CHEM3023: Spins, Atoms and Molecules

Lecture 6

Gaussian basis sets

C.-K. Skylaris

Learning outcomes

- Be able to convert expressions involving molecular orbitals to expressions involving basis functions and expansion coefficients and vice versa
- Be able to distinguish the characteristics and quality of different Gaussian basis sets
- Be able to select the basis set that provides the best compromise between required accuracy and computational efficiency for a given problem

UNIVERSITY OF Southampton
School of Chemistry

The SCF procedure for Hartree-Fock calculations

- A Hartree-Fock calculation provides the optimum molecular orbitals for a Slater determinant wavefunction
- These orbitals are the eigenfunctions of the Hartree-Fock eigenvalue equation, which involves the Fock operator

$$\hat{f}[\{\psi_i\}_{i=1}^{N_{elec}/2}]\ \psi_j(\mathbf{r}) = \varepsilon_j\, \psi_j(\mathbf{r})$$

- The Fock operator is very unusual: it is constructed from its own eigenfunctions
- An iterative scheme called the "Self Consistent Field" (SCF) procedure has been developed to find its eigenfunctions and hence the optimum molecular orbitals of Hartree-Fock theory

Basis functions

- Even for small molecules, the SCF procedure involves a huge amount of computation
- However, a computer performs operations on numbers rather than mathematical functions
- We need to transform the Hartree-Fock equations in a format that involves operations with numbers rather than functions
- To achieve this we write the molecular orbitals as linear combinations of some **fixed** set of functions which we call “basis functions”
- Often the basis functions are atomic orbitals: in this case a molecular orbital is a Linear Combination of Atomic Orbitals (LCAO)

molecular
(spatial) orbital

$$\psi_i(\mathbf{r}) = \sum_{\alpha=1}^{N_{BF}} G_\alpha(\mathbf{r})\, C_{\alpha i}$$

basis function
(fixed)

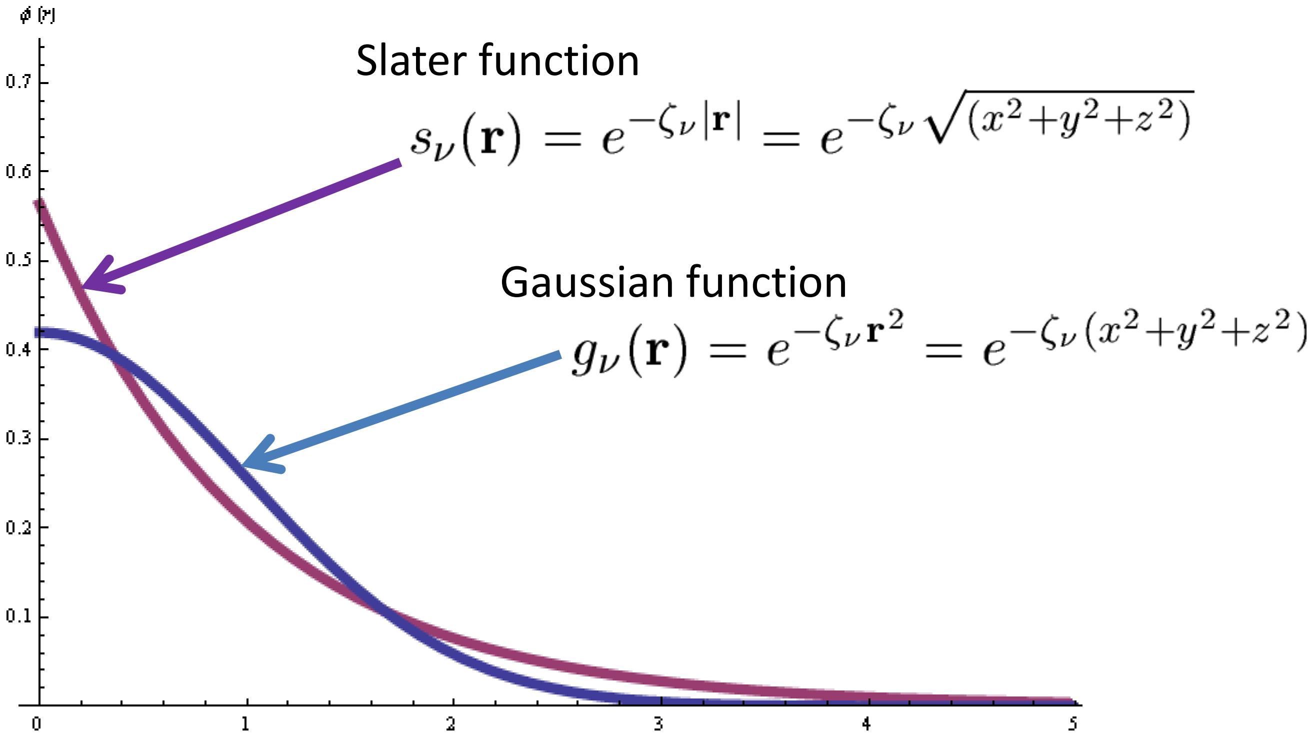
coefficient
(a number, to be
determined by the
SCF calculation)

Gaussian basis functions

Functions that resemble hydrogen atomic orbitals (Slater functions) are a very good basis set because they have the correct shape:

- Near the nucleus (cusp)
- Far from the nucleus (decay like e^{-ar})

Even though Gaussian functions do not behave that well, they are preferred in practice because they allow for efficient computation of molecular integrals



Slater function

$$s_\nu(\mathbf{r}) = e^{-\zeta_\nu |\mathbf{r}|} = e^{-\zeta_\nu \sqrt{(x^2+y^2+z^2)}}$$

Gaussian function

$$g_\nu(\mathbf{r}) = e^{-\zeta_\nu \mathbf{r}^2} = e^{-\zeta_\nu (x^2+y^2+z^2)}$$

Gaussian functions $g_\nu(\mathbf{r}) = x^k y^m z^n e^{-\zeta_\nu \mathbf{r}^2} = x^k y^m z^n e^{-\zeta_\nu (x^2+y^2+z^2)}$

Similar to atomic orbitals, we define **Cartesian primitive Gaussian** atomic functions by an angular momentum "quantum number" ℓ. We have $\ell = k + m + n$ which leads to the following types of function:

s functions

$$l = 0 \quad \left\{ \; e^{-\zeta \mathbf{r}^2} \right.$$

p functions

$$l = 1 \quad \left\{ \begin{array}{l} x e^{-\zeta \mathbf{r}^2} \\ y e^{-\zeta \mathbf{r}^2} \\ z e^{-\zeta \mathbf{r}^2} \end{array} \right.$$

d functions

$$l = 2 \quad \left\{ \begin{array}{l} x^2 e^{-\zeta \mathbf{r}^2} \\ y^2 e^{-\zeta \mathbf{r}^2} \\ z^2 e^{-\zeta \mathbf{r}^2} \\ xy e^{-\zeta \mathbf{r}^2} \\ xz e^{-\zeta \mathbf{r}^2} \\ zy e^{-\zeta \mathbf{r}^2} \end{array} \right.$$

etc.

Contracted Gaussian functions

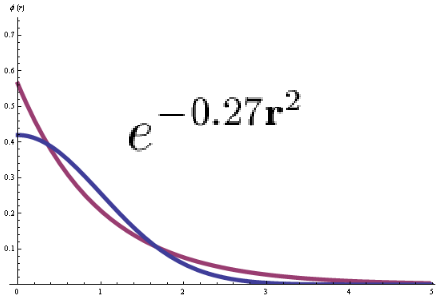
- In practice, fixed linear combinations of "primitive" Gaussian functions are used
- These are called "Contracted Gaussians" (**CGs**):

$$G_\alpha(\mathbf{r}) = \sum_{\nu=1}^{N_\alpha} c_\nu \, g_\nu(\mathbf{r})$$

- The simplest kind of CGs are the **STO-nG** basis sets
- These basis sets approximate Slater-type orbitals (STOs) by **n** primitive Gaussians

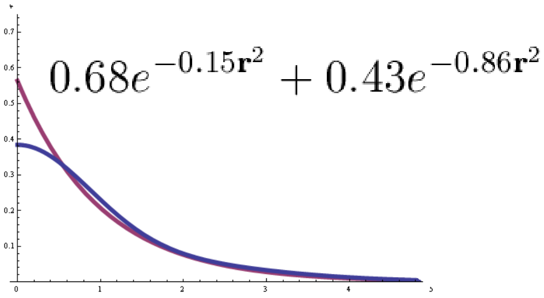
E.g. STO-nG functions for the 1s orbital of a hydrogen atom

STO-1G



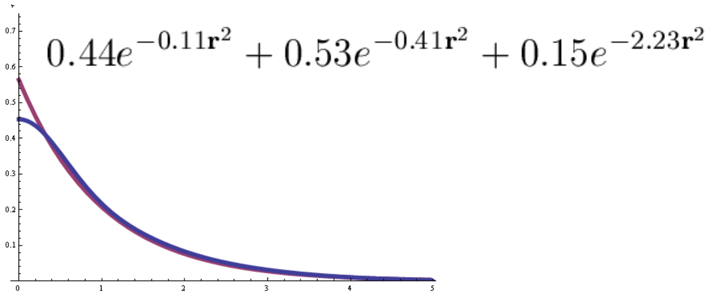
$$e^{-0.27\mathbf{r}^2}$$

STO-2G



$$0.68e^{-0.15\mathbf{r}^2} + 0.43e^{-0.86\mathbf{r}^2}$$

STO-3G



$$0.44e^{-0.11\mathbf{r}^2} + 0.53e^{-0.41\mathbf{r}^2} + 0.15e^{-2.23\mathbf{r}^2}$$

Gaussian basis sets

The STO-nG basis sets are rather unsatisfactory as they include only one contracted Gausssian (CG) per atomic orbital. Improved basis sets are obtained by including:

- More than one CG per atomic orbital, e.g.: **DZ** (“double zeta”), **TZ** (“triple zeta”), **QZ** (“quadruple zeta”)

- One CG per “core” atomic orbital and more than one for the valence atomic orbitals, e.g.: **SV**, **3-21G**, **4-31G**, **6-31G**, **6-311G**

Write down how many CGs some of the above basis sets will include for the following atoms: H, C, S

And for the following molecules: H_2O, CH_4

Polarisation and diffuse functions

Increasing the number of CGs per atomic orbital will never result in a good quality basis set. Other types of CGs need to be included, such as:

- CGs of angular momentum higher than in the valence orbitals of each atom. These "**polarisation functions**" enhance the "flexibility" of atoms to form chemical bonds in any direction and hence improve calculated molecular structures. Examples: **3-21G*, 6-31G*, 6-31G**, DVP, TZP, cc-pVDZ, cc-pVTZ**

- CGs which extend further from the nucleus than the atomic orbitals. Such "**diffuse functions**" improve the predicted properties of species with extended electronic densities such as anions or molecules forming hydrogen bonds. Examples: **4-31+G, 6-31+G**

- Basis sets are considered "balanced" when they include both polarisation and diffuse functions. Examples: **6-31+G*, 6-311++G**, aug-cc-pVDZ**

Write down how many CGs some of the above basis sets will include for the following atoms: H, F, S and molecules: H_2O, CH_4

The complete basis set limit

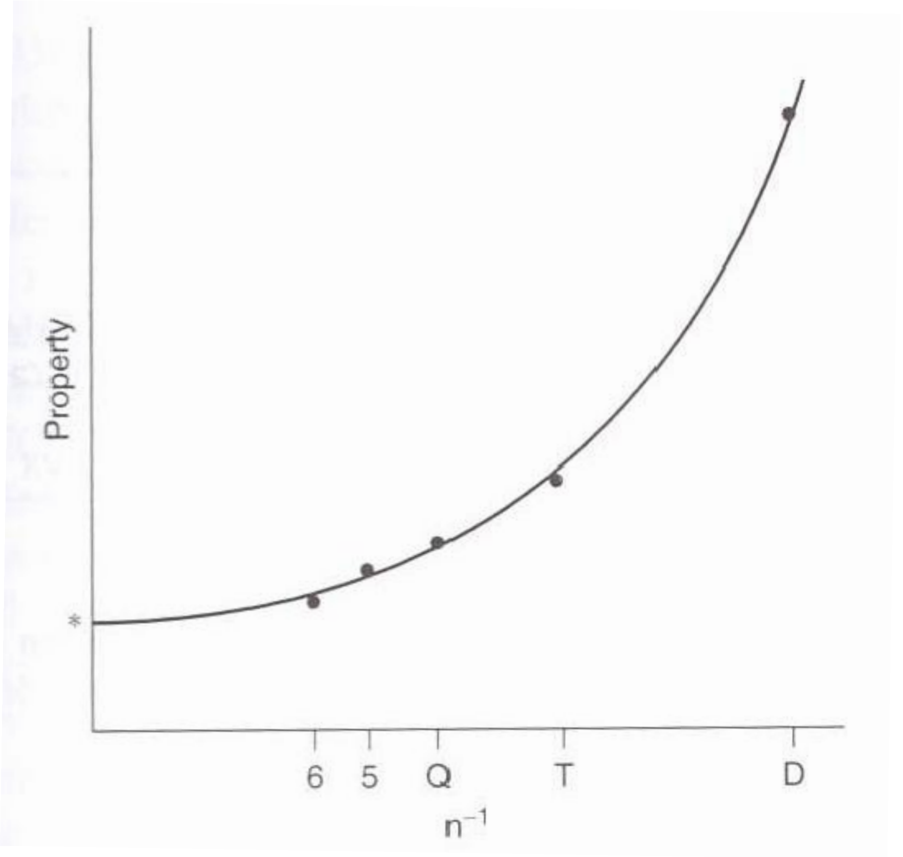
- Basis sets are an approximation introduced in order to solve equations for molecular orbitals on a computer
- The molecular orbitals obtained are solutions of the Hartree-Fock equations only within the “function space” of the basis set used (so solutions within the STO-3G set of functions, or the 6-31G set, etc.)
- Improving the quality of the basis set results in increasing the number of CGs
- A **complete** basis set can represent exactly **any molecular orbital**
- Unfortunately, complete basis sets tend to have an infinite number of functions and are therefore not practical for calculations

Extrapolation to the complete basis set limit

- We can estimate the complete basis set result by systematically increasing the number of basis functions and extrapolating to an infinite-size basis set

- In Hartree-Fock calculations, the complete basis set solutions are called the "**Hartree-Fock limit**"

- The cc-pVDZ, cc-pVTZ, cc-pVQZ, etc, basis sets are an example of a systematic series of basis sets that can be extrapolated to the complete basis set limit



Examples of H-F calculations with Gaussian basis sets

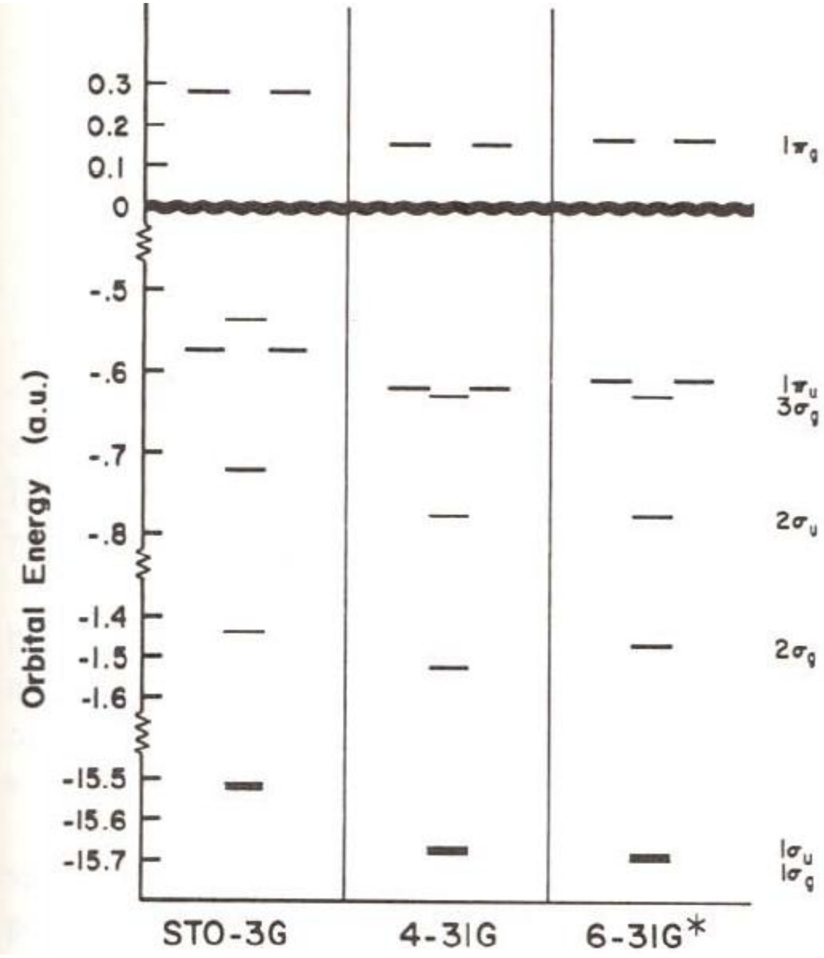
SCF total energies (a.u.) of N_2 and CO with the standard basis sets

Basis set	N_2	CO
STO-3G	−107.496	−111.225
4-31G	−108.754	−112.552
6-31G*	−108.942	−112.737
HF-limit[a]	−108.997	−112.791

SCF equilibrium bond angles for NH_3 and H_2O

Basis set	NH_3	H_2O
STO-3G	104.2	100.0
4-31G	115.8	111.2
6-31G*	107.5	105.5
6-31G**	107.6	106.0
Near-HF-limit	107.2[c]	106.1[b]
Experiment	106.7	104.5

Orbital energies of N2 with Gaussian basis sets

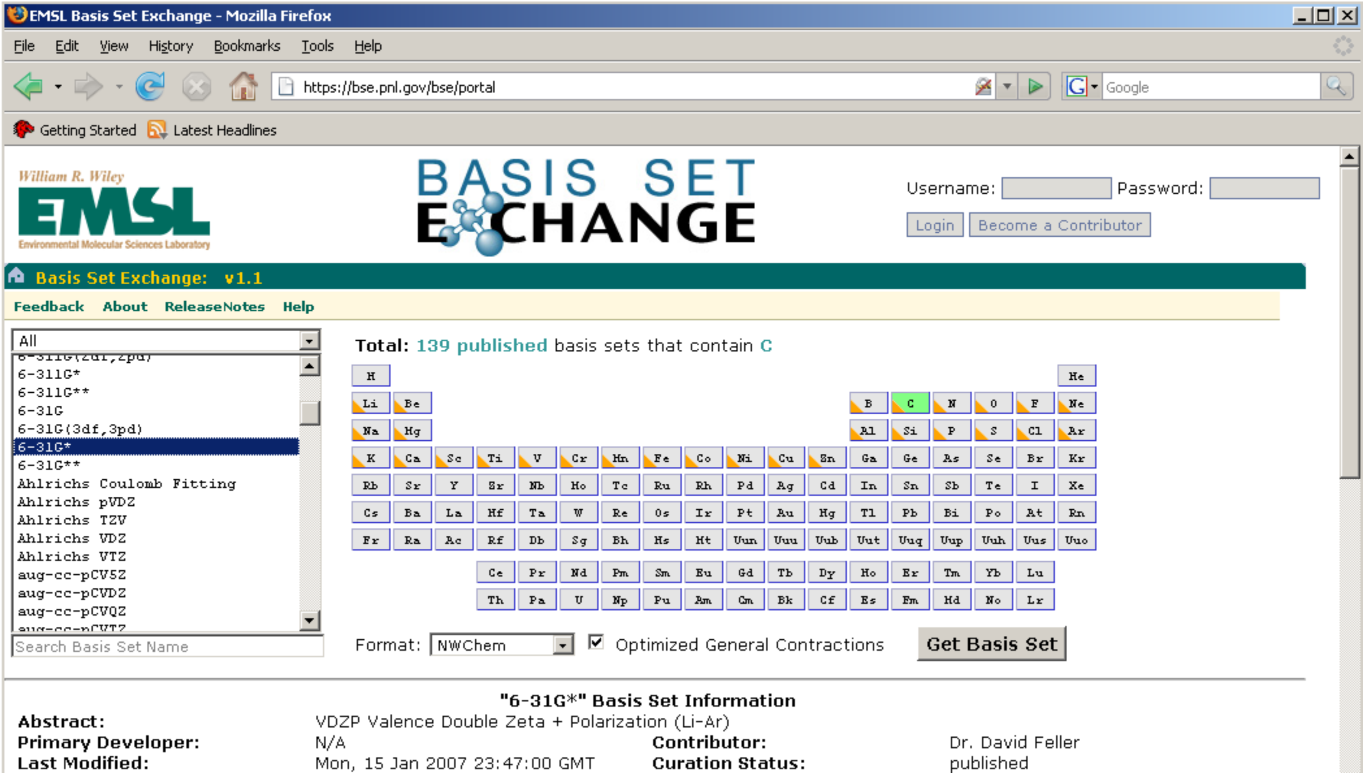


Example: available basis sets in a quantum chemistry program

3-21g
3-21++g
3-21gs
3-21++gs
3-21gsp
3-21gs_polarization
4-22gsp
4-31g
6-311g
6-311++g2d_2p
6-311g2df_2pd
6-311++g3df_3pd
6-311gs
6-311+gs
6-311gs_polarization
6-311gss
6-311++gss
6-311gss_polarization
6-31g
6-31++g
6-31g3df_3pd
6-31g-blaudeau
6-31gs
6-31+gs
6-31++gs
6-31gs-blaudeau
6-31gs_polarization
6-31gss
6-31++gss
6-31gss_polarization
ahlrichs_coulomb_fitting
ahlrichs_polarization
ahlrichs_pvdz
ahlrichs_tzv
ahlrichs_vdz
ahlrichs_vtz
aug-cc-pcv5z
aug-cc-pcvdz
aug-cc-pcvqz
aug-cc-pcvtz
aug-cc-pv5+dz
aug-cc-pv5+dz_diffuse
aug-cc-pv5z
aug-cc-pv5z_diffuse
aug-cc-pv6+dz
aug-cc-pv6+dz_diffuse
aug-cc-pv6z
aug-cc-pv6z_diffuse
aug-cc-pvd+dz
aug-cc-pvd+dz_diffuse
aug-cc-pvdz
aug-cc-pvdz_diffuse
aug-cc-pvq+dz
aug-cc-pvq+dz_diffuse
aug-cc-pvqz
aug-cc-pvqz_diffuse
aug-cc-pvt+dz
aug-cc-pvt+dz_diffuse
aug-cc-pvtz
aug-cc-pvtz_diffuse
bauschlicher_ano
binning-curtiss_1d_polarization
binning-curtiss_df_polarization
binning_curtiss_sv
binning_curtiss_svp
binning_curtiss_vtz
binning_curtiss_vtzp
blaudeau_polarization
cc-pcv5z
cc-pcv6z
cc-pcvdz
cc-pcvqz
cc-pcvtz
cc-pv5+dz
cc-pv5z
cc-pv5z_dk
cc-pv5zfi_sf_fw
cc-pv5zfi_sf_lc
cc-pv5zfi_sf_sc
cc-pv5zpt_sf_fw
cc-pv5zpt_sf_lc
cc-pv5zpt_sf_sc
cc-pv6+dz
cc-pv6z
cc-pvd+dz
cc-pvdz
cc-pvdz_dk
cc-pvdzfi_sf_fw
cc-pvdzfi_sf_lc
cc-pvdzfi_sf_sc
cc-pvdz-fit2-1
cc-pvdzpt_sf_fw
cc-pvdzpt_sf_lc
cc-pvdzpt_sf_sc
cc-pvdzseg-opt
cc-pvq+dz
cc-pvqz
cc-pvqz_dk
cc-pvqzfi_sf_fw
cc-pvqzfi_sf_lc
cc-pvqzfi_sf_sc
cc-pvqzpt_sf_fw
cc-pvqzpt_sf_lc
cc-pvqzpt_sf_sc
cc-pvqzseg-opt
cc-pvt+dz
cc-pvtz
cc-pvtz_dk
cc-pvtzfi_sf_fw
cc-pvtzfi_sf_lc
cc-pvtzfi_sf_sc
cc-pvtz-fit2-1
cc-pvtzpt_sf_fw
cc-pvtzpt_sf_lc
cc-pvtzpt_sf_sc
cc-pvtzseg-opt
chipman_dzp_+_diffuse
core_val._functions_cc-pcv5z
core_val._functions_cc-pcv6z
core_val._functions_cc-pcvdz
core_val._functions_cc-pcvqz
core_val._functions_cc-pcvtz
crenbl_ecp
crenbs_ecp
d-aug-cc-pv5z
d-aug-cc-pv5z_diffuse
d-aug-cc-pv6z
d-aug-cc-pv6z_diffuse
d-aug-cc-pvdz
d-aug-cc-pvdz_diffuse
d-aug-cc-pvqz
d-aug-cc-pvqz_diffuse
d-aug-cc-pvtz
d-aug-cc-pvtz_diffuse
demon_coulomb_fitting
dgauss_a1_dft_coulomb_fitting
dgauss_a1_dft_exchange_fitting
dgauss_a2_dft_coulomb_fitting
dgauss_a2_dft_exchange_fitting
dhms_polarization
dunning-hay_diffuse
dunning-hay_double_rydberg
dunning-hay_rydberg
dz_+_double_rydberg_dunning-hay
dz_dunning
dzp_+_diffuse_dunning
dzp_dunning
dzp_+_rydberg_dunning
dz_+_rydberg_dunning
dzvp2_dft_orbital
dzvp_dft_orbital
feller_misc._cvdz
feller_misc._cvqz
feller_misc._cvtz
gamess_pvtz
gamess_vtz
glendening_polarization
hay-wadt_mb_n+1_ecp
hay-wadt_vdz_n+1_ecp
hondo7_polarization
huzinaga_polarization
lanl2dzdp_ecp
lanl2dzdp_ecp_polarization
lanl2dz_ecp
mclean_chandler_vtz
midi!
midi_huzinaga
mini_huzinaga
mini_scaled
nasa_ames_ano
nasa_ames_cc-pcv5z
nasa_ames_cc-pcvqz
nasa_ames_cc-pcvtz
nasa_ames_cc-pv5z
nasa_ames_cc-pvqz
nasa_ames_cc-pvtz
partridge_uncontr._1
partridge_uncontr._2
partridge_uncontr._3
pople_2d_2p_polarization
pople_2df_2pd_polarization
pople_3df_3pd_polarization
pople-style_diffuse
pv6z
qmmm_zhang_3-21g_ecp
qmmm_zhang_6-31gs_ecp
sadlej_pvtz
sbkjc_vdz_ecp
sdb-aug-cc-pvqz
sdb-aug-cc-pvqz_diffuse
sdb-aug-cc-pvtz
sdb-aug-cc-pvtz_diffuse
sdb-cc-pvqz
sdb-cc-pvtz
sto-2g
sto-3g
sto-3gs
sto-3gs_polarization
sto-6g
stuttgart_rlc_ecp
stuttgart_rsc_1997_ecp
stuttgart_rsc_ano_ecp
stuttgart_rsc_segmented_ecp
sv_+_double_rydberg_dunning-hay
sv_dunning-hay
svp_+_diffuse_dunning-hay
svp_+_diffuse_+_rydberg
svp_dunning-hay
svp_+_rydberg_dunning-hay
sv_+_rydberg_dunning-hay
tz_dunning
tzvp_dft_orbital
wachters+f
wtbs

Basis sets on the Web

- All kinds of basis sets have been developed over the years
- Most are available for download from websites, such as
  - https://bse.pnl.gov/bse/portal
  - http://www.emsl.pnl.gov/forms/basisform.html



Total: 139 published basis sets that contain C

Format: NWChem ☑ Optimized General Contractions

"6-31G*" Basis Set Information

Abstract: VDZP Valence Double Zeta + Polarization (Li-Ar)

Primary Developer: N/A **Contributor:** Dr. David Feller

Last Modified: Mon, 15 Jan 2007 23:47:00 GMT **Curation Status:** published

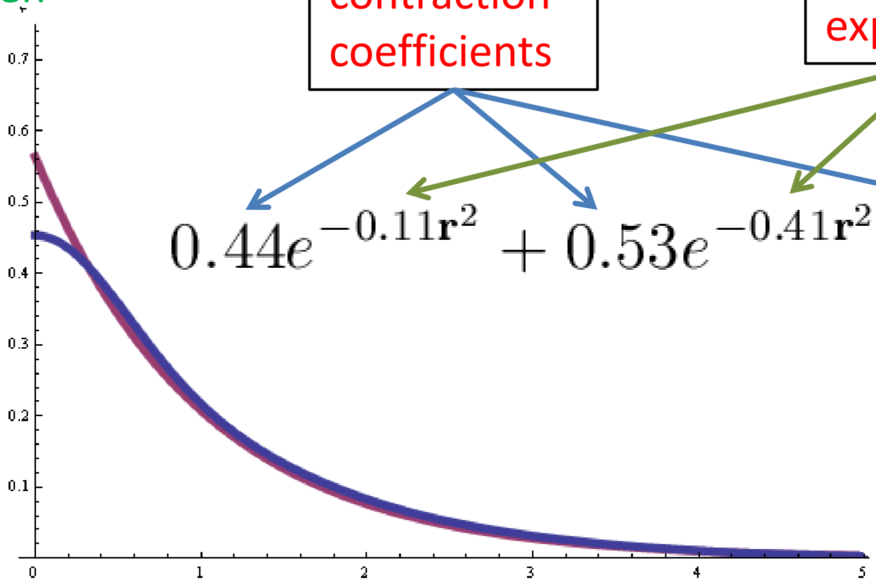
Exponents and contraction coefficients

Basis sets are essential input data for calculations. Gaussian bases are represented by two kinds of numbers:

1) Exponents
2) Contraction coefficients

$$G_\alpha(\mathbf{r}) = \sum_{\nu=1}^{N_\alpha} c_\nu \, e^{-\zeta_\nu \mathbf{r}}$$

Example: STO-3G basis for hydrogen



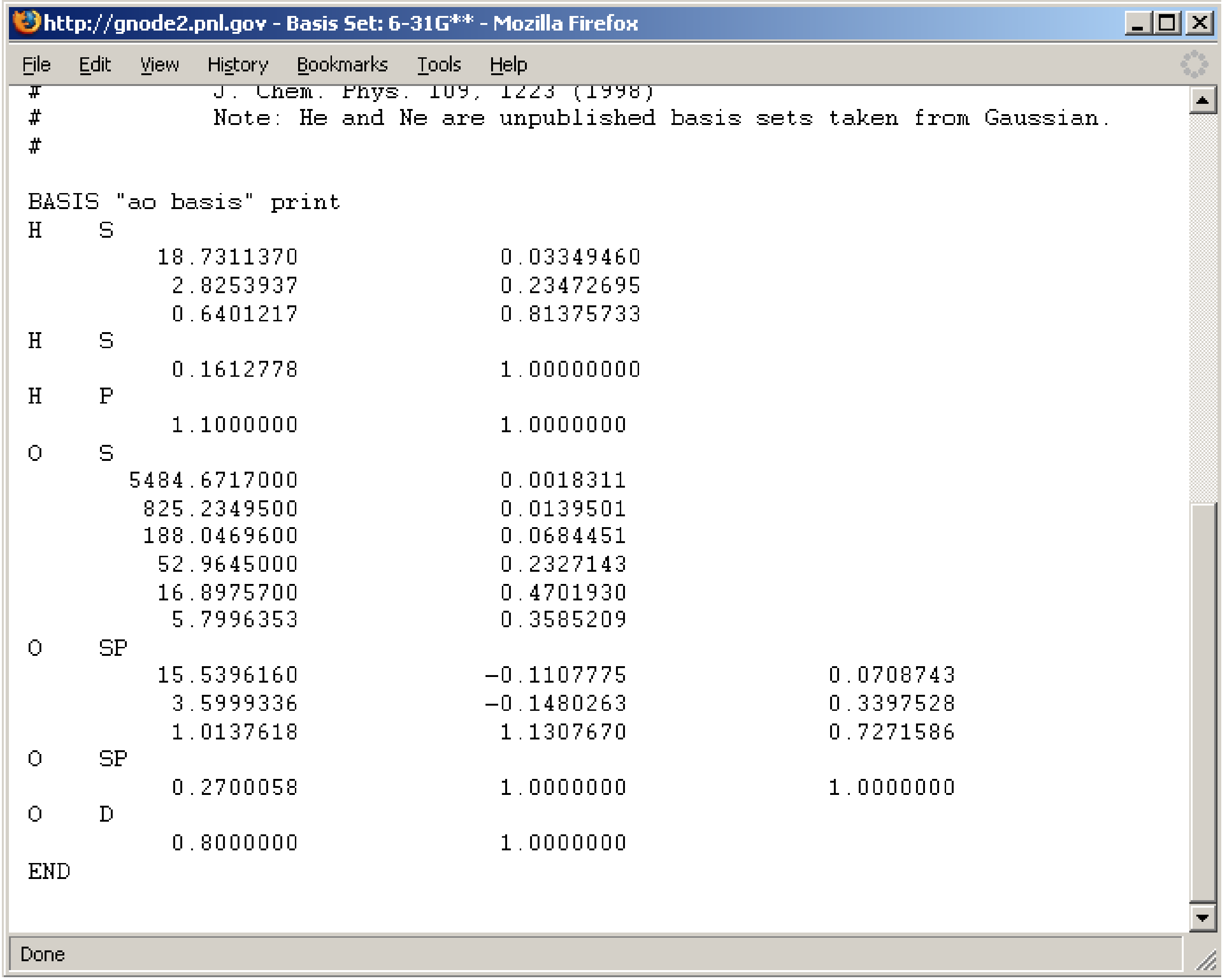
contraction coefficients

exponents

$$0.44e^{-0.11\mathbf{r}^2} + 0.53e^{-0.41\mathbf{r}^2} + 0.15e^{-2.23\mathbf{r}^2}$$

Downloading basis sets

Example: Data needed for calculation on water with 6-31G** basis



```
#            J. Chem. Phys. 109, 1223 (1998)
#            Note: He and Ne are unpublished basis sets taken from Gaussian.
#

BASIS "ao basis" print
H    S
         18.7311370             0.03349460
          2.8253937             0.23472695
          0.6401217             0.81375733
H    S
          0.1612778             1.00000000
H    P
          1.1000000             1.0000000
O    S
       5484.6717000             0.0018311
        825.2349500             0.0139501
        188.0469600             0.0684451
         52.9645000             0.2327143
         16.8975700             0.4701930
          5.7996353             0.3585209
O    SP
         15.5396160            -0.1107775             0.0708743
          3.5999336            -0.1480263             0.3397528
          1.0137618             1.1307670             0.7271586
O    SP
          0.2700058             1.0000000             1.0000000
O    D
          0.8000000             1.0000000
END
```

Summary / Reading assignment

- LCAO approach (Cramer, pages 111-113)
- Gaussian basis sets (Cramer, pages 166-178)