

# SCF methods, basis sets, and integrals

## Lecture III: Basis sets

Wim Klopper

Abteilung für Theoretische Chemie, Institut für Physikalische Chemie  
Karlsruher Institut für Technologie (KIT)

ESQC 2022, 11–24 September 2022

# Overview of basis functions

We may try to solve the Hartree–Fock or Kohn–Sham equations on a real-space grid in 3D. Accurate numerical Hartree–Fock methods exist for atoms and diatomic molecules.

Alternatively, we may expand the MOs or crystal wavefunctions in a set of **basis functions**. Examples include:

- Numerical atomic functions
- Finite elements (FEM)
- Wavelets
- Plane and spherical waves
- Slater-type orbitals (STOs)
- **Gaussian-type orbitals (GTOs)**

# Numerical atomic orbitals

- It is possible to use purely numerical atomic functions that are defined on a real-space grid in three dimensions.
- In density-functional theory (DFT), integrals are computed by a numerical quadrature in 3D.
- **DMol<sup>3</sup>** and **SIESTA** are DFT programs that use numerical atomic orbitals.
- The basis sets used by **DMol<sup>3</sup>** are denoted **Minimal, DN, DND, DNP, TNP**. Also **SIESTA** uses multiple-zeta and polarisation functions.
- In these programs, DFT is easily implemented in the local-density (LDA) and generalised-gradient approximations (GGA). **Hybrid functionals with Hartree–Fock exchange are more difficult.**
- The potentials  $V_{\text{ne}}(\mathbf{r})$ ,  $J(\mathbf{r})$  and  $v_{\text{xc}}(\mathbf{r})$  are **local**.

## Numerical atomic orbitals

- In DFT, without exact exchange, all potentials are local, and the Coulomb potential at a **grid point  $\mathbf{r}_p$**  can be computed as

$$J(\mathbf{r}_p) \approx \sum_{q=1}^{n_{\text{grid}}} w_q \sum_{i=1}^{n_{\text{occ}}} \frac{\varphi_i^*(\mathbf{r}_q)\varphi_i(\mathbf{r}_q)}{|\mathbf{r}_p - \mathbf{r}_q|} = \sum_{q=1}^{n_{\text{grid}}} w_q \frac{\rho(\mathbf{r}_q)}{|\mathbf{r}_p - \mathbf{r}_q|}$$

- The  $w_q$  are the appropriate weights of the quadrature.
- Matrix elements of the Coulomb and local exchange–correlation potentials can be computed as

$$\langle \chi_\mu | \hat{J} | \chi_\nu \rangle = \int \chi_\mu^*(\mathbf{r}) J(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r} \approx \sum_{p=1}^{n_{\text{grid}}} w_p \chi_\mu^*(\mathbf{r}_p) J(\mathbf{r}_p) \chi_\nu(\mathbf{r}_p)$$

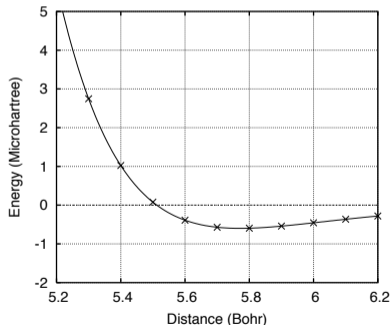
$$\langle \chi_\mu | v_{\text{xc}} | \chi_\nu \rangle = \int \chi_\mu^*(\mathbf{r}) v_{\text{xc}}(\mathbf{r}) \chi_\nu(\mathbf{r}) d\mathbf{r} \approx \sum_{p=1}^{n_{\text{grid}}} w_p \chi_\mu^*(\mathbf{r}_p) v_{\text{xc}}(\mathbf{r}_p) \chi_\nu(\mathbf{r}_p)$$

# Numerical molecular orbitals

- Some DFT implementations (such as **Octopus**) attempt to describe the molecular Kohn–Sham orbitals on a real-space grid.
- A 3D simulation box is chosen together with a grid spacing, for example  $0.5 a_0$ . Then, a grid in 3D is constructed and the SCF equations are solved on the grid.
- **This is different from an MO-LCAO expansion in numerical AOs!**
- Pseudopotentials are inevitable for real-space grid methods, but they are not required when numerical AOs are used.
- A great advantage of the use of numerical AOs as in **DMol<sup>3</sup>** is that the method is free of the basis-set superposition error (BSSE).
- Because **exact atomic orbitals** are used, the atoms in a molecule cannot improve their orbitals artificially using basis functions from other atoms.

## Basis-set superposition error (BSSE)

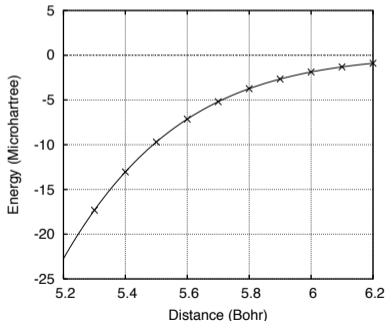
- A famous example of BSSE is the **Hartree–Fock** calculation of the He...He potential curve in a two-function **3-21G** basis:



- The RHF/3-21G calculation of He...He yields an interaction energy of  $-0.6 \mu E_h$  at  $R = 5.77 a_0$ .
- The Hartree–Fock curve should be purely repulsive!
- Accidentally, the Hartree–Fock minimum is close to the true minimum at  $5.60 a_0$ . The true well depth amounts to ca.  $-35 \mu E_h$ .
- The RHF/3-21G energy of the He atom is in error by  $26 mE_h$ .

## Basis-set superposition error (BSSE)

- Let us compute the RHF/3-21G energy of one He atom while another 3-21G basis (**without atom**) is approaching.



- Shown is the computed energy relative to the RHF calculation in only the atom's own 3-21G basis.
- This is the BSSE: artificial energy lowering due to neighbouring functions.**
- At  $R = 5.77 a_0$ , the artificial energy lowering is  $-4.1 \mu E_h/\text{atom}$  ( $-8.2 \mu E_h$  for both atoms).
- We should add  $8.2 \mu E_h$  to the computed interaction energy of  $-0.6 \mu E_h$ .

## The counterpoise correction

- Thus, at  $R = 5.77 a_0$ , we obtain a **repulsive potential of  $+7.6 \mu E_h$**  at the RHF/3-21G level if we correct for BSSE.
- This correction is known as **counterpoise correction**. It consists of computing not only the system of interest but also its fragments in the basis set of the whole system.
- The interaction energy is computed by subtracting the energies of the fragments computed in the whole basis.



- In practice, the basis set in a counterpoise calculation is most easily defined by setting the nuclear charge of the corresponding atom to zero (**ghost atom**).



# The counterpoise correction

- The CP-corrected interaction energy is directly obtained by calculating both the system and the fragments in the same basis,

$$\Delta E_{\text{CP corrected}} = E_{\text{AB}} - E_{\text{A+ghost(B)}} - E_{\text{B+ghost(A)}}$$

- The CP corrections to fragments A and B are defined as follows:

$$\delta_{\text{CP}}(\text{A}) = E_{\text{A}} - E_{\text{A+ghost(B)}}, \quad \delta_{\text{CP}}(\text{B}) = E_{\text{B}} - E_{\text{B+ghost(A)}}$$

- Hence, the CP-corrected interaction energy can also be computed from

$$\Delta E_{\text{CP corrected}} = \Delta E_{\text{CP uncorrected}} + \delta_{\text{CP}}(\text{A}) + \delta_{\text{CP}}(\text{B})$$

$$\Delta E_{\text{CP uncorrected}} = E_{\text{AB}} - E_{\text{A}} - E_{\text{B}}$$

- Using numerical AOs,  $E_{\text{A}} = E_{\text{A+ghost(B)}} = E_{\text{A}}(\text{exact})!$

# Counterpoise corrected binding energies

- Usually, **free fragments** have another geometry than in the complex (such as the H<sub>2</sub>O dimer).



- The **binding energy** is the energy of the complex or supermolecule in its optimized geometry relative to the energies of the dissociation products **in their own, optimized geometries**,

$$E_{\text{binding energy}} = \Delta E^{(1)} + \Delta E^{(2)} = \Delta E_{\text{CP corrected}} + \Delta E^{(1)}$$

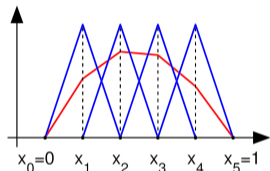
- $\Delta E^{(1)}$  is a one-body term. It contains the **relaxation energy of the dissociation products**,

$$\Delta E^{(1)} = E_{\text{A, complex geom.}} - E_{\text{A, relaxed geom.}} + \text{same for B}$$

- As defined here, the binding energy is a negative quantity. Often, however, it is reported as a positive value.

# Finite elements methods (FEM)

- The finite-element method is an expansion method which uses a strictly local, piecewise polynomial basis.

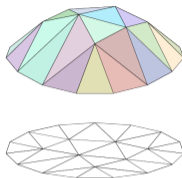


$$f(x) = \sum_{k=1}^4 c_k f_k(x)$$

- It combines the advantages of basis-set and real-space grid approaches.
- A finite element is a **basis function**, which takes the value 1 at a grid point in real space, but which is 0 at its neighbouring grid points and at all other grid points.
- In its simplest form, the basis function is linear between two grid points  $x_k$  and  $x_{k+1}$ .

# Finite elements methods (FEM)

- In 2D, the space is divided up in triangles and the surface is approximated by piecewise linear functions (see figure).
- FEM is also applicable in 3D.



- FEM has been used for **benchmark Hartree–Fock and MP2 (2<sup>nd</sup>-order Møller–Plesset perturbation theory) calculations of atoms** (e.g., with partial waves up to  $L = 12$ ).
- FEM has also been used for benchmark calculations of one-electron diatomics and for benchmark DFT calculations of diatomic systems.
- Modern techniques: Hermite interpolation functions, adaptive curvilinear coordinates, separable norm-conserving pseudopotentials, periodic boundary conditions, multigrid methods.

# Wavelets

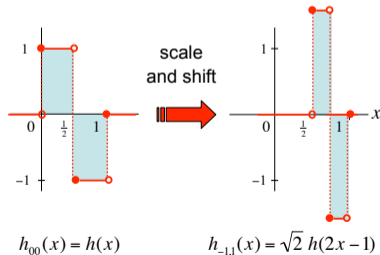
- Wavelets are a relatively new basis set in electronic structure calculations.
- Being localised both in real and in Fourier space, wavelets combine the advantages of local basis-set and plane waves.
- Localised orbitals and density matrices can be represented in a very compact way, and wavelets therefore seem an ideal basis set for  $\mathcal{O}(N)$  schemes.
- There exist **fast wavelet transforms (FWT)**.
- As an example, we shall consider the **Haar wavelets**, but there are many others (e.g., **Daubechies wavelets**, which can be used in electronic-structure theory).
- The Haar transform is very useful in image compression (JPEG).
- To the author's knowledge, an efficient general-purpose DFT program is not yet available.

# Wavelets

- A simple example is the **Haar wavelet**,

$$h_{mn}(x) = 2^{-m/2} h(2^{-m}x - n) \quad \text{with} \quad h(x) = \begin{cases} 1, & \text{if } 0 \leq x < 1/2 \\ -1, & \text{if } 1/2 \leq x < 1 \\ 0, & \text{otherwise} \end{cases}$$

- $h(x)$  is denoted as **mother wavelet**.
- The wavelets  $\{h_{mn}(x)\}$  form an **orthonormal basis**.

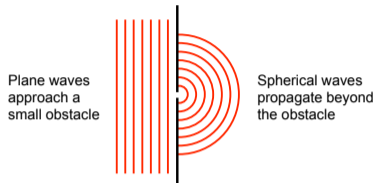


# Plane (and spherical) waves

- Plane (and spherical) waves are used in DFT codes that treat the electronic structure of **condensed matter**.
- CPMD, FLEUR, VASP and Wien2K are programs using plane waves.
- The basis functions can be written as

$$U_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \text{ (plane wave),} \quad \text{and} \quad U_k(\mathbf{r}) = \frac{e^{ikr}}{r} \text{ (spherical wave)}$$

- **Advantage of plane wave codes:** After defining a 3D box, the number of plane waves and the basis-set quality is controlled by a single energy-cutoff value. Basis functions up to that energy level are considered.

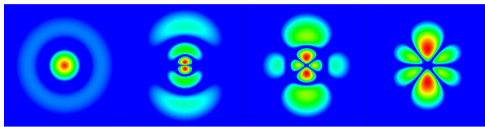


# Pseudopotentials (PPs)

- **Disadvantage of plane wave approaches:** It is very difficult for plane waves to describe the electronic structure near the nuclei.
- One solution to this problem consists of using (ultra-soft) pseudopotentials (US-PP).
- The idea is that with PPs, the (remaining) eigenstates and the electron density are much smoother than without. Plane waves can only handle a smooth potential well.
- Typical cutoff values range from 10–20  $E_h$  for **Vanderbilt ultra-soft pseudopotentials**, 30–50  $E_h$  for **Troullier–Martins norm-conserving pseudopotentials** to 40–100  $E_h$  for **Goedecker pseudopotentials** (*i.e.*, higher values for less soft PPs).
- With PPs, the number of plane waves is of the order of 100 per atom. Modern programs can treat thousands of valence electrons.



# Hydrogen atom eigenfunctions



- The **hydrogenic functions** seem to form a natural basis for the MO-LCAO Ansatz.
- These are the true atomic functions of hydrogen and H-like ions. The *bounded eigenfunctions* may be written as

$$\psi_{nlm} = R_{nl}(r) Y_l^m(\vartheta, \varphi)$$
$$R_{nl}(r) = \left(\frac{2Z}{n}\right)^{3/2} \sqrt{\frac{(n-l-1)!}{2n(n+l)!}} \left(\frac{2Zr}{n}\right)^l L_{n-l-1}^{2l+1} \left(\frac{2Zr}{n}\right) \exp\left(-\frac{Zr}{n}\right)$$

- The radial part contains an **associated Laguerre polynomial**  $L_{n-l-1}^{2l+1}$  in  $2Zr/n$  and an **exponential in**  $-Zr/n$ .

# Hydrogen atom eigenfunctions

- The H-atom eigenfunctions are the exact solutions for a one-electron Coulombic system, but the functions  $\psi_{nlm}$  are **not useful** as basis functions for many-electron atoms or molecules.
- In 1928, it was already recognised by Born and Hylleraas that the He atom could not be described by a CI expansion using the H-like *bound-state eigenfunctions*.
- To constitute a **complete set**, the bound-state eigenfunctions must be supplemented by the **unbounded continuum states**.
- Furthermore, the H-like functions spread out rapidly and become quickly too diffuse for calculations of the core and valence regions of a many-electron atom.

$$\langle \psi_{nlm} | r | \psi_{nlm} \rangle = \frac{3n^2 - l(l+1)}{2Z}$$

- They may be useful to describe **Rydberg states**.

# Hydrogen atom eigenfunctions

- The problem with the H-atom eigenfunctions is that the **exponent  $Z/n$**  in the exponential **decreases when  $n$  increases**,

$$\psi_{nlm} \propto (r/n)^l L_{n-l-1}^{2l+1}(2Zr/n) \exp(-Zr/n)$$

- It seems a good idea to change to functions of the type

$$\chi_{nlm} \propto (\zeta r)^l L_{n-l-1}^{2l+2}(2\zeta r) \exp(-\zeta r)$$

- These **Laguerre functions** form a complete, orthonormal set in  $L^2(\mathbb{R}^3)$ .
- Laguerre functions are very useful for highly accurate work on atoms.

## Nodeless Slater-type orbitals (STOs)

- We can expand the Hartree–Fock orbital of He in a basis of Laguerre functions,

$$\varphi_{\text{He}}(\mathbf{r}) = \sum_{n=1}^{n_{\max}} c_n L_{n-1}^2(2\zeta r) \exp(-\zeta r)$$

- There is one nonlinear parameter ( $\zeta$ , which could be determined via  $\langle \hat{V} \rangle = -2\langle \hat{T} \rangle$ ) and we must choose the expansion length.
- Can we fix  $n$  and use variable exponents?

$$\varphi_{\text{He}}(\mathbf{r}) = \sum_{k=1}^{k_{\max}} c_k \exp(-\zeta_k r)$$

- Can we even take variable exponents *and* variable powers in  $r$ ?

$$\varphi_{\text{He}}(\mathbf{r}) = \sum_{n=1}^{n_{\max}} \sum_{k=1}^{k_{\max}(n)} c_{nk} r^{n-1} \exp(-\zeta_{nk} r)$$

# Slater-type orbitals (STOs)

The figure shows the radial distribution

$$4\pi r^2 [\Phi_{2s}(r)]^2$$

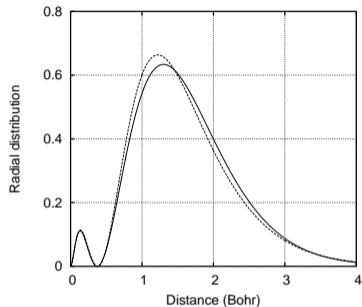
of the C atom from a **minimal 2s1p basis** (solid line) and from an **extended 6s4p basis** (dashed line).

In the minimal basis:

$$\varphi_{2s}(r) = -0.231 N_{1s} \exp(-5.58 r) + 1.024 N_{2s} r \exp(-1.46 r)$$

In the extended basis:

$$\varphi_{2s}(r) = \sum_{k=1,2} c_{k1} N_{1s} \exp(-\zeta_{k1} r) + \sum_{k=1,4} c_{k2} N_{2s} r \exp(-\zeta_{k2} r)$$



## Slater-type orbitals (STOs)

- Clementi–Roothaan–Yoshimine  $6s4p$  STO basis for carbon:

STO type	Exponents	Coefficients		
		$1s$	$2s$	$2p$
$1s$ STO	9.2683	0.07657	-0.01196	
	5.4125	0.92604	-0.21041	
$2s$ STO	4.2595	0.00210	-0.13209	
	2.5897	0.00638	0.34624	
	1.5020	0.00167	0.74108	
$2p$ STO	1.0311	-0.00073	0.06495	
	6.3438			0.01090
	2.5873			0.23563
	1.4209			0.57774
	0.9554			0.24756

$$\varphi_{2s}(r) = -0.01196 N_{1s} \exp(-9.2683 r) + \dots + 0.06495 N_{2s} r \exp(-1.0311 r)$$

- The **extended basis** contains  $2 + 4 + 4 \times 3 = 18$  basis functions.
- The (Hartree–Fock) coefficients are given with respect to **normalised** basis functions.
- The linear combinations with the Hartree–Fock coefficients can be used as a **minimal basis** comprising  $1 + 1 + 1 \times 3 = 5$  basis functions (**contractions**).

# Slater-type orbitals (STOs)

## Advantages of STOs:

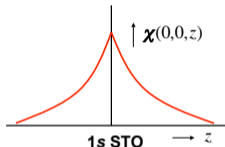
- Correct description of the **cusp** at the nucleus. For a one-electron system, for example, we have

$$\varphi_{1s} \propto Rr, \quad \left. \frac{\partial R(r)}{\partial r} \right|_{r=0} = -Z R(0) \neq 0$$

- STOs have the correct asymptotic long-range behaviour,

$$\varphi_{\text{HOMO}} \propto \exp(-\zeta r), \quad \zeta = \sqrt{2 \cdot \text{IP}} = \sqrt{2 \cdot |\varepsilon_{\text{HOMO}}|}$$

- Accurate calculations are possible for atoms and diatomics.



# Slater-type orbitals (STOs)

## Disadvantages of STOs:

- No efficient program available to evaluate the many-centre two-electron STO integrals.
- Long-range behaviour of the density is correct only if the smallest STO exponent is  $\zeta_{\min} = \sqrt{2 \cdot \text{IP}}$ . Stable molecules have  $\text{IP} > 5$  eV. Hence,  $\zeta$  should not be smaller than  $0.6 a_0^{-1}$ , but lower values are often required for accurate work on molecules.

A program that uses STOs is **ADF**.

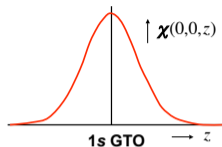
- The basis sets used by this program are denoted **SZ, DZ, DZP, TZP, TZ2P**.



# Gaussian-type orbitals (GTOs)

- In molecular calculations, the many-centre integrals are much easier to compute with Gaussian-type orbitals,

$$\chi(\mathbf{r}) \propto x^k y^l z^m \exp(-\alpha r^2)$$



- GTOs have no cusp at the nucleus, but this is not a main concern in chemical applications.
- The cusp occurs with point charges. For more realistic nuclei with finite extension, the Gaussian shape is actually more realistic.
- GTOs have the wrong asymptotic long-range behaviour, but the error due to falling off too quickly is less severe than the too long tail of an STO with too small exponent.
- Accurate calculations are possible for polyatomic molecules!**
- In terms of accuracy/effort, GTOs win over STOs.

# Gaussian basis sets: Overview

- Minimal basis sets (STO- $n$ G)
- Double-zeta basis sets (DZ, SV, 6-31G)
- Pople basis sets (6-311G\*, 6-311+G(2*df*,2*pd*), etc.)
- Karlsruhe “def2” basis sets
- Polarisation-consistent basis sets (pc- $n$ )
- Atomic natural orbital (ANO) basis sets
- Correlation-consistent basis sets (cc-pVXZ)
- Special-purpose basis sets (IGLO, Sadlej)
- Effective core potentials (*e.g.*, LANL2DZ)
- Auxiliary basis sets (RI- $J$ , RI- $JK$ , “cbas”, “cabs”)

## Gaussian basis sets: Purpose

- Choosing the right basis depends much on the type of calculation that we want to perform.
- Be aware that different basis sets are needed for Hartree–Fock and DFT calculations on the one hand and electron-correlation calculations ( $MP_n$ , CI, CC) on the other.
- The electron density of negative ions may be extended in space and GTOs with small exponents are required (**diffuse functions**).
- For some properties, the region near the nucleus is important (e.g., electric field gradient at the nucleus, Fermi contact term). Then, GTOs with large exponents are required (**tight functions**).
- Van der Waals intermolecular interactions need diffuse functions and are different from strongly covalently bound molecules.
- Be aware of the BSSE.

## STO- $n$ G basis sets

- The STO- $n$ G basis sets are **minimal basis sets**.
- The idea is to represent a Slater-type orbital (STO) by a linear combination of GTOs.
- In the STO-3G basis, for example,

$$N \exp(-\zeta r) \approx \sum_{k=1}^3 c_k N_k \exp(-\alpha_k r^2)$$

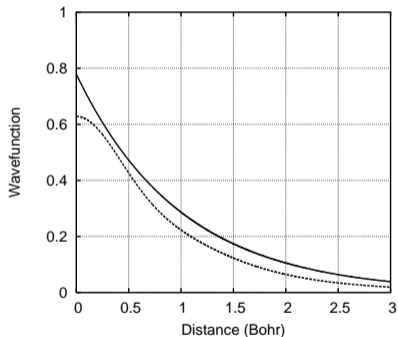
- For hydrogen, the following STO-3G basis represents the standard STO with exponent  $\zeta = 1.24 a_0^{-1}$ :

$k$	1	2	3
$\alpha_k/a_0^{-2}$	3.42525091	0.62391373	0.1688554
$c_k$	0.15432897	0.53532814	0.4446345

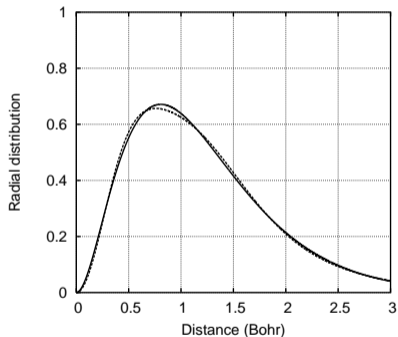
- The **exponents**  $\alpha_k$  and **contraction coefficients**  $c_k$  are obtained by a least-squares fit. A **contraction** is one single basis function, which itself is a fixed linear combination of (**primitive**) GTOs.

## STO- $n$ G basis sets

The H-atom STO-3G function (dashed line) replaces an STO with  $\zeta = 1.24$  (solid line).



$$\chi(r)$$

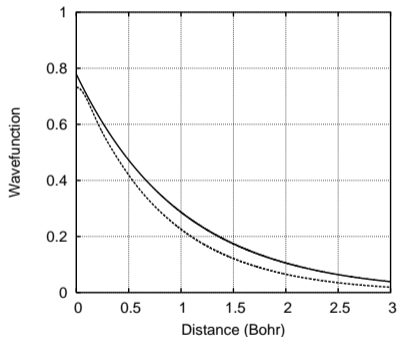


$$4\pi r^2 \chi^2(r)$$

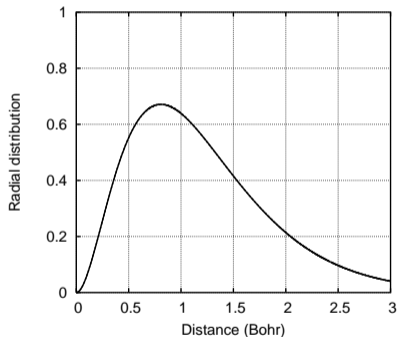
The figure on the left shows that the STO-3G basis function has no cusp at  $r = 0$ .

## STO- $n$ G basis sets

The H-atom STO-6G function (dashed line) replaces an STO with  $\zeta = 1.24$  (solid line).



$$\chi(r)$$



$$4\pi r^2 \chi^2(r)$$

The figure on the left shows that the STO-6G basis function has no cusp at  $r = 0$ .

## STO- $n$ G basis sets

- STO-3G basis sets exist for the atoms H–I.
- STO-6G basis sets exist for all atoms H–Kr.
- The exponents of the primitive Gaussians are chosen in a special manner. **The same exponents are chosen for the various angular momenta in an atomic shell.**
- For example, the same three exponents 7.295991196, 2.841021154 and 1.250624506 are used to replace the  $4s$ ,  $4p$  and  $4d$  STOs of iodine by Gaussians.
- **Choosing the same exponents may speed up the integral evaluation significantly**, but not all programs exploit this opportunity.
- If a certain STO- $n$ G basis function substitutes an STO with exponent  $\zeta$ , then a similar STO- $n$ G basis function with exponents  $\alpha'_k = \alpha_k \times (\zeta'/\zeta)^2$  replaces an STO with exponent  $\zeta'$ .

# Cartesian versus spherical-harmonic GTOs

- We may want to use **Cartesian GTOs**—centred at the centre  $\mathbf{A}$  (usually an atom)—of the form

$$\chi(\mathbf{r}; \alpha, k, l, m, \mathbf{A}) = N_{klm,\alpha} (x - x_A)^k (y - y_A)^l (z - z_A)^m \exp(-\alpha |\mathbf{r} - \mathbf{A}|^2)$$

- A set of  $f$ -type functions ( $l = 3$ ) is then defined by all combinations with  $k + l + m = 3$ . This yields **10 Cartesian  $f$ -type functions**. Similarly, there are **6 Cartesian  $d$ -type functions**, etc.
- The linear combination of 3 of the 6 Cartesian  $d$ -type functions corresponds to an  $3s$ -type function ( $x^2 + y^2 + z^2$ ). Similarly, the 10-component  $f$ -set contains three  $4p$ -type functions:  $(x^2 + y^2 + z^2)x$ , etc.
- It is much better to use the **spherical-harmonic GTOs** ( $5d, 7f, 9g$ , etc.) in the place of Cartesian GTOs to avoid near-degeneracies in the basis set. Most programs do this, but note that some standard basis-set definitions imply that they are Cartesian.



## Double-zeta and split-valence basis sets

- The **double-zeta (DZ)** basis set consists of **two basis functions** per atomic orbital and is twice as large as the minimal basis set.
- The **split-valence (SV)** basis is a **minimal basis for core orbitals** and is of **double-zeta quality for the valence shell**.
- Examples of SV basis sets are the **3-21G** (atoms H–Cs), **4-31G** (atoms H–Ne, P–Cl) and **6-31G** (H–Zn) basis sets.
- The notation “6-31G” means that 6 primitive GTOs are contracted to one basis function to describe the core orbitals. Furthermore, 3 primitive GTOs are contracted to the first basis function for the valence shell while another GTO is used as second basis function.
- Also in (most of) these basis sets, the exponents are constraint to be equal in  $ns$  and  $np$  shells.

# Polarisation functions

- The inclusion of a set of **polarisation functions** is often indicated by “P” or by an asterisk.
- Polarisation functions are basis functions with angular momentum that is not occupied in the atom, for example, *p*-type functions of H or *d*-type functions on O.
- **Polarisation functions are important when polarisation is important.**
- For example, the dipole moment of H<sub>2</sub>O amounts to  $0.96 ea_0$  in the **SV** basis but to  $0.83 ea_0$  in the **SVP** basis.
- Another example is the barrier to rotation in H<sub>2</sub>O<sub>2</sub>. The interaction between the dipoles along the polar OH bonds must be described accurately with polarisation functions.
- The polarisation functions are not always added to the H atoms. They *are* in sets denoted as **6-31G\*\*** and **SVP** but *not* in sets denoted **6-31G\*** and **SV(P)**.

## Valence triple-zeta plus polarisation

- **Recommended for molecular SCF calculations:** basis sets such as SV(P), SVP, 6-31G\* or 6-31G\*\*.
- For accurate SCF calculations, triple-zeta basis sets may be used. They are usually used with **polarisation functions**,
  - **6-311G\***: three contractions (311) for the valence shell, no polarisation functions on H.
  - **6-311G\*\***: same as 6-311G\* but with pol. func. on H.
  - **6-311G(2df,2pd)**: same as 6-311G\* but with  $2p1d$  polarisation set on H and  $2d1f$  set on other atoms.
  - **6-311G(3df,3pd)**: same as 6-311G(2df,2pd) but with 3  $d$  and 3  $p$  sets.
  - **def2-TZVP**: valence triple-zeta plus  $1p$  polarisation for H,  $2d1f$  for B–Ne and Al–Ar,  $1p1d1f$  for Sc–Zn.
  - **def2-TZVPP**: similar to def2-TZVP but with  $2p1d$  polarisation for H.

# Recommendations for Hartree–Fock and DFT

- **For routine work:** SV(P) or 4-31G\* or pc-1.
- **For accurate work:** def2-TZVP or 6-311G\* or pc-2.
- **For very accurate work:** def2-TZVPP or 6-311G\*\* or 6-311G(2df,2pd) or pc-3.

For some applications, diffuse functions must be added to obtain accurate (or even meaningful) results.

- A plus sign is added to the basis (6-311+G\*, 6-311+G(2df,2pd), etc.) when diffuse functions are added to the nonhydrogen atoms.
- Two plus signs are added when also the H atoms carry diffuse functions (6-311++G\*\*, 6-311++G(2df,2pd), etc.)
- Diffuse functions are for instance required for **anions**, **polar bonds**, **weak intermolecular interactions**, **Rydberg orbitals** and **excitation energies**.

## “def2” sets from the Turbomole basis-set library

- The “def2” basis sets form a system of **segmented contracted basis sets** for the **elements H–Rn** for different levels of flexibility/accuracy.
- The basis sets are denoted **def2-SV(P) to def2-QZVPP**. They are designed to give similar errors all across the periodic table for a given basis-set type.
- **At the Hartree–Fock and DFT levels**, the extended QZVPP basis yields atomisation energies (per atom) with an error  $< 1$  kJ/mol **with respect to the basis-set limit**. Other sets yield (in kJ/mol):

Basis	Hartree–Fock		DFT (BP-86)	
	mean	$\sigma$	mean	$\sigma$
def2-SV(P)	–14.5	15.3	–5.8	9.8
def2-SVP	–8.9	10.4	–2.0	8.8
def2-TZVP	–3.7	3.4	–2.6	2.1
def2-TZVPP	–2.0	2.2	–1.1	1.7
def2-QZVP	–0.2	0.6	–0.1	0.4

## Polarisation-consistent basis sets (pc- $n$ )

- Higher angular momentum functions are included based on energetical importance in Hartree–Fock calculations.

Atom	pc-0	pc-1	pc-2	pc-3	pc-4
C	$3s2p$	$3s2p1d$	$4s3p2d1f$	$6s5p4d2f1g$	$8s7p6d3f2g1h$
Si	$4s3p$	$4s3p1d$	$5s4p2d1f$	$6s5p4d2f1g$	$7s6p6d3f2g1h$

- Systematic basis sets (pc- $n$  with  $n = 0, 1, 2, 3, 4$ ) for which results converge monotonically to the Hartree–Fock limit. The Hartree–Fock energy obtained in a basis with angular momentum functions up to  $L$  is well described by

$$E_L = E_\infty + A(L + 1) \exp(-B\sqrt{L})$$

- The pc- $n$  basis sets are available for the elements H–Ar and can be **augmented with diffuse functions (aug-pc- $n$ )**.
- These basis sets use a **general contraction scheme**.

## Segmented versus general contractions

- Consider the **pc-1 basis for carbon ( $3s2p1d$ )**, which is of “double-zeta plus polarisation (DZP)” quality.

```
S-TYPE FUNCTIONS
  7   3
1252.600000000  0.005573400  0.000000000  0.000000000
 188.570000000  0.041492000  0.000277450  0.000000000
  42.839000000  0.182630000  0.002560200  0.000000000
  11.818000000  0.461180000  0.033485000  0.000000000
   3.556700000  0.449400000  0.087579000  0.000000000
   0.542580000  0.000000000 -0.537390000  0.000000000
   0.160580000  0.000000000  0.000000000  1.000000000

P-TYPE FUNCTIONS
  4   2
  9.142600000  0.044464000  0.000000000
  1.929800000  0.228860000  0.000000000
  0.525220000  0.512230000  0.000000000
  0.136080000  0.000000000  1.000000000

D-TYPE FUNCTIONS
  1   1
  0.800000000  1.000000000
```

- The input for a program that cannot handle general contractions must list an *s*-type CGTO built from the first 5 primitive GTOs, a second *s*-type CGTO built from the primitives 2–6, etc.

## Performance of various basis sets (test set)

- The table shows **mean absolute deviations** in  $r_e$  (pm),  $\omega_e$  (cm<sup>-1</sup>) and intensity (km/mol) **relative to the Hartree–Fock limit**.

Basis	Size	$\delta(r_e)$	$\delta(\omega_e)$	$\delta(\text{Intensity})$
STO-3G	9	5.5	142.3	22.8
pc-0	13	8.2	60.9	19.0
SVP	18	1.6	14.1	5.2
6-31G*	18	1.5	11.9	7.6
pc-1	18	1.8	11.8	5.4
cc-pVTZ	34	0.7	4.9	2.3
<b>pc-2</b>	<b>34</b>	<b>0.3</b>	<b>3.1</b>	<b>4.3</b>
cc-pVQZ	59	0.3	2.5	1.2
pc-3	64	< 0.1	0.3	0.9



## Performance of various basis sets for S<sub>2</sub>

The table shows deviations in  $D_e$  (kJ/mol),  $r_e$  (pm) and  $\omega_e$  (cm<sup>-1</sup>) relative to the ROHF Hartree–Fock limit.

Basis	Size	$\delta(D_e)$	$\delta(r_e)$	$\delta(\omega_e)$
pc-0	13	-220	20.3	-148
pc-1	18	-60	2.1	-17
pc-2	34	-19	0.5	-6
pc-3	64	-1	< 0.1	< 1
SV	13	-235	17.1	-178
def2-SVP	18	-47	1.7	-3
def2-TZVP	37	-7	0.2	-2
def2-TZVPP	42	-6	0.2	-2
def2-QZVP	70	-2	< 0.1	< 1

- No significant difference between basis sets of similar size.

## Relevance of basis-set errors

The table shows the Hartree–Fock value and various further contributions to the harmonic vibrational frequency of N<sub>2</sub>.

Contribution	$\omega_e / \text{cm}^{-1}$
<b>Near Hartree–Fock limit</b>	<b>2 730.5</b>
fc-CCSD(T) contribution (near basis-set limit)	–367.1
fc-CCSDTQ contribution (cc-pVTZ basis)	–9.1
fc-CCSDTQ5 contribution (cc-pVDZ basis)	–3.9
Core-correlation contribution	9.8
Relativistic correction (Dirac-Coulomb)	–0.8
Breit correction	–0.5
<b>Calculated value</b>	<b>2 358.9</b>
<b>Experimental value</b>	<b>2 358.6</b>

- Hartree–Fock theory tends to overestimate vibrational frequencies (by ca. 10%). Basis-set errors of the order of 1% are therefore fully acceptable.

## Concluding remarks on CGTO basis sets for SCF

- It is recommended to run applications in a “double-zeta plus polarisation”-type basis (DZP). For example,
  - **def2-SV(P)**: for H–Rn and programs that work efficiently with segmented contractions.
  - **pc-1**: for H–Ar and programs that work efficiently with general contractions.
- It is recommended to investigate basis-set effects by repeating the DZP calculation in a “triple-zeta plus polarisation”-type basis. For example,
  - **def2-TZVP**: for H–Rn and segmented contractions.
  - **pc-2**: for H–Ar and general contractions.
- Similar procedures apply to STOs (DZP and TZP in **ADF** and numerical AOs (DNP and TNP in **DMol<sup>3</sup>**).
- Need for diffuse functions must be checked.

## Atomic natural orbital (ANO) basis sets

- ANO basis sets are available for the atoms **H–Cm**.
- These are large **generally contracted basis sets** that are particularly useful in **electron-correlation** (also denoted post-Hartree–Fock) calculations.
- The contraction coefficients are the **natural orbitals** obtained from atomic post-Hartree–Fock calculations (*e.g.*, CISD, MCPF).
- Various states (also of ions) are averaged. Examples are:

	Primitives	CGTOs	Hartree–Fock range
H	$8s4p3d$	$6s4p3d$	$2s1p - 3s2p1d$
O	$14s9p4d3f$	$7s7p4d3f$	$3s2p1d - 4s3p2d1f$
S	$17s12p5d4f$	$7s7p5d4f$	$4s3p2d - 5s4p3d2f$
Zn	$21s15p10d6f4g$	$8s7p6d5f4g$	$5s3p2d - 6s5p4d3f2g$

- Can be systematically enlarged and BSSE is small.

# Correlation-consistent basis sets

- Analogous to ANOs, the aim of the correlation-consistent basis sets is to form **systematic sequences of basis sets of increasing size and accuracy**.
- Usually, the correlation-consistent basis sets have **generally contracted inner shells**.
- They are particularly useful in **electron-correlation calculations**.
- Polarisation functions are added in groups that contribute almost equally to the **correlation energy**.
- In their simplest form, they are denoted **cc-pVXZ**, with  $X = D, T, Q, 5, 6$ . “D” for “double-zeta”, “T” for “triple-zeta”, and so on.
- Diffuse functions can be added (**aug-cc-pVXZ**) as well as function to correlate the inner shells (**aug-cc-pCVXZ**, **aug-cc-pwCVXZ**).
- Basis sets such as **aug-cc-pV(X+d)Z**, **cc-pVXZ-PP** and **cc-pVXZ-F12** exist for selected atoms.

## MP2 correlation energies

- Valence-shell MP2 correlation energies of benzene. The basis-set limit is estimated as  $\Delta E_{\text{MP2}} = -1.0575 \pm 0.0005 E_h$ .

Basis	Size	$\Delta E_{\text{MP2}}/\%$	$\Delta E_{\text{MP2-F12}}/\%$
aug-cc-pVDZ	192	76.8	98.4
aug-cc-pVTZ	414	91.2	99.6
aug-cc-pVQZ	756	96.1	99.9
aug-cc-pV5Z	1242	97.9	100.0
aug-cc-pV6Z	1896	98.8	
def2-TZVP	222	88.2	99.1
def2-TZVPP	270	89.7	99.3
def2-QZVP	522	95.3	99.8

- Slater-type geminals of the form  $c_{ij}^{kl} \varphi_k(\mu) \varphi_l(\nu) \exp(-1.5 r_{\mu\nu})$  were used in the MP2-F12 method for each orbital pair  $ij$ .
- With standard MP2, extremely large basis sets are required to capture 98%.

## Special-purpose basis sets / ECPs

- Most basis sets have been optimised with respect to the total energy of an atom (or molecule).
- There exist basis sets that have been developed for the calculations of optical, electric or magnetic properties.
- Examples are the **Sadlej basis sets** for electric properties (dipole moment, polarisability) or the **IGLO basis sets** for NMR chemical shifts.
- In general, calculations of electric properties require **diffuse functions**. When those are added to all angular-momentum shells of a given basis, the prefix **aug** is added to the basis (**aug-cc-pVXZ**, **aug-pc-*n***).
- Sometimes, still more diffuse sets are required (d-aug- and t-aug- sets for polarisabilities and hyperpolarisabilities).
- Tight functions must be added when the wavefunction close to a nucleus is important (e.g., electric-field gradient).

## Auxiliary basis sets

- Thus far, we have discussed basis sets for the expansion of MOs and the electronic wavefunction.
- It is possible to save lots of computer time in DFT calculations when the electron density is expanded in a basis set,

$$\rho(\mathbf{r}) \approx \sum_P c_P^\rho \chi_P(\mathbf{r})$$

- In **Turbomole** nomenclature, such a basis is denoted **jbas** auxiliary basis.
- When also orbital products  $\varphi_i \chi_\kappa$  are expanded to build the exchange matrix, a **jkbas** auxiliary basis is needed.
- For the products  $\varphi_i \varphi_a$  that occur in MP2/CC2 theory, a **cbas** auxiliary basis is used.
- Again other auxiliary basis sets are used in explicitly-correlated methods (**cabs**).



## Closing remarks on basis sets

- For Hartree–Fock (and DFT), the ANO and correlation-consistent basis sets have no advantages over SVP/pc-1 respectively TZVPP/pc-2.
- Basis sets of at least quadruple-zeta quality are required for electron-correlation treatments.
- For very accurate electron-correlation calculations, basis sets larger than cc-pVQZ etc. are needed, in conjunction with basis-set extrapolation.
- Experience with explicitly-correlated theory using Slater-type geminals (two-particle basis functions) indicates that basis sets beyond triple-zeta quality are no longer needed.
- **Recipes:**
  - def2-SV(P) for DFT, check results with def2-TZVP.
  - def2-TZVPP or cc-pVTZ-F12 for MP2-F12, CCSD-F12 etc., check results with def2-QZVPP or cc-pVQZ-F12.