Introduction to Computational Chemistry

Vesa Hänninen
Laboratory of Physical Chemistry
Chemicum 4th floor
vesa.hanninen@helsinki.fi

September 25, 2012

Basis sets

A basis set is a set of functions used to create the molecular orbitals, which are expanded as a linear combination with coefficients to be determined.

- Usually these functions are centered on atoms, but functions centered in bonds or lone pairs have been used.
- Additionally, basis sets composed of sets of plane waves are often used, especially in calculations involving systems with periodic boundary conditions.

Quantum chemical calculations are typically performed within a finite set of basis functions.

- These basis functions are usually not the exact atomic orbitals, even for the corresponding hydrogen-like atoms, due to approximations and simplifications of their analytic formulas.
- If the finite basis is expanded towards an infinite complete set of functions, calculations using such a basis set are said to approach the basis set limit.

Slater type orbitals (STOs)

In the early days of quantum chemistry so-called Slater type orbitals (STOs) were used as basis functions due to their similarity with the eigenfunctions of the hydrogen atom. Their general definition is

$$S^{\zeta}_{nlm}(r,\vartheta,\varphi) = N r^{n-1} e^{-\zeta r} Y_l^m(\vartheta,\varphi) \tag{1}$$

with N being a normalization factor and Y_l^m being the spherical harmonics.

- STOs have an advantage in that they have direct physical interpretation and thus are naturally good basis for molecular orbitals.
- From a computational point of view the STOs have the severe shortcoming that most of the required integrals needed in the course of the SCF procedure must be calculated numerically which drastically decreases the speed of a computation.
- Recently, new numerical methods have been developed that allow very efficient SCF calculations using STOs

Gaussian type orbitals (GTOs)

STOs can be approximated as linear combinations of Gaussian orbitals. Gaussian type orbitals (GTOs) are defined as

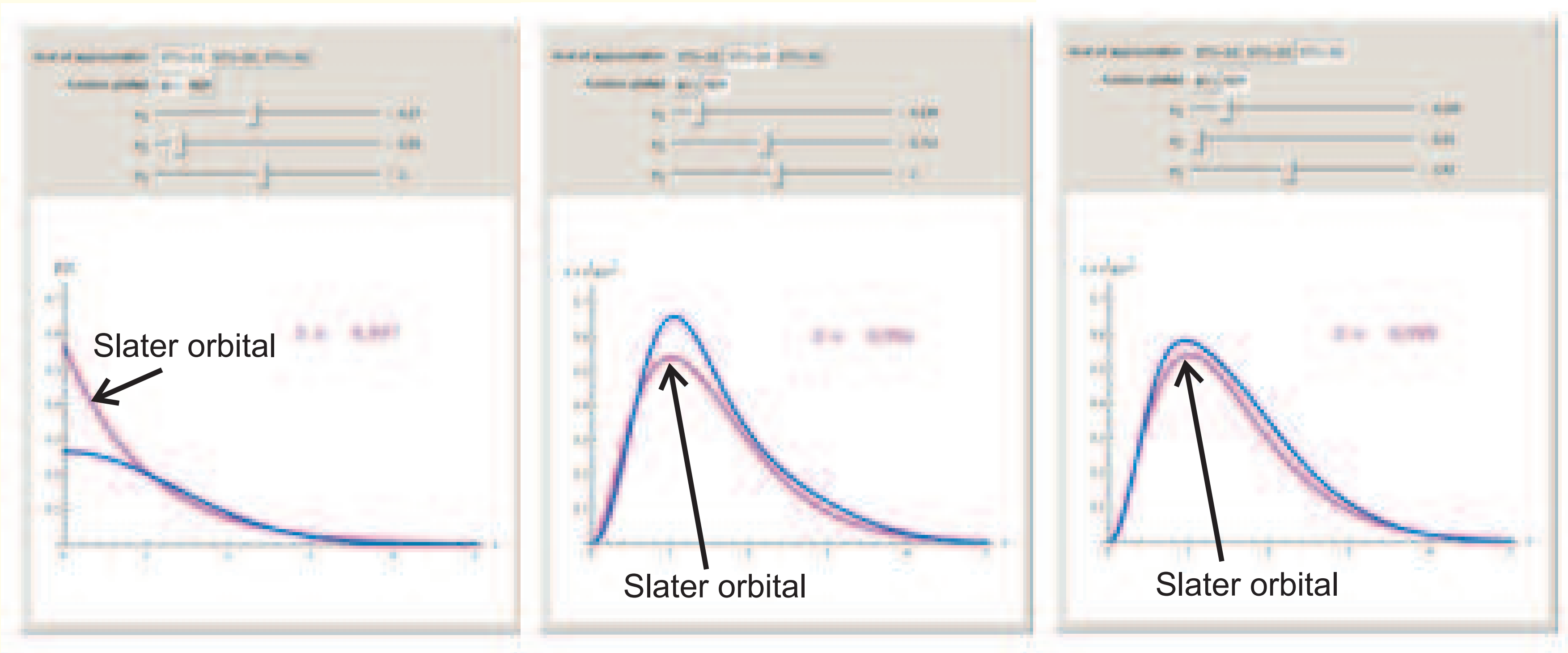
$$G_{ijk}^{\alpha,\mathbf{R}}(\mathbf{r}) = N_{ijk}^{\alpha}(x - R_1)^i (y - R_2)^j (z - R_3)^k e^{-\alpha(\mathbf{r}-\mathbf{R})^2}. \qquad (2)$$

N_{ijk}^{α} is a normalization factor, $\mathbf{R}$ and α are called the center and the exponent of the Gaussian function, respectively. GTOs are not really orbitals. They are simpler functions and are frequently called **gaussian primitives.**

- Gaussian primitives are usually obtained from quantum calculations on atoms (i.e. Hartree-Fock or Hartree-Fock plus some correlated calculations, e.g. CI).
- Typically, the exponents are varied until the lowest total energy of the atom is achieved.
- For molecular calculations, these gaussian primitives have to be contracted, i.e., certain linear combinations of them will be used as basis functions.
- Such a basis function will have its coefficients and exponents fixed.

Gaussian type orbitals (GTOs)

The main difference to the STOs is that the variable r in the exponential function is squared. Generally the inaccuracy at the center nor the qualitatively different behaviour at long distances from the center have a marked influence on the results.



STO-1G STO-2G STO-3G

Figure 1: Slater type s- and p-orbitals approximated by Gaussian orbitals.

Gaussian type orbitals (GTOs)

To understand why integrals over GTOs can be carried out when analogous STO-based integrals are much more difficult, one must consider orbital products $\psi_a\psi_b$ and $\psi_c\psi_d$, where a, b, c, and d refer to different atomic centers. These products give rise to multi-center integrals, which often arise in polyatomic-molecule calculations, and which can not be efficiently performed when STOs are employed. For orbitals in the GTO form, such products involve $e^{-\alpha_a(\mathbf{r}-\mathbf{R}_a)^2}e^{-\alpha_c(\mathbf{r}-\mathbf{R}_c)^2}$ which can be rewritten as

$$e^{-\alpha_a(\mathbf{r}-\mathbf{R}_a)^2}e^{-\alpha_c(\mathbf{r}-\mathbf{R}_c)^2} = e^{-(\alpha_a+\alpha_c)(\mathbf{r}-\mathbf{R}')^2}e^{-\alpha'(\mathbf{R}_a-\mathbf{R}_c)^2} \tag{3}$$

where

$$\mathbf{R}' = \frac{\alpha_a\mathbf{R}_a + \alpha_c\mathbf{R}_c}{\alpha_a+\alpha_c} \tag{4}$$

and

$$\alpha' = \frac{\alpha_a\alpha_c}{\alpha_a+\alpha_c} \tag{5}$$

Thus, the product of two GTOs on different centers is equal to a single other GTO at center $\mathbf{R}'$ between the original centers. As a result, even a four-center integral over GTOs can be written as two-center two-electron integral. A similar reduction does not arise for STOs.

Gaussian type orbitals (GTOs)

The sum of the exponents of the Cartesian coordinates, $L = i + j + k$, is used to mark functions as s-type (L=0), p-type (L=1), d-type (L=2), and so on

- Unfortunately GTOs are not eigenfunctions of the squared angular momentum operator L^2.
- However, combinations of gaussians are able to approximate correct nodal properties of atomic orbitals by taking them with different signs. For example combining three d-type Cartesian GTOs yields a Cartesian GTO of s-type:

$$G_{200}^{\alpha,\mathbf{R}} + G_{020}^{\alpha,\mathbf{R}} + G_{002}^{\alpha,\mathbf{R}} \propto G_{000}^{\alpha,\mathbf{R}}. \tag{6}$$

- Today, there are hundreds of basis sets composed of GTOs. The smallest of these are called minimal basis sets, and they are typically composed of the minimum number of basis functions required to represent all of the electrons on each atom. The largest of these can contain literally dozens to hundreds of basis functions on each atom.

Minimal basis sets

A minimum basis set is one in which a single basis function is used for each orbital in a Hartree-Fock calculation on the free atom.

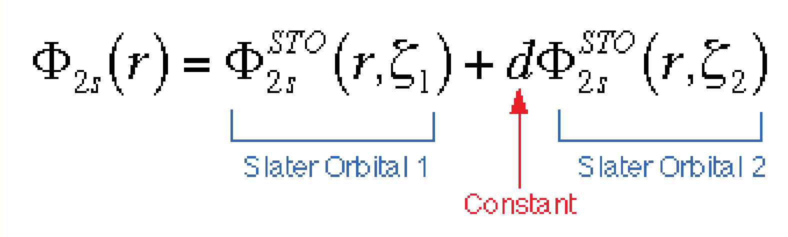
- However, for atoms such as lithium, basis functions of p type are added to the basis functions corresponding to the $1s$ and $2s$ orbitals of the free atom. For example, each atom in the first row of the periodic system (Li - Ne) would have a basis set of five functions (two s functions and three p functions).

The most common minimal basis set is STO-nG, where n is an integer. This n value represents the number GTOs used to approximate STO for both core and valence orbitals.

- Minimal basis sets typically give rough results that are insufficient for research-quality publication, but are much cheaper than their larger counterparts.
- Commonly used minimal basis sets of this type are: STO-3G, STO-4G, STO-6G

Extended basis sets

- The minimal basis sets are not flexible enough for accurate representation of orbitals
- **Solution:** Use multiple functions to represent each orbital
- For example, the double-zeta basis set allows us to treat each orbital separately when we conduct the Hartree-Fock calculation.



$$\Phi_{2s}(r) = \underbrace{\Phi_{2s}^{STO}(r,\zeta_1)}_{\text{Slater Orbital 1}} + \underset{\substack{\uparrow \\ \text{Constant}}}{d}\,\underbrace{\Phi_{2s}^{STO}(r,\zeta_2)}_{\text{Slater Orbital 2}}$$

Figure 2: $2s$ atomic orbital expressed as the sum of two STOs. The two equations are the same except for the value of ζ which accounts for how large the orbital is. The constant d determines how much each STO will count towards the final orbital.

Extended basis sets

The triple and quadruple-zeta basis sets work the same way, except use three and four Slater equations instead of two. The typical trade-off applies here as well, better accuracy...more time/work.
There are several different types of extended basis sets

- split-valence
- polarized sets
- diffuse sets
- correlation consistent sets

Split-valence basis sets

Pople's basis sets n-ijG or n-ijkG.

- n - number of primitives for the inner shells; ij or ijk - number of primitives for contractions in the valence shell. The ij notations describes sets of valence double zeta quality and ijk sets of valence triple zeta quality.
- The s and p contractions belonging to the same electron shell are folded into a sp-shell. In this case, number of s-type and p-type primitives is the same, and they have identical exponents. However, the coefficients for s- and p-type contractions are different.

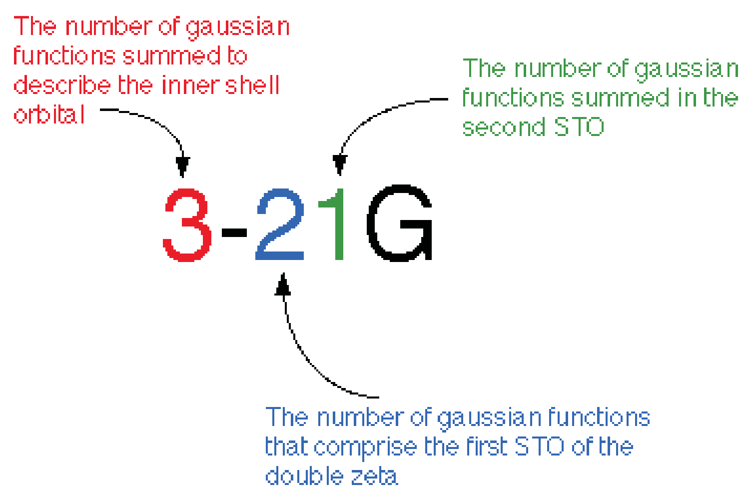


Figure 3: Pople's 3-21G basis set notation

Example: Hydrogen $1s$ orbital

Four s-type gaussians used to represent $1s$ orbital of hydrogen as:

$$\psi_{1s} = 0.50907 N_1 e^{-0.123317 r^2} + 0.47449 N_2 e^{-0.453757 r^2} \tag{7}$$

$$+ 0.13424 N_3 e^{-2.0133 r^2} + 0.01906 N_4 e^{-13.3615 r^2} \tag{8}$$

where N_i is a normalization constant for a given primitive. These primitives may be grouped in 2 contractions. The first contraction contains only 1 primitive:

$$\phi_1 = N_1 e^{-0.123317 r^2} \tag{9}$$

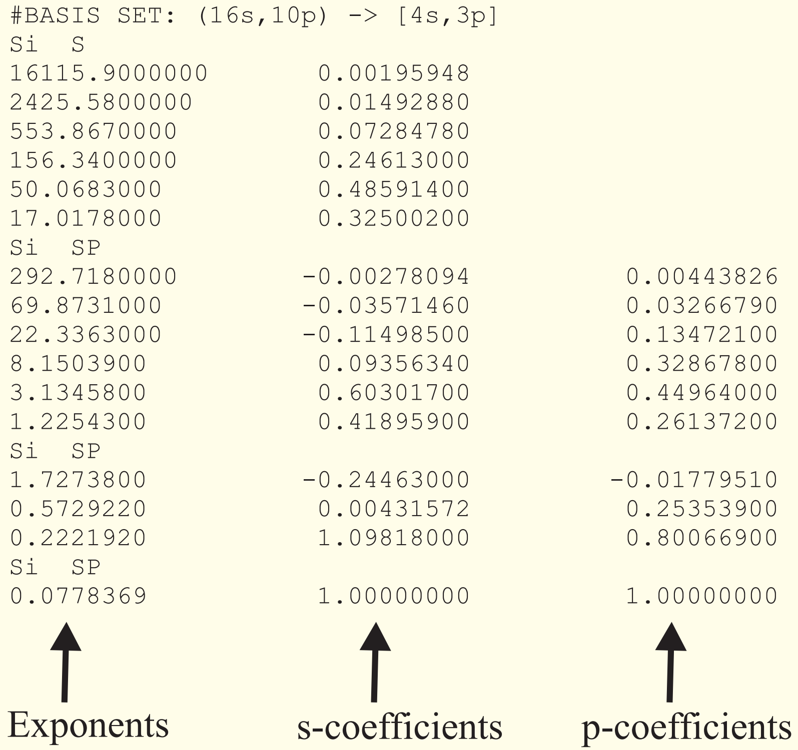
3 primitives are present in the second contraction:

$$\phi_2 = N[0.47449 N_2 e^{-0.453757 r^2} + 0.13424 N_3 e^{-2.0133 r^2} + 0.01906 N_4 e^{-13.3615 r^2}] \tag{10}$$

where N is a normalization constant for the whole contraction. In this case, 4 primitives were contracted to 2 basis functions. It is frequently denoted as (4s) → [2s] contraction. The coefficients in function are then fixed in subsequent molecular calculations.

Example: Silicon 6-31G basis set

- The corresponding exponents for s- and p-type contractions are equal but coefficients in s- and p-type contractions are different.
- Gaussian primitives are normalized here since coefficients for basis functions consisting of one primitive (last row) are exactly 1.
- The basis set above represents the following contraction $(16s,10p) \rightarrow [4s,3p]$



```
#BASIS SET: (16s,10p) -> [4s,3p]
Si   S
16115.9000000        0.00195948
2425.5800000         0.01492880
553.8670000          0.07284780
156.3400000          0.24613000
50.0683000           0.48591400
17.0178000           0.32500200
Si   SP
292.7180000         -0.00278094           0.00443826
69.8731000          -0.03571460           0.03266790
22.3363000          -0.11498500           0.13472100
8.1503900            0.09356340           0.32867800
3.1345800            0.60301700           0.44964000
1.2254300            0.41895900           0.26137200
Si   SP
1.7273800           -0.24463000          -0.01779510
0.5729220            0.00431572           0.25353900
0.2221920            1.09818000           0.80066900
Si   SP
0.0778369            1.00000000           1.00000000
```

Figure 4: 6-31G basis set for silicon.

Polarized basis sets

- Polarization functions denoted in Pople's sets by an asterisk
- Two asterisks, indicate that polarization functions are also added to light atoms (hydrogen and helium).
- Polarization functions have one additional node.
- For example, the only basis function located on a hydrogen atom in a minimal basis set would be a function approximating the $1s$ atomic orbital. When polarization is added to this basis set, a p-function is also added to the basis set. The 6-31G** is synonymous to 6-31G(d,p).

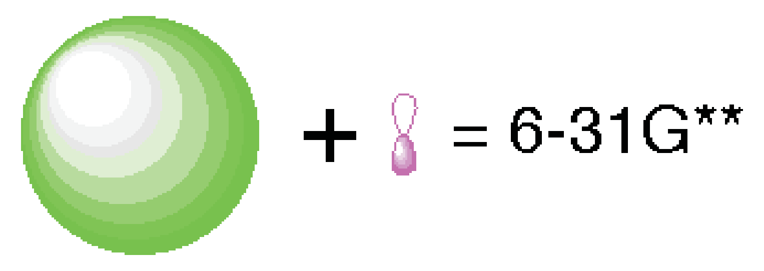


Figure 5: p-function is added to the $1s$ orbital

Polarized basis sets

- Polarization functions add flexibility within the basis set, effectively allowing molecular orbitals to be more asymmetric about the nucleus.
- This is an important for accurate description of bonding between atoms, because the precence of the other atom distorts the environment of the electrons and removes its spherical symmetry.

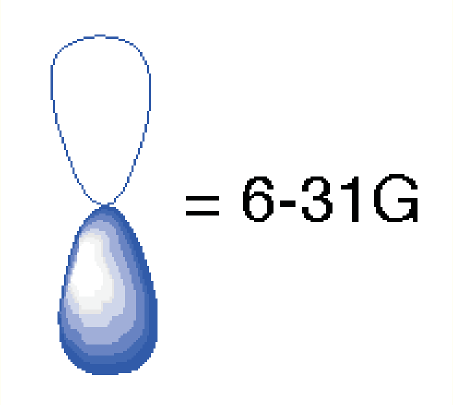


Figure 6: p orbital

Polarized basis sets

- Similarly, d-type functions can be added to a basis set with valence p orbitals, and so on.
- High angular momentum polarization functions (d,f,...) are important for heavy atoms

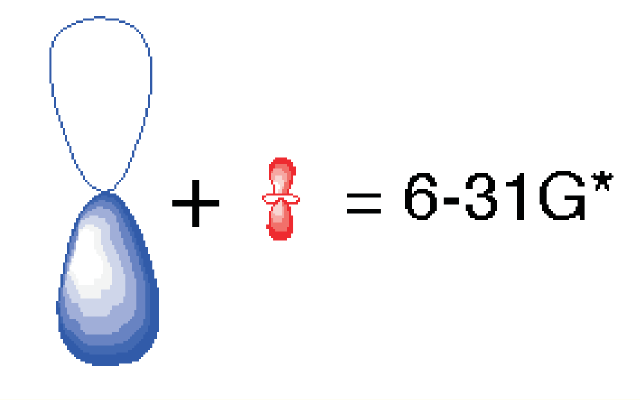


Figure 7: d-function is added to the p orbital

Polarized basis sets

Some observations concerning polarization functions:

- The exponents for polarization functions cannot be derived from Hartree-Fock calculations for the atom, since they are not populated.
- In practice, these exponents are estimated "using well established rules of thumb" or by using a test set of molecules.
- The polarization functions are important for reproducing chemical bonding.
- They should be included in all calculations where electron correlation is important.
- Adding them is costly. Augmenting basis set with d type polarization functions adds 5 basis function on each atom while adding f type functions adds 7.

Diffuse functions

The basis sets are also frequently augmented with the so-called diffuse functions.

- These gaussians have very small exponents and decay slowly with distance from the nucleus.
- Diffuse gaussians are usually of s and p type.
- Diffuse functions are necessary for correct description of anions and weak bonds (e.g. hydrogen bonds) and are frequently used for calculations of properties (e.g. dipole moments, polarizabilities, etc.).

For the Pople's basis sets the following notaton is used:

- n-ij+G, or n-ijk+G when 1 diffuse s-type and p-type gaussian with the same exponents are added to a standard basis set on heavy atoms.
- The n-ij++G, or n-ijk++G are obtained by adding 1 diffuse s-type and p-type gaussian on heavy atoms and 1 diffuse s-type gaussian on hydrogens.

Diffuse functions

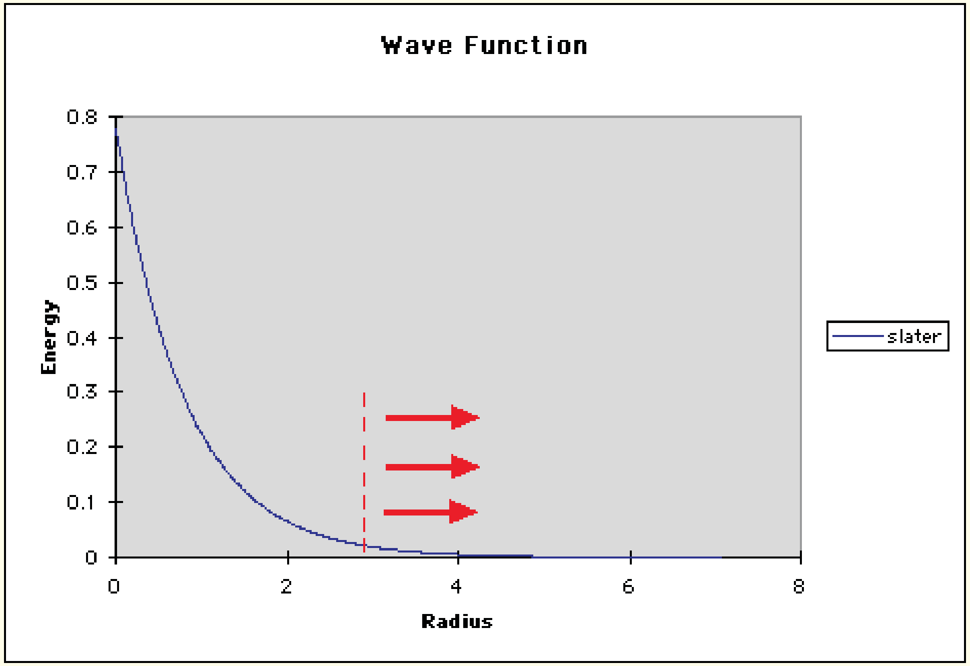


Figure 8: The area which is modelled by diffuse functions.

Diffuse functions are very shallow Gaussian basis functions, which more accurately represent the "tail" portion of the atomic orbitals, which are distant from the atomic nuclei.

Correlation consistent basis sets

Widely used basis sets are those developed by Dunning and co. These basis sets have become the current state of the art for correlated calculations

- Designed to converge systematically to the complete basis set (CBS) limit using extrapolation techniques
- For first- and second-row atoms, the basis sets are cc-pVNZ where N=D,T,Q,5,6,... (D=double-zeta, T=triple-zeta, etc.)
- The 'cc-p', stands for 'correlation consistent polarized' and the 'V' indicates they are valence only basis sets.
- They include successively larger shells of polarization (correlating) functions (d, f, g, etc.).
- The prefix 'aug' means that the basis is augmented with diffuse functions
- Examples: cc-pVTZ, aug-cc-pVDZ, aug-cc-pCV5Z

Correlation consistent basis sets

- Correlation consistent basis sets are built up by adding shells of functions to a core set of atomic Hartree-Fock functions. **Each function in a shell contributes very similar amounts of correlation energy in an atomic calculation.**

- For the 1st and 2nd row atoms, the cc-pVDZ basis set adds 1s, 1p, and 1d function. The cc-pVTZ set adds another s, p, d, and an f function, etc.

- For third-row atoms, additional functions are necessary; these are the cc-pV(N+d)Z basis sets.

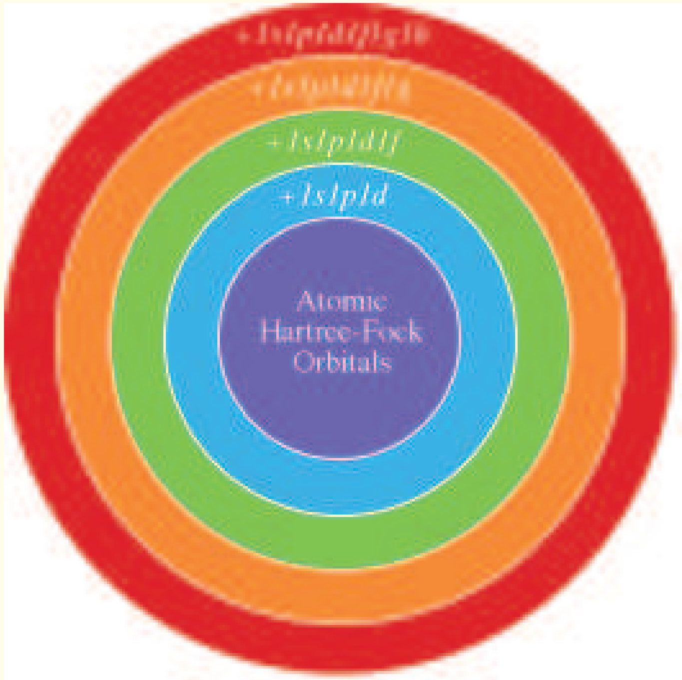


Figure 9: Adding shells of functions to a core set of atomic Hartree-Fock functions.

Basis set superposition error

Calculations of interaction energies are susceptible to basis set superposition error (BSSE) if they use finite basis sets.

- As the atoms of interacting molecules or two molecules approach one another, their basis functions overlap. Each monomer "borrows" functions from other nearby components, effectively increasing its basis set and improving the calculation.

- The counterpoise approach (CP) calculates the BSSE employing "ghost orbitals". In the uncorrected calculation of a dimer AB, the dimer basis set is the union of the two monomer basis sets. The uncorrected interaction energy is

  $$V_{AB}(G) = E_{AB}(G, AB) - E_A(A) - E_B(B) \tag{11}$$

  where G denotes the coordinates that specify the geometry of the dimer and $E_{AB}(G, AB)$ the total energy of the dimer AB calculated with the full basis set AB of the dimer at that geometry. Similarly, $E_A(A)$ and $E_B(B)$ denote the total energies of the monomers A and B, each calculated with the appropriate monomer basis sets A and B, respectively.

Basis set superposition error

The counterpoise corrected interaction energy is

$$V_{AB}^{cc}(G) = E_{AB}(G, AB) - E_A(G, AB) - E_B(G, AB) \qquad (12)$$

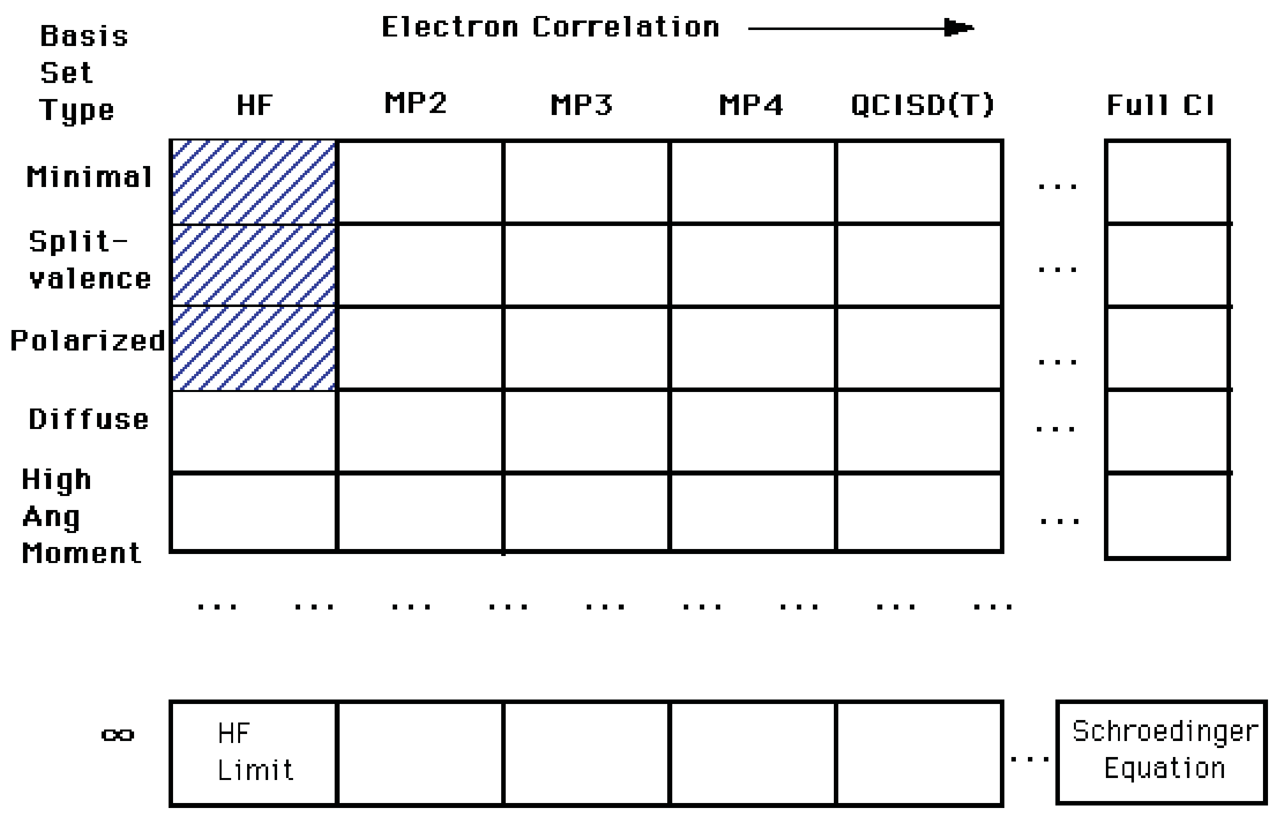
where $E_A(G, AB)$ and $E_B(G, AB)$ denote the total energies of monomers A and B, respectively, computed with the dimer basis set at geometry G, i.e. in the calculation of monomer A the basis set of the "other" monomer B is present at the same location as in dimer A, but the nuclei of B are not. In this way, the basis set for each monomer is extended by the functions of the other monomer.

The counterpoise corrected energy is thus

$$E_{AB}^{cc} = E_{AB}(G, AB) + V_{AB}^{cc}(G) - V_{AB}(G) \qquad (13)$$

- The counterpoise correction provides only an estimate of the BSSE.
- BSSE is present in all molecular calculations involving finite basis sets but in practice its effect is important in calculations involving weakly bound complexes. Usually its magnitude is few kJ/mol to binding energies which is often very significant.

Final picture



Electron Correlation →

Basis Set Type	HF	MP2	MP3	MP4	QCISD(T)		Full CI
Minimal	▨					…	
Split-valence	▨					…	
Polarized	▨					…	
Diffuse						…	
High Ang Moment						…	
…	…	…	…	…	…	…	
∞	HF Limit					…	Schroedinger Equation

Figure 10: Towards exact solution of Scrödinger equation