Basis sets
What are they?

- Mathematically, a basis set is a collection of vectors which spans (defines) a space in which a problem is solved;
- i, j, k define a Cartesian, 3D linear vector space;
- The Fourier series constitutes an orthogonal basis set for periodic functions;
- In quantum chemistry, the “basis set” often refers to any set of (usually nonorthogonal) one-particle functions used to build molecular orbitals.

Some of this is from Sherrill’s computational chemistry course at GA Tech.
What are they?

In Quantum Chemistry, various types are used:

- LCAO-MO approximation: MO’s built from AO’s
- An “orbital” is a one-electron function
- AO’s represented by atom-centered Gaussians in most quantum chemistry programs - why Gaussians? (GTO’s)
- Some older programs used Slater functions (STO’s)
- Newer basis sets include correlation-consistence and set size convergence
- Physicists and materials chemists like plane wave basis sets
Slater-type orbitals

- Atomic Orbitals (because they are centered on atoms)
  \[ \psi_{\zeta,n,l,m}(r, \theta, \phi) = NY_{l,m}(\theta, \phi)r^{n-1}e^{-\zeta r} \]
- \(\zeta\) controls the width of the orbital (large \(\zeta\) gives a tight function, small \(\zeta\) gives a diffuse function) - see the exponential
- H-like, but lack radial nodes - linear combinations to get actual wave functions
- Correct short- and long-range behaviour
- Not used very much - Gaussians have better integration behaviour. Some revival happening now.
- Sometimes written as linear combinations of Gaussians (STO-3G).
Gaussian-type orbitals

\[ \psi_{\zeta,n,l,m}(r, \theta, \phi) = NY_{l,m}(\theta, \phi)r^{2n-2-l}e^{-\zeta r^2} \]

\[ \psi_{\zeta,l_x,l_y,l_z}(x, y, z) = Nx^{l_x}y^{l_y}z^{l_z}e^{-\zeta r^2} \]

- \( l_x, l_y, l_z \) determine the type of orbital
- Not H-like anymore
- Much easier to compute - products of Gaussians are Gaussians centered between atom centres - four-centre integrals can be reduced to single centre.
- They have poor performance near the nucleus and at long distances - more GTOs necessary to achieve a given accuracy than STOs.
Contracted GTOs

- The GTO inaccuracies can be solved to some extent by forming STOs by linear combination of GTO’s.
- Unrealistic GTO behaviour near nucleus means that a lot of time during calculation spent optimizing core representation.
- Contracted Gaussian-type orbitals: STO-nG basis.

\[ \psi_{\zeta,l_x,l_y,l_z}^{CGTO}(x, y, z) = N \sum_{i=1}^{n} c_i x^{l_x} y^{l_y} z^{l_z} e^{-\zeta_i r^2} \]

coefficients \( c_i \) are calculated with the basis set, not obtained variationally during the calculation - significant speedup.

Figure 3.3 Comparison of the quality of the least-squares fit of a 1s Slater function (\( \zeta = 1.0 \)) obtained at the STO-1G, STO-2G, and STO-3G levels.

Figure from Szabo and Ostlund, Modern Quantum Chemistry.
Minimal and expanded basis sets

- **Minimal** basis set: one function (STO, GTO, CGTO) per AO
  - H atom: 1s AO, one basis set function
  - C atom: 5 AO (1s, 2s, 3x2p) - five basis functions
- **Double zeta (DZ)**: two basis functions (different $\zeta$) per AO
  - allows treatment of spatially different bonds at the same atom.
  - C atom: 2 basis functions per AO, so 10 in total
- **Triple zeta (TZ)**: three basis functions per AO.
- **Quadruple (QZ)**, quintuple (pentuple - PZ, 5Z) zeta ..... Having different size exponentials allows the orbital to change size with perturbations from approaching atoms
- **Split valence**: only one basis function for core AOs, and more basis functions for valence AOs
  - C atom valence double zeta (VDZ): 9 basis functions - why?
### Minimal basis set: STO-3G

For C:

```plaintext
BASIS="STO-3G"

<table>
<thead>
<tr>
<th>Atom</th>
<th>n</th>
<th>1.00</th>
<th>( \zeta )</th>
<th>( c_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>3</td>
<td>1.00</td>
<td>71.6168370</td>
<td>0.154328970</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13.0450960</td>
<td>0.535328140</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.53051220</td>
<td>0.444634540</td>
</tr>
<tr>
<td>S</td>
<td>3</td>
<td>1.00</td>
<td>2.94124940</td>
<td>-0.999672300E-01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.683483100</td>
<td>0.399512830</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.222289900</td>
<td>0.700115470</td>
</tr>
<tr>
<td>P</td>
<td>3</td>
<td>1.00</td>
<td>2.94124940</td>
<td>0.155916270</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.683483100</td>
<td>0.607683720</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.222289900</td>
<td>0.391957390</td>
</tr>
</tbody>
</table>
```

The problem with minimal basis is its lack of flexibility. When bonds are formed, orbitals usually contract (or expand), and the minimal basis cannot adequately describe such changes in orbitals. The solution is to represent each orbital with 2 and more functions. That's the idea behind \( n \)-tuple-\( \zeta \) (double-zeta, triple-zeta, quadruple-zeta, etc.) basis sets.
Expanded basis sets

- **Double zeta (DZ):** two basis functions per AO
  - allow treatment of spatially different bonds at the same atom.

  - C-H $\sigma$-bond: H 1s orbital and C $2p_z$.
  - CN $\pi$-bond: C and N $2p_x$ (and $2p_y$) AOs.
  - $\pi$-bond is more diffuse: optimal $\zeta$ for $p_x$ ($p_y$) is smaller than for more localized $p_z$.

DZ describe charge distribution in both parts of the molecule: optimized AO coefficient (in MO expansion) of ‘tighter’ inner $p_z$ function on carbon will be larger in the C-H bond. More diffuse outer $p_x$ and $p_y$ functions will have larger AO coefficients in the $\pi$-bond.
Polarization

- As other atoms approach, an atom’s orbitals might want to shift to one side or the other (polarization).
- An s orbital can polarize in one direction if it’s mixed with a p orbital. p orbitals can polarize if mixed with d orbitals.
- In general, to polarize a basis function with angular momentum $l$, one mixes it with basis functions of angular momentum $l + 1$.
- This gives polarized double-zeta, or double-zeta plus polarization (DZP) basis sets, etc.
- Polarization functions particularly important in achieving basis set convergence with correlated methods.
Polarization

H-C σ-bond:

- Electron distribution along CH bond is different from the perpendicular direction. H 1s orbital does not describe this behavior well.

- If p-functions are added to H, then $p_z$ AO can improve the description of the CH bond.

- p-functions induce a polarization of s-orbitals.
In Gaussian notation, the C-atom DZP basis set is:

```
BASIS="DZP (Dunning)"
C 0
S 6  1.00
  4232.61000  0.202900000E-02
  634.882000  0.155350000E-01
  146.097000  0.754110000E-01
  42.4974000  0.257121000
  14.1892000  0.596555000
  1.96660000  0.242517000
S 1  1.00
  5.14770000  1.00000000
S 1  1.00
  0.496200000  1.00000000
S 1  1.00
  0.153300000  1.00000000
P 4  1.00
  18.1557000  0.185340000E-01
  3.98640000  0.115442000
  1.14290000  0.386206000
  0.35940000  0.640089000
P 1  1.00
  0.11460000  1.00000000
D 1  1.00
  0.75000000  1.00000000
```
Diffuse functions

- Diffuse functions have small $\zeta$ exponents - the electron is held far away from the nucleus.
- Necessary for anions, Rydberg states, very electronegative atoms (fluorine) with a lot of electron density.
- Necessary for accurate polarizabilities or binding energies of van der Waals complexes (bound by dispersion).
- It is very bad to do computations on anions without using diffuse functions; your results could change completely!
Pople basis sets

- Contracted basis sets developed by the late Nobel Laureate, John Pople, and popularized by the Gaussian set of programs.
- STO-\(n\)G are minimal basis sets in which each AO is represented by \(n\) Gaussians, chosen to mimic the behavior of a STO.
- 3-21G is a split valence basis set where core orbitals are a contraction on 3 GTO’s, the inner part of the valence AOs is a contraction of 2 GTO’s and the outer part is given by 1 GTO.
- Pople’s split-valence double-zeta basis set is called 6-31G; the core orbital is a CGTO made of 6 Gaussians, and the valence is described by two orbitals - one CGTO made of 3 Gaussians, and one single Gaussian.
Pople basis sets

- 6-31G* [or 6-31G(d)] is 6-31G with added d polarization functions on non-hydrogen atoms; 6-31G** [or 6-31G(d,p)] is 6-31G* plus p polarization functions for hydrogen
- 6-311G is a split-valence triple-zeta basis; it adds one GTO to 6-31G
- 6-31+G is 6-31G plus diffuse s and p functions for non-hydrogen atoms; 6-31++G has diffuse functions for hydrogen also
- The main disadvantage of Pople basis sets is that they do not converge towards a basis set limit with increasing size.
- This is partly because no higher-l functions are included, and because quite different GTOs are added to larger set based on physical properties rather than in a consistent fashion.
## Pople basis sets

In Gaussian notation, the C-atom 6-31G basis set is:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>6</td>
<td>1.00</td>
<td>ζ</td>
<td>3047.52490</td>
<td>0.183470000E-02</td>
<td>457.369510</td>
<td>0.140373000E-01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ci</td>
<td>103.948690</td>
<td>0.688426000E-01</td>
<td>29.2101550</td>
<td>0.232184400</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.28666300</td>
<td>0.467941300</td>
<td>3.16392700</td>
<td>0.362312000</td>
</tr>
<tr>
<td>SP</td>
<td>3</td>
<td>1.00</td>
<td></td>
<td>7.86827240</td>
<td>-0.119332400</td>
<td>1.88128850</td>
<td>-0.160854200</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.544249300</td>
<td>1.14345640</td>
<td>0.689991000E-01</td>
<td></td>
</tr>
<tr>
<td>SP</td>
<td>1</td>
<td>1.00</td>
<td></td>
<td>0.168714400</td>
<td>1.000000000</td>
<td>0.316424000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.744308300</td>
<td>1.000000000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Not as good as Dunning DZ!
What is this?

C 0
S 6 1.00
  3047.52490  0.183470000E-02
  457.369510  0.140373000E-01
  103.948690  0.688426000E-01
  29.2101550  0.232184400
  9.28666300  0.467941300
  3.16392700  0.362312000

SP 3 1.00
  7.86827240 -0.119332400  0.689991000E-01
  1.88128850 -0.160854200  0.316424000
  0.544249300  1.14345640  0.744308300

SP 1 1.00
  0.168714400  1.00000000  1.00000000

D 1 1.0
  0.800000000  1.000000000
Correlation-consistent (Dunning) basis sets

- Basis sets optimized at HF level may not be appropriate for correlated methods.
- cc basis sets are optimized using CISD wavefns.
- Geared towards recovering the correlation energy of valence electrons.
- Size increases by adding shells (or sets) of fns that contribute similar amounts of correlation energy.
- Designed to converge smoothly towards a basis set limit
- aug-cc basis sets have diffuse functions added
- Some include core correlation as well
- cc-pVXZ means a Dunning correlation-consistent, polarized valence, X-zeta basis; X=D,T,Q,5,6,7. cc-pVDZ for C atom consists of 3s2p1d. cc-pVTZ would be 4s3p2d1f. cc-pVQZ would be 5s4p3d2f1g.
Plane waves

- Instead of using LCAO, use a basis of functions appropriate for the entire system;
- Particularly useful for periodic materials;
- In metals, for ex, it is appropriate to treat valence electrons as free electrons - use a basis set of free electron wavefunctions:

\[ \psi_k(r) = e^{i\mathbf{k} \cdot \mathbf{r}} \]

- Wave vector \( \mathbf{k} \) behaves just like \( \zeta \). Permissible values related to the unit cell translational vector: \( \mathbf{k} \cdot \mathbf{t} = 2\pi m \). Basis set size depends on unit cell size.
- For core electrons, need functions with rapid oscillations - high \( k \), but usually ECP’s are used.