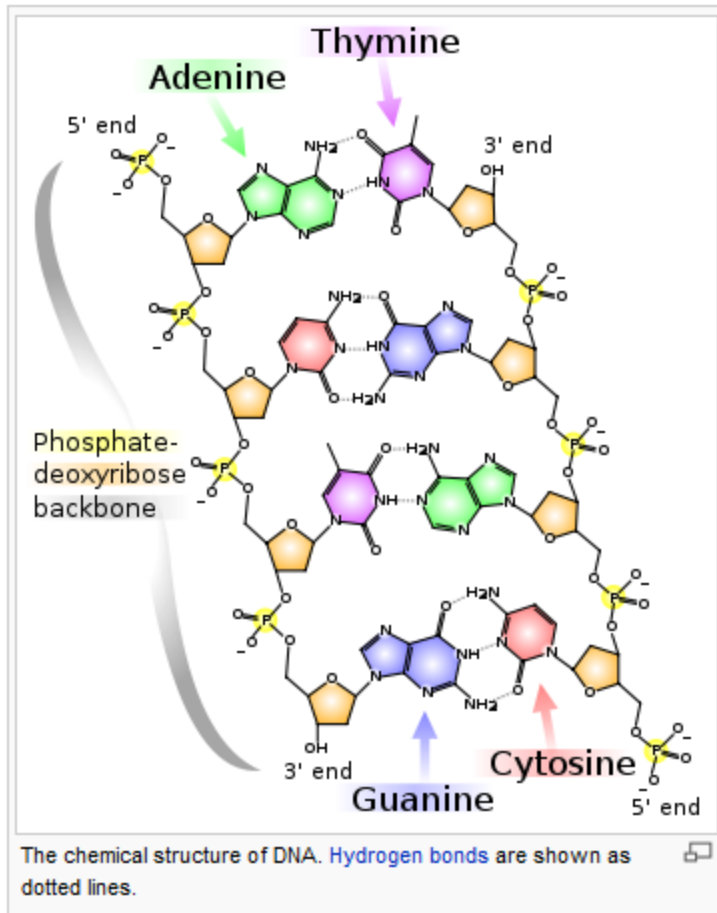
```
2PTDCI Popt BSSE correction
```

```
0 1
C      -1.99898000    1.18154000    0.99560500  1
C      -1.26309500    0.11795300    1.60980400  1
C       0.20480800    0.19836000    1.66469500  1
C       0.86397000    1.33836500    1.10266300  1
C       0.12808500    2.40195200    0.48846500  1
C      -1.33981800    2.32154400    0.43357300  1
C       0.98224400   -0.80350700    2.24885600  1
.....
C      -0.86397000   -1.33836500   -1.10266300  2
C      -0.12808500   -2.40195200   -0.48846500  2
C       1.33981800   -2.32154400   -0.43357300  2
C      -2.29155500   -1.41656400   -1.15604700  2
C      -2.95995400   -2.54054700   -0.60314400  2
C      -3.03552800   -0.36994800   -1.76176400  2
C      -0.20480800   -0.19836000   -1.66469500  2
C      -2.22846900   -3.55635000   -0.01463400  2
C      -4.43889800   -2.64137300   -0.64791900  2
2
.....
```

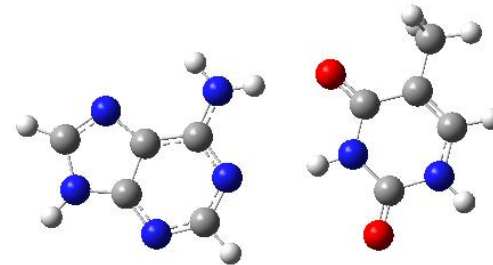


Exercise

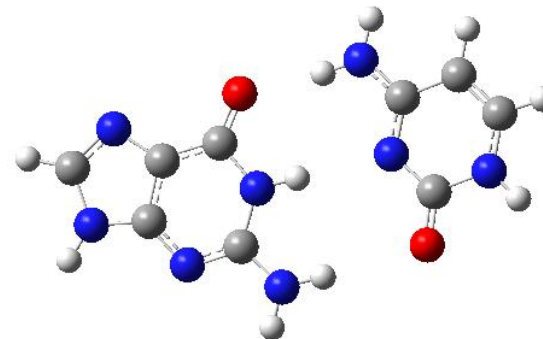
Calculate the interaction energies in the DNA base pairs Adenine-Thymine and Cytosine-Guanine. Consider the BSSE



You can look for pdb files of DNA bases at:
<http://www.biocheminfo.org/klotho/pdb/>



Adenine-Thymine base pair



Guanine-Cytosine base pair