Basis Sets Used in Molecular Orbital Calculations

1. Introduction

Each molecular orbital (one electron function) ψ_i is expressed as a linear combination of n basis functions Φ_μ. By convention, molecular orbitals are abbreviated by the greek letter ψ (psi), while the basis functions are symbolized by the greek letter Φ (phi).

$$\psi_i = \sum_{\mu=1}^{n} c_{\mu i} \Phi_\mu \qquad (2)$$

The coefficients $c_{\mu i}$ are called *molecular orbital expansion coefficients* or simply *MO coefficients*. Usually these basis functions are located at the center of atoms and are therefore often called atomic basis functions. This is, however, not strictly true and in order to avoid confusion, we will in the following use the term *basis function*.

The basis functions used in molecular orbital calculations are usually described through an abbreviation or acronym such as "6-31G(d)". The complete description of the theoretical model used in MO calculations is given in the format **"theoretical method/basis set"**. A calculation using Hartree-Fock theory in combination with the 6-31G(d) basis set would then be referred to as "HF/6-31G(d)". Often the molecular geometries are optimized at one level of theory (method 1) and relative energies are then recomputed using these geometries, but a different theoretical method (method 2). The conventional description of this level of theory is "**method 2//method 1**". If, for example, geometry optimization has been performed at the HF/6-31G(d) level of theory and relative energies have then been calculated at the MP2/6-31G(d) level, this approach would be described as "MP2/6-31G(d)//HF/6-31G(d)".

In principle, any mathematical function can be used to construct molecular orbitals. There are, however, only a few solutions, which have found widespread application.
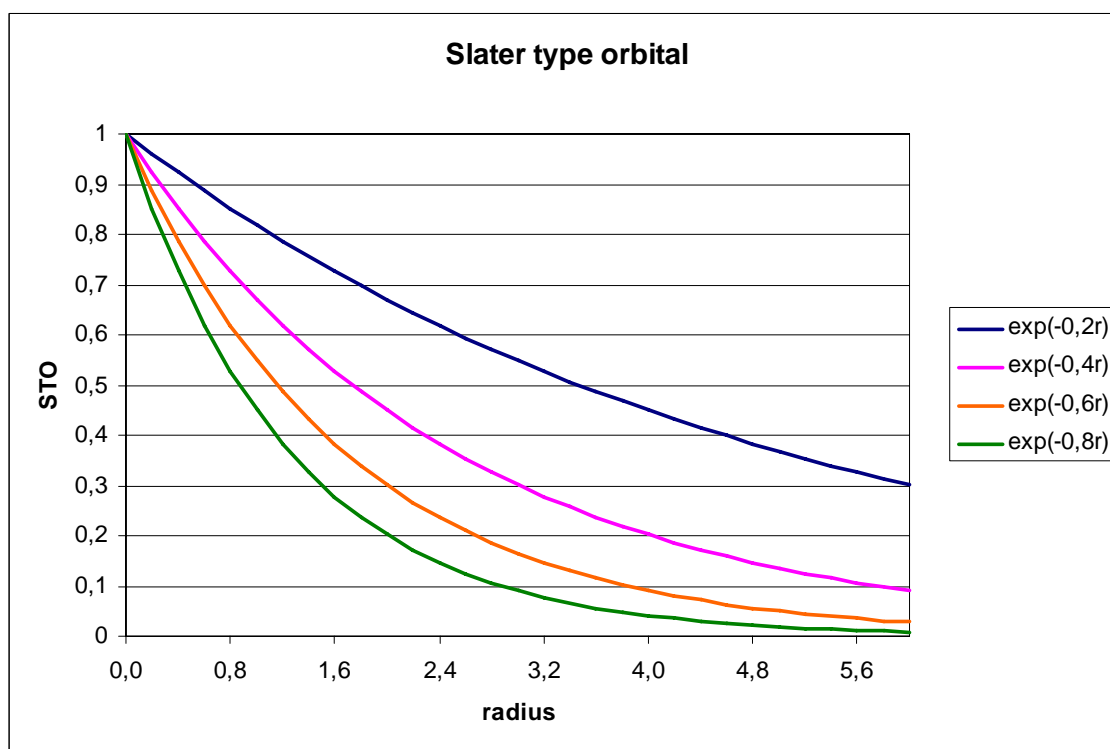
1.1 Slater type orbitals (STOs)

STOs are constructed from a radial part describing the radial extend of the orbital and an angular part describing the shape of the orbital (equ. 3).

$$\Phi_\mu = N \bullet r^{n-1} \exp(-\zeta r) \bullet Y_{lm} \qquad (3)$$

The **radial part $r^{n-1} \exp(-\zeta r)$** depends on the distance r from the origin of the basis function (usually the location of the nucleus), the orbital exponent ζ (greek: zeta), and the principal quantum number n. The **spherical part Y_{lm}** depends on the angular

quantum number *l* and the magnetic quantum number *m*. The normalization constant *N* is chosen such that the integral over the square of the basis function yields unity. Taking a 1s orbital as an example, the spherical part as well as the r^{n-1}-factor equate to unity and after evaluating the normalization constant, we get:

$$\Phi_{1s} = \left(\frac{\zeta_{1s}^3}{\pi}\right) \exp(-\zeta_{1s}r) \qquad (4)$$

Slater type orbital

Slater type orbitals are not very useful for electronic structure calculations, because two electron integrals over STOs are difficult to calculate.

1.2 Gaussian type orbitals (GTOs)

GTOs are also constructed from a radial and a spherical part, but the radial part now has a different dependence on *r* (equ. 5):

$$g = N \exp(-\alpha r^2) \, x^a \bullet y^b \bullet z^c \qquad (5)$$

The radial extend is proportional to ***exp(-αr²)***, α being the exponent. The normalization constant *N* serves a similar purpose as for STOs. The spherical part is now expressed through the cartesian coordinates *x*,*y*, and *z* in powers of *a*, *b*, and *c*, respectively. The exponents ***a, b, and c, should not be confused with quantum numbers***. The sum of these exponents L=a+b+c is used to define the angular momentum of the basis

functions: **s-type (L=0), p-type (L=1), d-type (L=2), f-type (L=3), g-type (L=4)** . . ..
For a simple Gaussian function representing an 1s-type orbital, we get:

$$g_s = \left(\frac{2\alpha_s}{\pi}\right)^{3/4} \exp(-\alpha r^2) \tag{6}$$

Gaussian type orbital
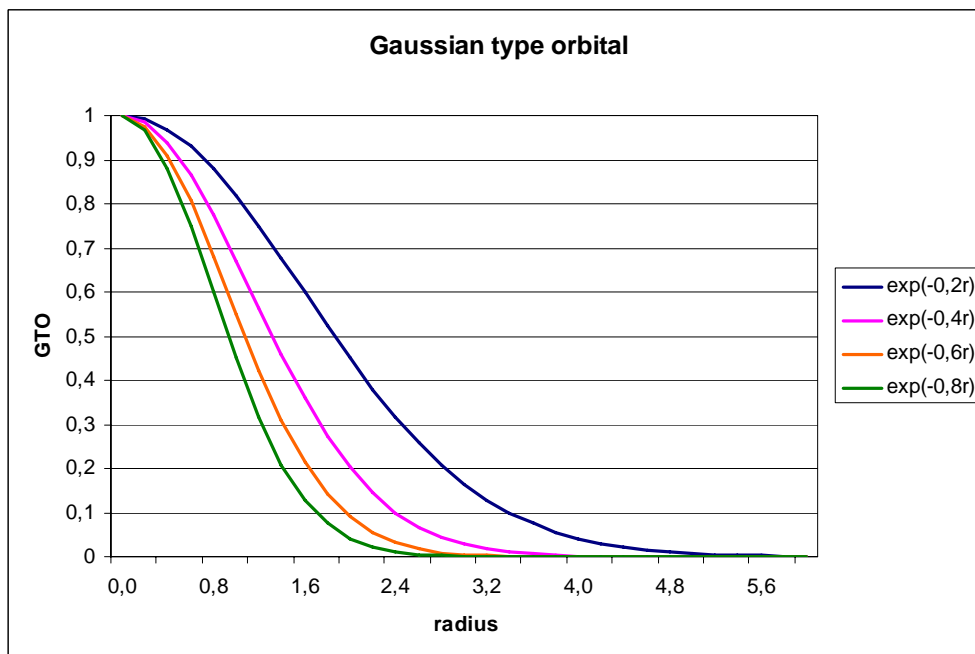
Gaussian type orbitals are very useful for electronic structure calculations, because two electron integrals over GTOs are NOT difficult to calculate. GTOs have the advantage that all the intergrals needed for the LCAO calculations can be done by using simple formulae. (S. F. Boys, *Proceedings of the Royal Society of London Series A*, **1950**, *200*, 542-554.)

Finally, a comparison between STOs and GTOs: There are two major differences (i) the behaviour for r = 0 and for (ii) r = ∞. The STO displays a cusp at r = 0 while the GTO has an incorrect behaviour at the nucleus. Moreover GTOs fall off too quickly as the electron gets further from the nucleus.

GTO vs STO
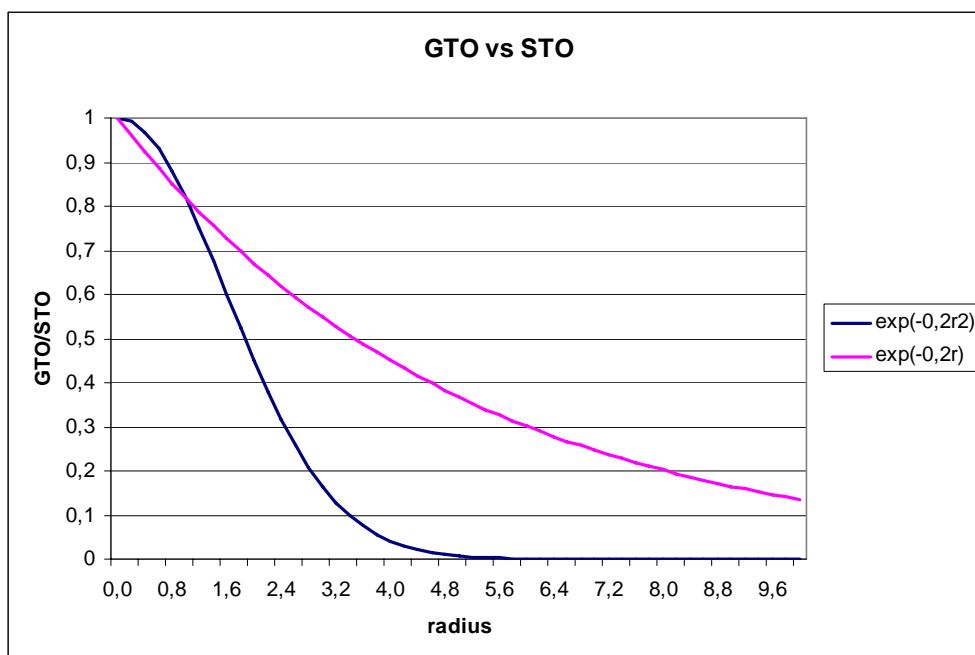
1.3 Contracted Gaussian type orbitals

1.3.1 Minimal basis sets

Single Gaussian functions as described by equ. (5) are not well suited to describe the spatial extend and nodal characteristics of atomic orbitals. To solve this problem, basis functions are described as a sum (a "**contraction**") of several Gaussian functions (primitives):

$$\Phi_\mu = \sum_s d_{\mu s}\, g_s \qquad (7)$$

The Carbon 1s Orbital (STO-3G)
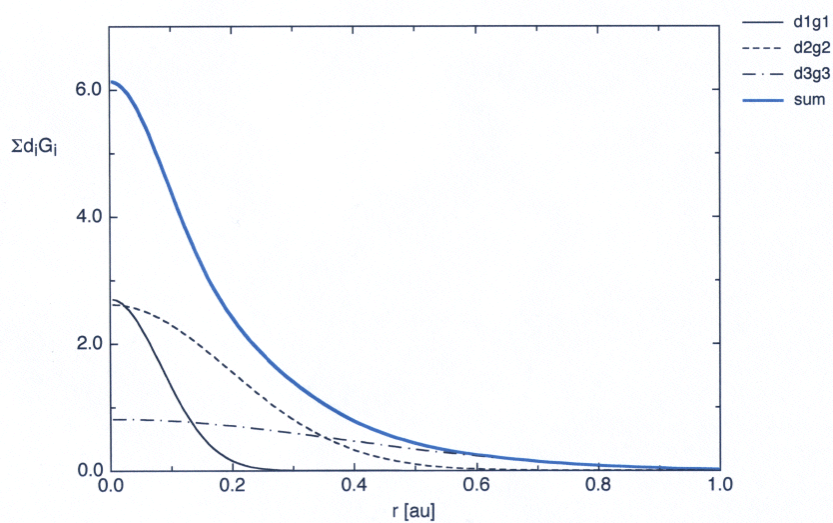
We will use the STO-3G basis set[1] for carbon as an example (here in a format produced by Gaussian 94):

C 0
```
 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
```

The format used is:

Shell_type, No. of primitive Gaussians x, Scale_factor

Orbital exponent α_x, Contraction coefficient d_x

(repeat x times)

The first line specifies that the carbon 1s orbital will be described as a sum of three Gaussian primitives with different exponents $\alpha_{1s,x}$ and coefficients $d_{1s,x}$. The following three lines contain the exponent $\alpha_{1s,x}$ and coefficient $d_{1s,x}$ for x=1, 2, and 3. Written in an explicit manner, the 1s basis function for carbon then equates to:

$$\Phi_{1s} = \quad d_{1s,1}\left(\frac{2\alpha_{1s,1}}{\pi}\right)\exp(-\alpha_{1s,1}\,r^2) \qquad (8)$$

$$+ d_{1s,2}\left(\frac{2\alpha_{1s,2}}{\pi}\right)\exp(-\alpha_{1s,2}\,r^2)$$

$$+ d_{1s,3}\left(\frac{2\alpha_{1s,3}}{\pi}\right)\exp(-\alpha_{1s,3}\,r^2)$$

The basis sets for the 2s and 2p orbitals are also composed of a contraction of three Gaussian primitives and are specified together in one single "SP"-block. For the sake of computational efficiency, the **2s and 2p orbitals share the same exponents $\alpha_{2,x}$, but have different contraction coefficients**. Each line of the basis set description therefore contains only one exponent ($\alpha_{2s,x} = \alpha_{2p,x}$), but two different coefficients $d_{2s,x}$ and $d_{2p,x}$. The nodal characteristics of the 2s-orbital are simulated using one coefficient with negative sign (-0.09996722919) and two coefficients with positive (+0.3995128261 and +0.7001154689) sign. Orbital exponents and coefficients have been optimized first for atomic calculations minimizing the energy of single atoms. Additional scale factors have, however, then been developed for use of these atomic basis functions in studies of small organic molecules. The STO-3G basis set thus contains an appreciable amount of "semiempirical" character!

How many basis functions do we need for a small organic molecule such as methanol (CH_4O)? For each of the atomic 1s, 2s and 2p-orbitals, the STO-3G basis set

uses a contraction of three Gaussians. We thus have 5 basis functions for each carbon and oxygen atom, and only one basis function for each hydrogen, yielding a total of 14 basis functions (and 14 MO-coefficients to vary in the SCF calculations). Each of these basis functions consists of three Gaussian primitives, and the number of primitives is therefore 3*14=42. In the output file of *Gaussian 98* the number of basis functions and primitives is given by link 301 together with the number of electrons and the core-core repulsion energy:

```
14 basis functions     42 primitive gaussians
 9 alpha electrons      9 beta electrons
   nuclear repulsion energy     40.7431450799 Hartrees.
```

How "good" is the STO-3G basis set? The answer to this question depends on the chemical problem at hand. Structural features of ground state molecules are reproduced quite often with surprisingly good accuracy at the HF/STO-3G level of theory. Table 1 lists structural parameters of methanol as predicted by a variety of theoretical methods.

Table 1. Structural features of methanol (C_s) as predicted by several theoretical methods.

Method	C-O [Å]	C-O-H[°]	bf[a]
HF/STO-3G	1.432999	103.856	14
HF/3-21G	1.440922	110.336	26
HF/3-21+G	1.454241	113.291	34
HF/6-31G	1.430503	113.410	26
HF/D95	1.437049	113.874	28
HF/6-31G(d)	1.399562	109.448	38
HF/6-31G(d,p)	1.398520	109.650	50
HF/6-31++G(d,p)	1.401386	110.542	62
HF/6-311G(d,p)	1.399240	109.355	60
HF/6-311++G(d,p)	1.400243	110.017	72
HF/6-311G(2df,2pd)	1.396113	109.836	116
HF/cc-pVDZ	1.397818	109.116	48
HF/cc-pVTZ	1.398133	109.937	116
MP2(fc)/cc-pVDZ	1.417284	106.283	48
MP2(fc)/cc-pVTZ	1.418481	107.431	116
exp.	1.421	108.0	

[a] number of basis functions

1.3.2 Split-valence basis sets I: the 3-21G basis set[2-4]

The description of the valence electrons can be significantly improved over that in the minimal STO-3G basis set, **if more than one basis function is used per valence electron**. Basis sets of this type are called "**split valence**" basis sets as the description

of valence orbitals are split into two (or more) basis functions. A related term is "**double zeta**" in reminiscence of the greek symbol used for the orbital exponents of STOs. The term "double zeta" (DZ) does not imply, however, whether two basis sets are use for all of the orbitals or only for the valence space.

One very economical, small split valence basis set is the **3-21G** basis set.[2-4] The non-valence electrons are described by single basis functions composed of a contraction of three Gaussians. **Each valence electron is described by two basis functions**. The first of these basis functions is composed of two Gaussian primitives while the second consists of a single uncontracted Gaussian primitive.

For carbon the 3-21G basis set is[2] (Gaussian 94 format):

```
C 0
 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
```

The main difference to the STO-3G listing before is that there are two sections now for the 2s and 2p orbitals. Again, s and p orbitals share the same exponents $\alpha_{2,x}$ in the first column, but have different contraction coefficients listed in the second and third column. The contraction coefficient of the outer SP shell is, of course, unity as the outer SP shell consists of only one Gaussian function. Inspection of the orbital exponents of the first SP shell (3.66 and 0.77) and the second SP shell (0.19) shows that the first SP shell has larger exponents and therefore describes electron density closer to the nucleus as compared to the second SP shell. *Consequently, the first and second SP shells are sometimes referred to as the "inner" and "outer" shells, respectively*.

The 3-21G basis set has somewhat of a curious developmental history. Straightforward variation of the orbital exponents and expansion coefficients as practiced in the development of other basis sets led to an "falling inward" of the inner valence shells. This is due to the fact that there are only three Gaussians available for the description of the core region and that addition of more primitives to the core region lowers the overall energy more than an adequate description of the valence shell. For most elements the basis set parameters were therefore first optimized using a much larger core space of six Gaussians (6-21G). At this stage all basis set parameters were varied to minimize the energy of the atoms in their electronic ground states at the unrestricted Hartree-Fock (UHF) level of theory. The large core was then replaced by a smaller one of only three Gaussians and the basis set parameters of the new core region

were reoptimized, keeping the valence region constant. A similar strategy has also been used for second row elements.[3,4]

How much more computational effort is required for the 3-21G basis set as compared to the minimal STO-3G solution? Each carbon and oxygen atom in methanol now needs 9 and each hydrogen atom needs 2 basis functions. For methanol this equates to 26 instead of 14 basis functions. This much larger number of basis functions and thus variable MO coefficients is achieved, however, with the same number of Gaussian primitives as for the STO-3G basis set: 42.

1.3.3 Split-valence basis sets II: the 6-31G basis sets

Development of the larger 4-31G and 6-31G split valence basis sets[5,6] predates that of the 3-21G basis set considerably. The main difference between 3-21G and 6-31G is that a much larger number of primitives is used in the latter in the core as well as the inner most valence shell. **The use of a contraction of six Gaussian primitives for each core orbital improves the description of the core region significantly**. The valence region is again described by two basis functions per atomic orbital. The inner shell is composed of a contraction of three Gaussians and the outer shell consists of one single Gaussian primitive. As in other basis sets developed by the **Pople group**, s and p functions share common exponents.

The 6-31G basis for carbon is:

```
C 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
```

The orbital exponents and expansion coefficients were optimized to yield the lowest possible UHF energies for the respective atoms in their electronic ground states. The exponents of the valence shell atoms have then been scaled uniformly using scale factors developed for the 4-31G basis set in order to achieve the best possible results in MO calculations on a set of small organic molecules.

For methanol, the 6-31G basis set includes 26 basis functions, which are composed of a total of 60 Gaussian primitives.

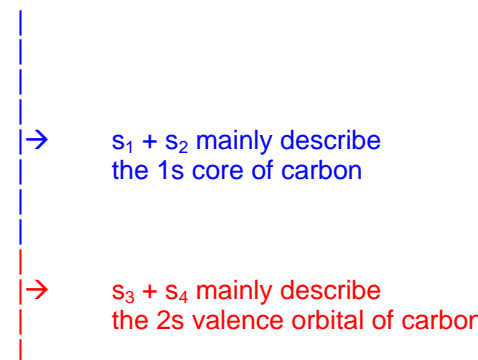### 1.3.4 Double zeta basis sets: Dunnings D95 basis set[7,8]

Dunnings D95 basis set has been derived from an already existing large atomic basis set of nine uncontracted Gaussian primitives of s- and five uncontracted Gaussian primitives of p-type.[8] Six of the nine s-type functions have then been grouped into a single contraction, while the other three s-type functions have been left alone. Similarly, four of the five p-type functions have been contracted into a single function, while one function was left uncontracted. Overall, this yields **a basis set of four s-type and two p-type basis function**s. In contrast to the split valence basis sets discussed before, the D95 basis set is a **full double zeta basis set** in that it allocates two basis functions for each atomic orbital of the core as well as the valence region occupied in the electronic ground state.

The D95 basis set for carbon is:

```
C 0
 S   6 1.00
  .4232610000D+04   .2029000000D-02          |
  .6348820000D+03   .1553500000D-01          |
  .1460970000D+03   .7541100000D-01          |
  .4249740000D+02   .2571210000D+00          |
  .1418920000D+02   .5965550000D+00          |→     s₁ + s₂ mainly describe
  .1966600000D+01   .2425170000D+00          |      the 1s core of carbon
 S   1 1.00                                  |
  .5147700000D+01   .1000000000D+01          |
 S   1 1.00                                  |
  .4962000000D+00   .1000000000D+01          |→     s₃ + s₄ mainly describe
 S   1 1.00                                  |      the 2s valence orbital of carbon
  .1533000000D+00   .1000000000D+01          |
 P   4 1.00
  .1815570000D+02   .1853400000D-01
  .3986400000D+01   .1154420000D+00
  .1142900000D+01   .3862060000D+00
  .3594000000D+00   .6400890000D+00
 P   1 1.00
  .1146000000D+00   .1000000000D+01
```

The format used here for the listing of orbital exponents and expansion coefficients is different from those used before for the Pople basis sets as s- and p-type functions of the D95 basis set do not share the same orbital exponents.

Using an uncontracted atomic basis set as the starting point for the development of contracted versions suitable for the treatment of larger systems is common practice. The standard nomenclature used specifies the uncontracted basis set in brackets and the resulting contracted version in square brackets. Using the D95 basis set as an example, the contraction can be described as **(9s5p) -> [4s,2p]**. This notation does not specify, how many primitives are contained in each contraction. This can be specified in more detail as **(6111,41)**, listing first the s-type functions (here distributed over four

contractions) and then the p-type functions. Basis sets, in which a given primitive appears in only one of the contractions, are termed *segmented*.

Why are contractions done? MO calculations using the uncontracted 9s5p atomic basis set would need to handle 9*1+5*3=24 MO coefficients for each carbon atom while only 4*1+2*3=10 MO coefficients are necessary for the [4s,2p] contraction. If proper care is taken during the contraction process, **calculations using the contracted basis sets can be performed with similar accuracy but dramatically reduced computational cost**.

For methanol, the D95 basis set uses 10 basis functions for each carbon and oxygen atom, and two functions for hydrogen, yielding a total of 28 basis functions. These basis functions are constructed from a total of 64 Gaussian primitives.

1.4 Polarization functions for double zeta basis sets

To further increase the flexibility of the orbital description, double zeta basis sets are frequently **augmented with basis functions of higher angular momentum**. A typical first step consists of the addition of a **set of d-type functions** to the basis sets of those atoms, which have occupied s- and p-shells in their electronic ground states. For hydrogen, this corresponds to the addition of a set of p-type functions. Two different notations exist to specify the addition of polarization functions. The first notation adds one asterisk to the basis set to specify addition of polarization functions to non-hydrogen atoms, while two asterisks symbolize the addition of polarization functions to all atoms (including hydrogen). The **6-31G**** basis set[10] is thus constructed from the split valence 6-31G basis set through **addition of one set of d-functions to all non-hydrogen atoms and one set of p-functions to all hydrogen atoms**. In the second (preferable) notation the polarization functions are specified through their angular quantum number explicitly. The 6-31G** basis set would then be termed "**6-31G(d,p)**". This latter notation is much more flexible as multiple sets of polarization functions can be specified much more easily.

Further characteristics of polarization functions can be discussed after inspection of an example. The 6-31G(d) basis set for carbon is:

```
C 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
```

```
 D   1 1.00
   .8000000000D+00   .1000000000D+01
```

The polarization functions for the 6-31G basis appear in the basis set listing as a single set of uncontracted d-type Gaussians. There are **six cartesian d-type Gaussians** (x^2, y^2, z^2, xy, yz, zx) exp ($-\alpha_d r^2$), which **are equivalent to five pure d-type functions** (xy, yz, zx, x^2-y^2, $3z^2$-r^2) exp ($-\alpha_d r^2$) plus one additional s-type function. In some basis sets (such as 6-31G(d)) calculations are performed with all six polarization functions, in other cases only the pure d-type functions are used. Unfortunately, this is not reflected in the notation currently used for polarization functions. Great care must therefore be taken when theoretical results using polarized basis sets are compared.

For the elements C, N, O, and F the orbital exponents of the d-type functions in the 6-31G(d) basis set have uniformly been set to a value of **0.8**.[10] The rational for this somewhat arbitrary assignment was that optimized orbital exponents of 0.73-0.92 were calculated for a small set of organic molecules and that the use of an uniform average value of 0.8 would be advantageous. For hydrogen, a standard orbital exponent of α_p=**1.1** has been adopted for the p-type Gaussian polarization functions of the 6-31G(d,p) basis set. The full 6-31G(d,p) basis set for hydrogen is:

```
H 0
 S   3 1.00
   .1873113696D+02   .3349460434D-01
   .2825394365D+01   .2347269535D+00
   .6401216923D+00   .8137573262D+00
 S   1 1.00
   .1612777588D+00   .1000000000D+01
 P   1 1.00
   .1100000000D+01   .1000000000D+01
```

As there is **no inner core** the 1s orbitals are described by two basis functions. The inner basis function is composed of three s-type Gaussians and the outer basis function is a single uncontracted s-type Gaussian with comparatively small exponent. The last two lines specify a single set of uncontracted p-type Gaussians with exponent 1.1.

Returning to the example of methanol, one can readily see that the addition of polarization functions increases the number of basis functions and primitives significantly: The number of basis functions is: 26 [6-31G], 38 [6-31G(d)], and 50 [6-31G(d,p)] and the number of primitives is: 64 [6-31G], 72 [6-31G(d)], 84 [6-31G(d,p)]. Calculations involving polarized basis sets are therefore much more time consuming than those using the corresponding unpolarized double zeta basis sets. The results in Table 1 show, however, that already the addition of a single set of polarization functions on non-hydrogen atoms goes a long way towards the results obtained with much more sophisticated basis sets. *The use of polarized basis sets is especially important for the proper description of bonds of strongly electronegative elements such as oxygen and fluorine and for theoretical studies using correlated methods*.[11]

1.5 Diffuse basis functions for split valence basis sets

  The theoretical description of negatively charged species is particularly challenging for *ab initio* MO theory. This is due to the fact that the excess negative charge spreads outward to a much larger degree than is typically the case for uncharged or positively charged molecules. The **description of such a diffuse charge distribution** is not very well possible with the typical split valence basis sets discussed before. Addition of very diffuse basis functions (with correspondingly **small orbital exponents**) cures this problem to a certain extend as it allows the description of electron density relatively far from the nucleus.[12,13] Diffuse basis functions are typically added as an additional set of uncontracted Gaussian functions of the same angular momentum as the valence electrons. To reflect the addition of diffuse basis functions on all non-hydrogen atoms, a +-sign is added to the standard basis set notation. If diffuse s-type functions are also added to the basis set of hydrogen atoms, a second +-sign is appended. Using carbon as an example, the combination of the 3-21G basis set with one set of diffuse sp-functions yields the improved "**3-21+G**" basis set:

```
C 0
 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
 SP  1 1.00
  .4380000000D-01   .1000000000D+01   .1000000000D+01
```

It can clearly be seen that the orbital exponent of the diffuse SP shell is significantly smaller than that of the outer valence shell (**0.0438** vs. 0.1959). The same diffuse functions developed for the small 3-21G basis set are also used with larger double and triple zeta basis sets. Recommended exponents[13] for the first row atoms and hydrogen are: 0.0360 (H), 0.0074 (Li), 0.0207 (Be), 0.0315 (B), 0.0438 (C), 0.0639 (N), 0.0845 (O), and 0.1076 (F). These values were obtained through minimizing the energy for a small set of hydrides at the UHF/6-21+G level of theory.

  A word on theoretical studies of negatively charged species appears in place at this point. *Numerous negatively charged species do not represent stable species in the gas phase with respect to spontaneous loss of an electron*. That MO calculations can be done at all on these species is simply due to the use of a limited basis set. It is therefore suggested to always verify in calculations on negatively charged species that the energy of the anionic system is lower than the energy of the neutral system obtained after

vertical neutralization. This does not guarantee that the negatively charged system represents a bound state, but certainly makes this more likely. A point in case is the formaldehyde radical anion. This molecule can be optimized at the UHF/3-21G level of theory and gives a result which is in line with expectation for ketyl radical anions: the C-O bond is longer than in formaldehyde (1.319Å vs. 1.207Å), most of the negative charge is located at the oxygen atom and most of the spin density is located on carbon. Comparison of the total energy of the radical anion (-113.132324 hartree) with that of neutral formaldehyde after vertical neutralization (-113.203973 hartree) shows, however, that the formaldehyde radical anion does not represent a bound species in the gas phase. The same result is obtained at other levels of theory.

The use of diffuse basis functions in calculations of neutrals has usually little consequences. As can be seen in Table I, the C-O bond length becomes slightly longer on inclusion of diffuse basis functions. This can easily be understood as a consequence of removal of some electron density from the internuclear region further away from the molecules. The addition of diffuse functions does, however, significantly increase the computational effort. For methanol, we have 26 basis functions and 42 primitives for the 3-21G, but 34 basis functions and 50 primitives for the 3-21+G basis set. The computational cost is further increased as SCF convergence is more difficult to achieve with diffuse basis functions.

1.6 Basis sets for correlated methods

All of the basis sets discussed up to now have been developed for use at the Hartree-Fock level of theory, even though corrections have been included in some cases to better reproduce experimental data. Basis sets specifically designed **for use with correlated methods** have been developed more recently. These basis sets are usually at least of **double zeta quality and also include polarization functions**. One typical example is the 6-311G(d,p) basis set.[14] This basis set consists of a core region of six contracted Gaussian primitives and a triply split valence basis set using, respectively, three, one, and one Gaussian primitives. Polarization functions on first-row elements consist of one set of *five* pure d-functions, while a set of three p-type orbitals are added on hydrogen. On total this yields 18 basis functions for each non-hydrogen atom of the first row and 6 basis functions for hydrogen. The basis sets for carbon and hydrogen are:

```
C 0
S   6 1.00
 .4563240000D+04  .1966650000D-02
 .6820240000D+03  .1523060000D-01
 .1549730000D+03  .7612690000D-01
 .4445530000D+02  .2608010000D+00
 .1302900000D+02  .6164620000D+00
 .1827730000D+01  .2210060000D+00
SP  3 1.00
 .2096420000D+02  .1146600000D+00  .4024870000D-01
 .4803310000D+01  .9199990000D+00  .2375940000D+00
 .1459330000D+01 -.3030680000D-02  .8158540000D+00
SP  1 1.00
 .4834560000D+00  .1000000000D+01  .1000000000D+01
SP  1 1.00
 .1455850000D+00  .1000000000D+01  .1000000000D+01
D   1 1.00
 .6260000000D+00  .1000000000D+01

 H 0
S   3 1.00
 .3386500000D+02  .2549380000D-01
 .5094790000D+01  .1903730000D+00
 .1158790000D+01  .8521610000D+00
S   1 1.00
 .3258400000D+00  .1000000000D+01
S   1 1.00
 .1027410000D+00  .1000000000D+01
P   1 1.00
 .7500000000D+00  .1000000000D+01
```

Orbital exponents and expansion coefficients of the 6-311G(d,p) basis set have been chosen such that the ground state UMP2(FC) atomic energies were minimized. The frozen core (FC) calculation eliminates the non-valence electrons in the inner core from the electron correlation calculations. Unlike the small 6-31G(d) basis set, the 6-311G(d,p) basis set uses different exponents for the polarization functions on different nuclei.

For methanol the 6-311G(d,p) basis set consists of 60 basis functions which are in turn composed of 94 primitive Gaussians.

1.7 correlation consistent basis sets

A series of basis sets for correlated calculations has also been developed by Dunning et al.[15] These basis sets are referred to as *correlation consistent* (or *cc*) and are designed such that a base set of sp functions is combined with correlation functions. These latter functions are chosen such that all functions in a given set lead to a similar lowering of the atomic correlation energy (calculated by CISD). The smallest member of this series and thus often the starting point for correlated calculations is the correlation consistent polarized double zeta basis set designated "**cc-pVDZ**".[15] The cc-pVDZ basis for oxygen is:

```
 O 0
 S   9 1.00
  .1172000000D+05  .7100000000D-03
  .1759000000D+04  .5470000000D-02
  .4008000000D+03  .2783700000D-01
  .1137000000D+03  .1048000000D+00
  .3703000000D+02  .2830620000D+00
  .1327000000D+02  .4487190000D+00
  .5025000000D+01  .2709520000D+00
  .1013000000D+01  .1545800000D-01
  .3023000000D+00 -.2585000000D-02
 S   9 1.00
  .1172000000D+05 -.1600000000D-03
  .1759000000D+04 -.1263000000D-02
  .4008000000D+03 -.6267000000D-02
  .1137000000D+03 -.2571600000D-01
  .3703000000D+02 -.7092400000D-01
  .1327000000D+02 -.1654110000D+00
  .5025000000D+01 -.1169550000D+00
  .1013000000D+01  .5573680000D+00
  .3023000000D+00  .5727590000D+00
 S   1 1.00
  .3023000000D+00  .1000000000D+01        „generic contractions"
 P   4 1.00
  .1770000000D+02  .4301800000D-01
  .3854000000D+01  .2289130000D+00
  .1046000000D+01  .5087280000D+00
  .2753000000D+00  .4605310000D+00
 P   1 1.00
  .2753000000D+00  .1000000000D+01
 D   1 1.00
  .1185000000D+01  .1000000000D+01
```

This smallest member of the *cc*-family is a [3s,2p,1d] contraction of a (9s,4p,1d) basis set. A (1s,1p,1d) subset of this basis set has been optimized such that the energy for the oxygen atom is minimized in a CISD calculation.

The next member of the family is obtained by starting from a somewhat larger (10s,5p) atomic basis set obtained at the Hartree-Fock level and adding 2d and 1f polarization functions to obtain the final [4s,3p,2d,1f] contraction. Addition of the two d-type polarization function yield in this case the same amount of correlation energy as addition of a single f function. The higher members of the cc-family of basis sets are obtained in a similar manner:

basis set	contraction	functions for O
cc-pVDZ	[3s2p1d]	14
cc-pVTZ	[4s3p2d1f]	30
cc-pVQZ	[5s4p3d2f1g]	55
cc-pV5Z	[6s5p4d3f2g1h]	91
cc-pV6Z	[7s,6p,5d,4f,3g,2h,1i]	140

An extension of the *cc*-basis sets by more diffuse functions is mandatory for charged systems. A systematic extension is represented by the aug-cc-pVxZ series which derives from the cc-pVxZ series through addition of one set of diffuse functions for each angular momentum type already present in the respective cc-basis. The cc-pVDZ thus adds a diffuse (1s1p1d) set to build the final [4s3p2d] aug-cc-pVDZ set. The complete family of **aug-cc-pVXZ** basis set is then:

basis set	contraction	functions for O
aug-cc-pVDZ	[4s3p2d]	23
aug-cc-pVTZ	[5s4p3d2f]	46
aug-cc-pVQZ	[6s5p4d3f2g]	80
aug-cc-pV5Z	[7s6p5d4f3g2h]	127

The exponents of the most diffuse basis functions in the aug-cc-pVXZ basis sets have been optimized through HF and CISD calculations **on the negatively charged atoms**. *These basis sets are generally recommend for use with positively and negatively charged systems* (Table 2).

Table 2. Proton Affinities for Hydrogen Fluoride [CCSD(T) calculations]

Basis set	PA [kcal/mol]
cc-pVDZ	129.14
cc-pVTZ	125.31
cc-pVQZ	122.97
cc-pV5Z	121.94
aug-cc-pVDZ	118.97
aug-cc-pVTZ	121.63
aug-cc-pVQZ	121.62
aug-cc.pV5Z	121.54
exp.	122±1

Despite the fact that these basis sets have been developed specifically for charged systems, the perfomance of the aug-cc-basis sets is often superior to the cc-basis sets even for neutral molecules. The complexation energies calculated for the water dimer (Table 3) might serve as an example.

Table 3. Interaction Energies for the Water Dimer [RI-MP2(FC) calculations]

Basis set	ΔE [kcal/mol]
cc-pVDZ	-7.54
cc-pVTZ	-6.08
cc-pVQZ	-5.47
cc-pV5Z	-5.09
aug-cc-pVDZ	-5.31
aug-cc-pVTZ	-5.18
aug-cc-pVQZ	-5.09
aug-cc.pV5Z	-4.98
$\Delta H(375)$[theo.]	-3.2 ± 0.1
$\Delta H(375)$[exp.]	-3.6 ± 0.5

2. Practical considerations

With the exception of semiemipirical methods such as AM1, MNDO, and PM3, and compound methods such as G2 any basis set can be combined with any quantum mechanical method (not all combinations are meaningfull, though). In most cases, the basis set will be called from a library contained in quantum mechanics programs through an acronym such as "6-31G(d)". In the input file format used in Gaussian 98, quantum mechanical method and basis set information are separated by a "/". The following example requests a Hartree-Fock calculation using the 6-31G(d) basis set:

#P HF/6-31G(d)

If a **non standard basis set** is to be used, then the "**GEN**" directive is used instead of the basis set information in the keyword line and the basis set exponents and coefficients are given in fixed format after the geometry information (separated by one blank line):

```
#HF/GEN 6D

HF sp H2O using the 6-31G(d) basis set

0 1
O1
H2  1  r2
H3  1  r2  2  a3

r2=0.94733729
a3=105.50796079

O 1
 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 -0.1107775490D+00  0.7087426820D-01
  0.3599933586D+01 -0.1480262620D+00  0.3397528390D+00
  0.1013761750D+01  0.1130767010D+01  0.7271585770D+00
 SP   1 1.00
  0.2700058226D+00  0.1000000000D+01  0.1000000000D+01
 D   1 1.00
  0.8000000000D+00  0.1000000000D+01
 ****
H 0
 S   3 1.00
  0.1873113696D+02  0.3349460434D-01
  0.2825394365D+01  0.2347269535D+00
  0.6401216923D+00  0.8137573262D+00
 S   1 1.00
  0.1612777588D+00  0.1000000000D+01
 ****
```

The basis set information is given for each element separated by a line containing four stars. If this type of input is used, the program does not check whether all atoms have received basis functions. *If, for example, the basis set information for hydrogen is deleted from the last example, the calculation will still run as usual, but without any basis functions at hydrogen - with desasterous results!* Also some care has to be taken in specifying the correct number of d-orbitals (five pure d-type orbitals vs. six cartesian d-functions). If no additional information is given Gaussian 98 assumes the use of five d-type functions. This can be specified more explicitely using the keywords "5D" and "6D". In a similar fashion, the program can be directed to use either "7F" or "10F" polarization functions. The combination of a standard basis set and some additional basis functions is most easily achieved using the **"GEN" keyword**. In the following example the 6-31G basis set is called from the basis set library for all carbon and hydrogen atoms and an extra d-type polarization function is then added onto atom No. 1 (oxygen). Using an exponent of 0.8 for the six d-type functions reproduces exactly what is otherwise described as Pople's "6-31G(d)" basis set.

```
#HF/GEN 6D

HF sp H2O using the 6-31G basis + additional d-type functions (6D)

0 1
O1
H2  1  r2
H3  1  r2  2  a3

r2=0.94733729
a3=105.50796079

O H 0
6-31G
 ****
1 0
 D    1 1.00
 0.8000000000D+00  0.1000000000D+01
 ****
```

A listing of basis sets in a format appropriate for input as a general basis set can be obtained in two ways. (i) The first is using *Gaussian* itself to produce a listing of the currently used basis set. This can be achieved using the keyword **gfinput** *e.g.*

#P HF/6-31G(d,p) **gfinput**

This results in basis set information given for each single center contained in the system.

(ii) An alternative source of basis set information is provided by the EMSL Gaussian Basis Set Library at **http://www.emsl.pnl.gov/forms/basisform.html** which also provides a number of basis sets not provided as a standard basis set by *Gaussian*.

Literatur

1) W. J. Hehre, R. F. Stewart, J. A. Pople, Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Orbitals. *J. Chem. Phys.* **1969**, *51*, 2657. [STO-nG basis sets]

2) J. S. Binkley, J. A. Pople, W. J. Hehre, Self-Consistent Molecular Orbital Methods. 21. Small Split-Valence Basis Sets for First-Row Elements. *J. Am. Chem. Soc.* **1980**, *102*, 939. [3-21G and 6-21G basis sets]

3) M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro, W. J. Hehre, Self-Consistent Molecular Orbital Methods. 22. Small Split-Valence Basis Sets for Second-Row Elements. *J. Am. Chem. Soc.* **1982**, *104*, 2797. [3-21G, 2nd row elements]

4) W. J. Pietro, M. M. Francl, W. J. Hehre, D. J. DeFrees, J. A. Pople, J. S. Binkley, Self-Consistent Molecular Orbital Methods. 24. Supplemental Small Split-Valence Basis Sets for Second-Row Elements. *J. Am. Chem. Soc.* **1982**, *104*, 5039. [Polarization for 3-21G, 2nd row elements]

5) R. Ditchfield, W. J. Hehre, J. A. Pople, Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724. [4-31G basis set]

6) W. J. Hehre, R. Ditchfield, J. A. Pople, Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257. [6-31G basis set]

7) T. H. Dunning, Jr., Gaussian Basis Functions for Use in Molecular Calculations. I. Contraction of (9s5p) Atomic Basis sets for the First-Row Atoms. *J. Chem. Phys.* **1970**, *53*, 2823. [D95 basis]

8) T. H. Dunning Jr, P. J. Hay, in *Modern Theoretical Chemistry*, Ed. H. F. Schaefer, III, Plenum Press, New York, **1976**, Vol. 3, p.1. [D95 basis]

9) S. Huzinaga, *J. Chem. Phys.* **1965**, *42*, 1293. [Uncontracted 9s5p basis]

10) P. C. Hariharan, J. A. Pople, The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. chim. Acta.* **1973**, *28*, 213. [polarization functions for 6-31G]

11) D. J. DeFrees, B. A. Levi, S. K. Pollack, W. J. Hehre, J. S. Binkley, J. A. Pople, Effect of Electron Correlation on Theoretical Equilibrium Geometries. *J. Am. Chem. Soc.* **1979**, *101*, 4085. [HF/6-31G(d) and MP2/6-31G(d) results for small organics]

12) G. W. Spitznagel, T. Clark, J. Chandrasekhar, P. v. R. Schleyer, Stabilization of Methyl Anions by First-Row Substituents. The Superiority of Diffuse Function-Augmented Basis Sets for Anion Calculations. *J. Comp. Chem.* **1982**, *3*, 363.

13) T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. v. R. Schleyer, Efficient Diffuse Function-Augmented Basis Sets for Anion Calculations. III. The 3-21+G Basis Set for First-Row Elements, Li-F. *J. Comp. Chem.* **1983**, *4*, 294.

14) R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, Self-Consistent Molecular Orbital Methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650. [6-311G** basis set]

15) T.H. Dunning, Jr., Gaussian basis stes for use in correlated molecular calculations I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007. [cc basis sets]

16) R. A. Kendall, T. H. Dunning, Jr., R. J. Harrison, *J. Chem. Phys.* **1992**, *96*, 6796. [augmented cc basis sets]