Summer Lectures in Electronic Structure Theory: Basis Sets and Extrapolation
July 6, 2010
Two limitations to solving the Schrödinger Equation
Two limitations to solving the Schrödinger Equation

Arrange Four Electrons in Six Orbitals

1. Ground Configuration: HF
Two limitations to solving the Schrödinger Equation

<table>
<thead>
<tr>
<th>1</th>
<th>Ground Configuration</th>
<th>HF</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Single Excitations</td>
<td>S</td>
</tr>
</tbody>
</table>
Two limitations to solving the Schrödinger Equation

Arrange Four Electrons in Six Orbitals

1  Ground Configuration       HF
32 Single Excitations        S
168 Double Excitations       D
Two limitations to solving the Schrödinger Equation

### Arrange Four Electrons in Six Orbitals

<table>
<thead>
<tr>
<th></th>
<th>Ground Configuration</th>
<th>HF</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>Single Excitations</td>
<td>S</td>
</tr>
<tr>
<td>168</td>
<td>Double Excitations</td>
<td>D</td>
</tr>
<tr>
<td>224</td>
<td>Triple Excitations</td>
<td>T</td>
</tr>
</tbody>
</table>
Two limitations to solving the Schrödinger Equation

Arrange Four Electrons in Six Orbitals

<table>
<thead>
<tr>
<th></th>
<th>Ground Configuration</th>
<th>HF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Single Excitations</td>
<td>S</td>
</tr>
<tr>
<td>32</td>
<td>Double Excitations</td>
<td>D</td>
</tr>
<tr>
<td>168</td>
<td>Triple Excitations</td>
<td>T</td>
</tr>
<tr>
<td>224</td>
<td>Quadruple Excitations</td>
<td>Q</td>
</tr>
<tr>
<td>495</td>
<td>Slater Determinants</td>
<td></td>
</tr>
</tbody>
</table>
Two limitations to solving the Schrödinger Equation

Arrange Four Electrons in Six Orbitals

<table>
<thead>
<tr>
<th>1</th>
<th>Ground Configuration</th>
<th>HF</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>Single Excitations</td>
<td>S</td>
</tr>
<tr>
<td>168</td>
<td>Double Excitations</td>
<td>D</td>
</tr>
<tr>
<td>224</td>
<td>Triple Excitations</td>
<td>T</td>
</tr>
<tr>
<td>70</td>
<td>Quadruple Excitations</td>
<td>Q</td>
</tr>
<tr>
<td>495</td>
<td>Slater Determinants</td>
<td></td>
</tr>
</tbody>
</table>

Multi-Electron Basis

Quality of a calculation depends upon the quality of basis sets

- Multi-electron basis: Slater determinants available to expand the wavefunction
- One-electron basis: atomic orbitals available to expand the molecular orbitals
- Full multi-electron basis recovers all correlation within a given one-electron basis
- Full one-electron basis recovers the SCF energy without limitation on the shape of MO
Quality of a calculation depends upon the quality of basis sets
- Multi-electron basis: Slater determinants available to expand the wavefunction
- One-electron basis: atomic orbitals available to expand the molecular orbitals
- Full multi-electron basis recovers all correlation within a given one-electron basis
- Full one-electron basis recovers the SCF energy without limitation on the shape of MO

Two limitations to solving the Schrödinger Equation
1. HF Limit
2. Full CI
Quality of a calculation depends upon the quality of basis sets

- Multi-electron basis: Slater determinants available to expand the wavefunction
- One-electron basis: atomic orbitals available to expand the molecular orbitals
- Full multi-electron basis recovers all correlation within a given one-electron basis
- Full one-electron basis recovers the SCF energy without limitation on the shape of MO
Basis set in quantum chemistry

- Generically, a basis set is a collection of vectors which spans (defines) a space in which a problem is solved.
- $\hat{i}, \hat{j}, \hat{k}$ define a Cartesian, 3D linear vector space.
- In quantum chemistry, the “basis set” usually refers to the set of (nonorthogonal) one-particle functions used to build molecular orbitals.
- 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)
- Physicists like plane wave basis sets.
Slater-type orbitals (STOs)

\[ \phi_{abc}^{STO}(x, y, z) = N x^a y^b z^c e^{-\zeta r} \]

- \( N \) is a normalization constant
- \( a, b, c \) control angular momentum, \( L = a + b + c \)
- \( \zeta \) (zeta) controls the width of the orbital (large \( \zeta \) gives tight function, small \( \zeta \) gives diffuse function)
- These are H-atom-like, at least for 1s; however, they lack radial nodes and are not pure spherical harmonics (how to get 2s or 2p, then?)
- Correct short-range and long-range behavior
Gaussian-type orbitals (GTOs)

$$\phi_{abc}^{GTO}(x, y, z) = N x^a y^b z^c e^{-\zeta r^2}$$

- Again, $a, b, c$ control angular momentum, $L = a + b + c$
- Again, $\zeta$ controls width of orbital
- No longer H-atom-like, even for 1s
- Much easier to compute (Gaussian product theorem)
- Almost universally used by quantum chemists

Radial Distribution
- too flat at r near 0
- falls off too quickly at large r
Contracted Gaussian-type orbitals (CGTOs)

- **Problem:** STO’s are more accurate, but it takes longer to compute integrals using them.

- **Solution:** Use a linear combination of enough GTO’s to mimic an STO.

- **Unfortunate:** A combination of $n$ Gaussians to mimic an STO is often called an “STO-nG” basis, even though it is made of CGTO’s...

\[
\phi_{\text{CGTO}}^{abc}(x, y, z) = N \sum_{i=1}^{n} c_i x^a y^b z^c e^{-\zeta_i r^2}
\]

(1)
Contracted Gaussian-type orbitals (CGTOs), cont.

Cost for HF scales as $N^4$

Optimizing all coefficients in the Gaussian expansion would be more accurate but costly

CGTOs permit boost in computational efficiency by limiting number of coefficients in Gaussian expansion

CGTOs used most commonly for inner orbitals & GTO for outer shells

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.
Designations of basis set size

- **Minimal**: one basis function (STO, GTO, or CGTO) for each atomic orbital in each atom
- **Double-zeta (DZ)**: two basis functions for each AO
- **Triple-zeta (TZ)**: three basis functions for each AO
- etc., for **quadruple-zeta (QZ)**, 5Z, 6Z, ...

The presence of different-sized functions allows the orbital to get bigger or smaller when other atoms approach it, adds flexibility to adequately describe anisotropic electron distribution in molecules.

**Adding**: add’l fns of varying sizes with valence ang. mom.
Designations of basis set size

- **Minimal**: one basis function (STO, GTO, or CGTO) for each atomic orbital in each atom
- **Double-zeta (DZ)**: two basis functions for each AO
- **Triple-zeta (TZ)**: three basis functions for each AO
- etc., for **quadruple-zeta (QZ), 5Z, 6Z, ...**

The presence of different-sized functions allows the orbital to get bigger or smaller when other atoms approach it, adds flexibility to adequately describe anisotropic electron distribution in molecules.

Virtual


Occupied


**Adding**: add’l fns of varying sizes with valence ang. mom.
Designations of basis set size

- **Minimal**: one basis function (STO, GTO, or CGTO) for each atomic orbital in each atom
- **Double-zeta (DZ)**: two basis functions for each AO
- **Triple-zeta (TZ)**: three basis functions for each AO
- etc., for quadruple-zeta (QZ), 5Z, 6Z, ...

The presence of different-sized functions allows the orbital to get bigger or smaller when other atoms approach it, adds flexibility to adequately describe anisotropic electron distribution in molecules

\[
\begin{array}{c}
\text{Virtual} \\
\begin{array}{c}
\_ \\
\_ \\
\_ \\
\_ \\
\_ \\
\end{array} \\
\text{Double-}\zeta\text{ Basis Set (4s2p)} \\
2\cdot(2s1p) = 10 \\
\text{Occupied} \\
\begin{array}{c}
\frac{1}{1} \\
\frac{1}{1} \\
\end{array}
\end{array}
\]

**Adding**: add’l fns of varying sizes with valence ang. mom.
Designations of basis set size

- **Minimal**: one basis function (STO, GTO, or CGTO) for each atomic orbital in each atom
- **Double-zeta (DZ)**: two basis functions for each AO
- **Triple-zeta (TZ)**: three basis functions for each AO
- etc., for quadruple-zeta (QZ), 5Z, 6Z, ...

- The presence of different-sized functions allows the orbital to get bigger or smaller when other atoms approach it, adds flexibility to adequately describe anisotropic electron distribution in molecules

Adding: add’l fns of varying sizes with valence ang. mom.
Designations of basis set size

- **Minimal**: one basis function (STO, GTO, or CGTO) for each atomic orbital in each atom
- **Double-zeta (DZ)**: two basis functions for each AO
- **Triple-zeta (TZ)**: three basis functions for each AO
- etc., for **quadruple-zeta (QZ), 5Z, 6Z, ...**

The presence of different-sized functions allows the orbital to get bigger or smaller when other atoms approach it, adds flexibility to adequately describe anisotropic electron distribution in molecules.

Adding: add’l fns of varying sizes with valence ang. mom.
Split-valence basis sets

- A split-valence basis uses only one basis function for each core AO and a large basis for the valence AOs.
- Flexibility more important for valence orbitals, which are chemically important.
Split-valence basis sets

- A split-valence basis uses only one basis function for each core AO and a large basis for the valence AOs.
- Flexibility more important for valence orbitals, which are chemically important.

Virtual

Occupied

Minimal Basis Set (2s1p)

Virtual

Occupied

Minimal Basis Set (2s1p)

Virtual

Occupied

Minimal Basis Set (2s1p)

1s + 1·(1s1p) = 5
Split-valence basis sets

- A split-valence basis uses only one basis function for each core AO and a large basis for the valence AOs.
- Flexibility more important for valence orbitals, which are chemically important.

### Double-ζ Basis Set (4s2p)
- Virtual: \( \frac{1}{4} \) \( \frac{1}{4} \)
- Occupied: \( \frac{1}{2} \) \( \frac{1}{2} \)

\[ 2 \cdot (2s1p) = 10 \]

### Double-ζ Split-Valence Basis Set (3s2p)
- Virtual: \( \frac{1}{3} \) \( \frac{1}{3} \)
- Occupied: \( \frac{1}{2} \) \( \frac{1}{2} \)

\[ 1s + 2 \cdot (1s1p) = 9 \]
Split-valence basis sets

- A split-valence basis uses only one basis function for each core AO and a large basis for the valence AOs.
- Flexibility more important for valence orbitals, which are chemically important.

Virtual

\[
\text{Triple-\(\zeta\) Basis Set (6s3p)}
\]

\[
3 \cdot (2s1p) = 15
\]

Occupied

Virtual

\[
\text{Triple-\(\zeta\) Split-Valence Basis Set (4s3p)}
\]

\[
1s + 3 \cdot (1s1p) = 13
\]

Occupied
Split-valence basis sets

- A split-valence basis uses only one basis function for each core AO and a large basis for the valence AOs.
- Flexibility more important for valence orbitals, which are chemically important.

### Quadruple-ζ Basis Set (8s4p)
4·(2s1p) = 20

### Quadruple-ζ Split-Valence Basis Set (5s4p)
1s + 4·(1s1p) = 17
Pitfall: Occupancy

- QC programs which use symmetry can yield valuable efficiency gains for symmetry molecules (prefactor 1/point group order)
- Since there is no coupling between each subgroup (e.g., $a_1$, $b_2$, $b_1$, $a_2$) of orbitals, the program must guess an occupancy
- If guess is incorrect, SCF energies will never match ground-state values from other programs, as you have effectively calculated an excited state. If you’re trying to do a frequency analysis, you may have imaginary frequencies from optimizing to a non-global minimum
- This pitfall is more likely with larger extended basis sets
- Beware different ordering of symmetry subgroups between QC programs, esp. $C_{2v}$

Pitfall: for extended basis sets, when using symmetry, a QC program may allocate electrons to a non-ground state configuration of orbitals.

Solution: (1) Specify the correct allocation by transferring from a QC program you trust. (2) Start the SCF with a small (STO-3G, 6-31G) basis, then direct the program to project into a larger (desired) basis set.
Polarization Functions

- As other atoms approach, an atom’s orbitals might want to shift to one side or the other (polarization). Polarization of an s orbital in one direction can be represented by mixing it with a p orbital.
- p orbitals can polarize if mixed with d orbitals.
- In general, to polarize a basis function with angular momentum \( l \), mix it with basis functions of angular momentum \( l + 1 \).
- This gives “polarized double-zeta” or “double-zeta plus polarization” basis sets.

Adding: add’d fns with higher-than-valence ang. mom.
Pitfall: Counting polarization functions

- We know there should be 5 \(d\) functions; these are called “pure angular momentum” functions (even though they aren’t really eigenfunctions of the AM operator)

\[
d_{x^2-y^2}, \quad d_{z^2}, \quad d_{xy}, \quad d_{xz}, \quad d_{yz}
\]

- Computers would prefer to work with 6 \(d\) functions; these are called “6 Cartesian \(d\) functions”

\[
d_{x^2}, \quad d_{y^2}, \quad d_{z^2}, \quad d_{xy}, \quad d_{xz}, \quad d_{yz}
\]

\(d_{x^2} + d_{y^2} + d_{z^2}\) looks like a \(s\) orbital

- Similar answers are obtained using 5 or 6 \(d\) functions

- For \(f\) functions, it’s 7 versus 10 \(f\) functions

- **Pitfall:** a basis set is designed by its author to use one or the other \((5/6 d \& 7/10 f)\) schemes; whereas, a QC program may also default to one or the other. If a QC program is aware of a basis set’s predilection, it may forget if the basis set is specified with variations or in combination. You can very easily get wrong answers that look about right

- **Solution:** (1) Pay attention to the output- check that the number of orbitals and the resultant energy is the same as for a QC program you are confident is handling the basis set correctly. (2) If the output is wrong, read the manual and find out how to specify the \(d\) and \(f\) functions.
Diffuse Functions

- When electron density exists far from nuclear centers, normal basis sets are inadequate
- Diffuse functions have small zeta exponents to hold the electron 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!

Adding: add’l fns with valence ang. mom. and smaller-than-valence exponents
Pitfall: Linear Dependency

Large basis sets (particularly, large basis sets with diffuse functions) are prone to “near linear dependency,” where the description of space spanned by the basis functions is overcomplete.

With a loss in uniqueness in the molecular orbital coefficients, SCF may be slow to converge or behave erratically.

Usually diagnosed from eigenvalues of the overlap matrix; thereafter near-degeneracies can be projected out according to a cutoff value.

Pitfall: (1) SCF is having difficulty converging and the output warns of linear dependency. (2) You are dealing with a potential energy curve or systems along a reaction path and linear dependency is being treated differently in each.

Solution: (1) Tighten up the cutoff value ($<10^{-5}$) for projecting out LD. (2) Adjust cutoff or turn off projection until same number of orbitals dropped along curve and potential energy surface is smooth.
A few basis sets

6-31G, 6-31G Blaudeau, 6-31+G, 6-31G*, 6-31G***, 6-31G Blaudeau, 6-31+G*, 6-31+G**,
6-31G(3df,3pd), 6-31LG, 6-31LG*, 6-31LG**, 6-311+G*, 6-31+G**, 6-311+G**(2d,2p),
6-311G(2df,2pd), 6-311+G(3df,3pd), MINI (Huzinaga), MINI (Scaled), MIDI (Huzinaga),
MIDIIV, SV (Dunning-Hay), SVP+Diffuse (Dunning-Hay), DZ (Dunning), DZP (Dunning),
DZP+Diffuse (Dunning), TZ (Dunning), Chipman DZP+Diffuse, cc-pVDZ, cc-pVTZ,
cc-pVQZ, cc-pV5Z, cc-pV6Z, vP6Z, pV7Z, cc-pVDZ (sp, opt), cc-pVTZ (sp, opt),
cc-pVQZ(sp, opt), cc-pCVTZ, cc-pCVQZ, cc-pCV5Z, aug-cc-pVDZ, aug-cc-pVTZ,
aug-cc-pVQZ, aug-cc-pV5Z, aug-cc-pV6Z, aug-cc-pV7Z, aug-cc-pCVTZ, aug-cc-pCVQZ,
aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z,
aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z,
aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z,
aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z,
aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z,
aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z,
aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z,
aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z,
aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z,
aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z,
aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z,
aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z,
aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z,
aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z,
aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z,
aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z,
aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z,
aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z,
aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z,
aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z,
aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z,
aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z,
aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z,
aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z,
aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z,
aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z,
aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z,
aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z,
aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z,
aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z,
aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z,
aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z,
aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z,
aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z,
aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z,
aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z,
aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z,
aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z,
aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z,
aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z,
aug-cc-pCV6Z, aug-cc-pCV7Z, aug-cc-pCVQZ, aug-cc-pCV5Z, aug-cc-pCV6Z,
aug-cc-pCV7Z, aug-cc-pCVQZ. 

A partial listing by a practitioner irritated by basis set acronyms
Meet the basis sets: Pople-style bases

- Developed by the late Nobel Laureate, John Pople, and popularized by the Gaussian set of programs
- Basis set notation looks like: k-nlm++G** or k-nlm++G(idf,jpd)
- Core/Valence functions
  - k primitive GTO for core electrons
  - n primitive GTO for inner valence orbitals
  - l primitive GTO for medium valence orbitals
  - m primitive GTO for outer valence orbitals
- Polarization functions
  - * indicates one set d polarization functions added to heavy atoms (non-H), alt. (d)
  - ** indicates one set d polarization functions added to heavy atoms and one set p polarization functions added to H atom, alt. (d,p)
  - idf indicates i set d and one set f polarization functions added to heavy atoms
  - idf,jpd indicates i set d and one set f polarization functions added to heavy atoms and j set p and one set d polarization functions added to H atom
- Diffuse functions
  - + indicates one set p diffuse functions added to heavy atoms (non-H)
  - ++ indicates one set p diffuse functions added to heavy atoms and 1 s diffuse function added to H atom

Examples:
- e.g., 3-21G, 6-31G, 6-311G
- e.g., 6-31G*, 6-311G(d,p)
- e.g., 6-311++G
Meet the basis sets: Dunning-style bases

- Pople-style basis sets generally optimized from HF calculations on atoms and small molecules
- Thom Dunning (UT, ORNL) utilized higher order correlated methods, CISD, hence “correlation-consistent”
- Designed to systematically converge the correlation energy to the basis set limit
- Basis set notation looks like: (aug)-cc-p(C)VXZ, X=D,T,Q,5,6,7
- Valence & Polarization functions
  - cc-pVXZ is a polarized valence X-zeta basis
  - functions are added in shells such that all orbitals with a similar contribution to the energy are added together. For C, cc-pVDZ is 3s2p1d, cc-pVTZ 4s3p2d1f, cc-pVQZ is 5s4p3d2f1g
- Diffuse functions
  - prefix “aug” indicates one set of diffuse functions added for every angular momentum present in the basis. For C, aug-cc-pVDZ has diffuse s,p,d
  - extra diffuse functions available with “d-aug”
- Core functions
  - letter “C” indicates presence of core correlation functions
    - e.g., cc-pCVTZ
Meet the basis sets: Others

Atomic Natural Orbitals (ANO)

- J. Amlof and P. R. Taylor at NASA Ames
- ANO designed to reproduce the natural orbitals for correlated (CISD) calculations on atoms
- Generally contracted- more complicated scheme to implement
- Well-suited for harmonic and anharmonic force fields of correlated methods
- Very large basis sets (consequently very expensive) but thorough and numerous levels of truncation are available

Properties Basis Sets

- PBS: polarized basis set of Sadlej, double-zeta + diffuse, developed for calculation of electrical properties, good for dipole moments, polarizabilities, excited states
- WMR: generally contracted basis sets by Widmark, Malmqvist, and Roos for atomic and molecular properties, rather large
Effective Core Potentials / Pseudopotentials

- Core orbitals generally not affected by changes in chemical bonding but require many basis functions to describe accurately

- Treat core electrons as averaged potentials rather than actual particles—pseudopotentials

- Advantages of greater efficiency since calculation not dominated by unnecessary integrals

- Advantage of incorporating relativistic effects

- Developed by removing core-dominated basis functions, then reoptimizing the remaining basis functions in the presence of the pseudopotential

- Potential is linear combination of Gaussians chosen to model the core electrons, orthogonal to valence electrons in basis functions

- Very important for heavy atoms, especially transition metals
Acquiring basis sets – the EMSL library

- Locate website by searching “EMSL Gaussian”
- Search by basis set name
- Search by element
- Available for many QC programs
- Stores basis sets and pseudopotentials
Acquiring basis sets— the EMSL library

A basis set of Contracted Gaussian-Type Orbitals (CGTO’s) needs to specify the exponents ($\zeta_i$’s) and the contraction coefficients ($c_i$’s). These are given below in the format used by the Gaussian program (exponents first).
Acquiring basis sets– the EMSL library

BASIS SET EXCHANGE v1.2.2

<table>
<thead>
<tr>
<th>Elements</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na – Mg</td>
<td>D.E. Woon and T.H. Dunning, Jr. (to be published)</td>
</tr>
</tbody>
</table>

```plaintext
****
C  0
S  8  1.00
   6665.00000000  0.0006920
   1000.00000000  0.0053290
     228.00000000  0.0270770
     64.71000000  0.1017180
     21.06000000  0.2747400
      7.49500000  0.4485640
      2.79700000  0.2850740
      0.52150000  0.0152040
S  8  1.00
   6665.00000000 -0.0001460
   1000.00000000 -0.0011540
     228.00000000 -0.0057250
     64.71000000 -0.0233120
     21.06000000 -0.0639550
      7.49500000 -0.1499810
      2.79700000 -0.1272620
      0.52150000  0.5445290
S  1  1.00
   0.15960000  1.0000000
P  3  1.00
   9.43900000  0.0381090
   2.00200000  0.2094800
   0.54560000  0.5085570
P  1  1.00
   0.15170000  1.0000000
D  1  1.00
   0.55000000  1.0000000
****
```
Basis set superposition error (BSSE)

“In a system comprising interacting fragments A and B, the fact that in practice the basis sets on A and B are incomplete means that the fragment energy of A will necessarily be improved by the basis functions on B (and vice versa), irrespective of whether there is any genuine binding interaction in the compound system or not.”

Correct with Counterpoise method of Boys and Bernardi (use ghost atoms to obtain monomer energies in dimer basis)

\[ E_{unCP} = E_{dimer}^{AB} - E_{mono}^A - E_{mono}^B \]

\[ E_{CP} = E_{dimer}^{AB} - E_{mono}^{AB} - E_{mono}^{AB} \]
Properties and basis sets

Convergence of Bond Lengths for BH, CH+, and NH

Convergence of Vibrational Frequencies for BH, CH+, and NH

*Figure 1.* Convergence of CCSD(T) \( r_e \) and \( \omega_e \) toward the CBS limit derived for valence-only (cc-pVNZ) basis sets.

Balancing methods and basis sets

<table>
<thead>
<tr>
<th>Method</th>
<th>DZ</th>
<th>TZ</th>
<th>QZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF</td>
<td>cc-pVDZ HF</td>
<td>cc-pVTZ HF</td>
<td>cc-pVQZ HF</td>
</tr>
<tr>
<td>MP2</td>
<td>cc-pVDZ MP2</td>
<td>cc-pVTZ MP2</td>
<td>cc-pVQZ MP2</td>
</tr>
<tr>
<td>MP3</td>
<td>cc-pVDZ MP3</td>
<td>cc-pVTZ MP3</td>
<td>cc-pVQZ MP3</td>
</tr>
<tr>
<td>MP4</td>
<td>cc-pVDZ MP4</td>
<td>cc-pVTZ MP4</td>
<td>cc-pVQZ MP4</td>
</tr>
<tr>
<td>CCSD</td>
<td>cc-pVDZ CCSD</td>
<td>cc-pVTZ CCSD</td>
<td>cc-pVQZ CCSD</td>
</tr>
<tr>
<td>CCSD(T)</td>
<td>cc-pVDZ CCSD(T)</td>
<td>cc-pVTZ CCSD(T)</td>
<td>cc-pVQZ CCSD(T)</td>
</tr>
<tr>
<td>CISD</td>
<td>cc-pVDZ CISD</td>
<td>cc-pVTZ CISD</td>
<td>cc-pVQZ CISD</td>
</tr>
</tbody>
</table>

FIG. 3. Normal distributions $\rho(R)$ for the errors in the calculated bond distances. The distributions have been calculated from the mean errors and the standard deviations in Table IV (pm). For easy comparison, all distributions have been normalized to one and plotted on the same horizontal and vertical scales.
Basis Set Limit

- As the number of basis functions increases, the wavefunction is better represented and the energy decreases to approach the complete basis set (CBS) limit.

- Convergence with respect to size is very slow.

- Actually employing an infinite number of basis functions is impossible, but we can try to estimate the energy at the CBS limit.

- Can use hierarchical basis sets to extrapolate to CBS.

---

Fig. 2. Error in the valence MP2 (black lines) and CCSD (gray lines) correlation energies of Ne (full lines), H₂O (dashed lines), and N₂ (dotted lines), as a function of 1/N, where N is the number of basis functions.
Basis Set Extrapolation

- Extrapolations between two AM adjacent basis sets are good (e.g., cc-pVTZ & cc-pVQZ or heavy-aug-cc-pVDZ & heavy-aug-cc-pVTZ)
- Extrapolate HF and correlation energies separately
  \[ E_{CCSD(T)}^{CBS} = E_{HF}^{CBS} + E_{corr}^{CBS} \]
- Two schemes
  - Exponential Form
    \[ E_X^{HF} = E_{CBS}^{HF} + A \exp(-\alpha X) \text{ and } E_X^{corr} = E_{CBS}^{corr} + BX^{-3} \]
  - Power Form
    \[ E_X^{HF} = E_{CBS}^{HF} + BX^{-3} \text{ and } E_X^{corr} = E_{CBS}^{corr} + BX^{-\beta}. \]
- \( X^{-3} \) term remedies only basis set incompleteness error, not BSSE, so CP-correction recommended
- DZ is erratic so avoid when possible
- Two-step extrapolation is excellent when a high-quality method like CCSD(T) is affordable at TZ & QZ, but this is often impossible, and correlation extrapolation at the MP2 level is not the best
- Use three-step extrapolation
  \[ E_{CCSD(T)}^{CBS} \approx E_{MP2}^{CBS} + \Delta CCSD(T) = E_{HF}^{CBS} + E_{corr MP2}^{CBS} + (E_{small}^{corr CCSD(T)} - E_{small}^{corr MP2}). \]
- HF and MP2 correlation extrapolation as before with large basis sets
- deltaCCSD(T) “coupled-cluster correction” with smaller basis to recover correlation between MP2 and CCSD(T)
- With large enough basis sets (6Z) can achieve within 0.1 mHartrees of the CBS limit
Basis Set Extrapolation

- Extrapolations between two AM adjacent basis sets are good (e.g., cc-pVTZ & cc-pVQZ or heavy-aug-cc-pVDZ & heavy-aug-cc-pVTZ)
- Extrapolate HF and correlation energies separately
  \[ E_{CCSD(T)}^{CBS} = E_{HF}^{CBS} + E_{corr}^{CBS} \]
- Two schemes
  - Exponential Form
    \[ E_X^{HF} = E_{CBS}^{HF} + A \exp(-\alpha X) \] and \[ E_X^{corr} = E_{CBS}^{corr} + BX^{-3} \]
  - Power Form
    \[ E_X^{HF} = E_{CBS}^{HF} + BX^{-\alpha} \] and \[ E_X^{corr} = E_{CBS}^{corr} + BX^{-\beta} \]

- The question arises at which level the extrapolation should be performed. The choice of a method is determined by the available computer power and affordable at TZ & QZ level is not the best
- Use three-step extrapolation
  - Use HF and MP2 correlation energies followed by application of a "couple-clusters correction"
  - \( E_{exp}^{corr} = E_{small}^{corr} + \Delta E_{corr} \) at the MP2 level
  - \( \Delta E_{corr} = (E_{corr}^{CCSD(T)} - E_{corr}^{MP2}) \)
  - Recover correlation between MP2 and CCSD(T) extrapolation at the MP2 level
  - With large enough extrapolation limit

Fig. 3. Normal distributions of the two-parameter fits \( E_X^{corr} = E_x^{corr} + AX^{-3} \) as a function of the data fitted (from \( X_{min} \) to \( X_{max} \)).
Basis Set Extrapolation

Extrapolations between two AM adjacent basis sets are good (e.g., cc-pVTZ & cc-pVQZ or heavy-aug-cc-pVDZ & heavy-aug-cc-pVTZ)

Extrapolate HF and correlation energies separately

\[ E_{\text{CBS}}^{\text{CCSD(T)}} = E_{\text{HF}}^{\text{CBS}} + E_{\text{corr}}^{\text{CCSD(T)}} \]

Two schemes

- Exponential Form
  \[ E_X^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + A \exp(-\alpha X) \text{ and } E_X^{\text{corr}} = E_{\text{CBS}}^{\text{corr}} + BX^{-3} \]

- Power Form
  \[ E_X^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + BX^{-\alpha} \text{ and } E_X^{\text{corr}} = E_{\text{CBS}}^{\text{corr}} + BX^{-\beta} \]

The \(X^{-3}\) term remedies only basis set incompleteness error, not BSSE, so CP-correction recommended

DZ is erratic so avoid when possible

Two-step extrapolation is excellent when a high-quality method like CCSD(T) is affordable at TZ & QZ, but this is often impossible, and correlation extrapolation at the MP2 level is not the best

Use three-step extrapolation

\[ E_{\text{CBS}}^{\text{CCSD(T)}} \approx E_{\text{MP2}}^{\text{CBS}} + \Delta \text{CCSD(T)} = E_{\text{HF}}^{\text{MP2}} + E_{\text{corr MP2}}^{\text{CBS}} + (E_{\text{small corr CCSD(T)}}^{\text{CBS}} - E_{\text{small corr MP2}}^{\text{CBS}}) \]

HF and MP2 correlation extrapolation as before with large basis sets

deltaCCSD(T) “coupled-cluster correction” with smaller basis to recover correlation between MP2 and CCSD(T)

With large enough basis sets (6Z) can achieve within 0.1 mHartrees of the CBS limit


**General Comments**

- The bigger the basis, the better? Usually — need to balance with correlation method; e.g., cc-pVQZ is great for CCSD(T), but overkill for Hartree-Fock

- STO-3G should not be used: too small

- Hard to afford more than polarized double-zeta basis sets except for small molecules

- Anions must have diffuse functions

- In our experience, cc-pVDZ is not necessarily better than 6-31G(d,p); however, cc-pVTZ *is* better than 6-311G(d,p) or similar

- Convergence of *ab initio* results is disappointingly slow with respect to basis set for non-DFT methods (see, for example, papers by Helgaker or Dunning)

- DFT is less dependent on basis set size than wavefunction-based methods (see, for example, papers by Angela Wilson)