Basis Sets in Quantum Chemistry

C. David Sherrill

School of Chemistry and Biochemistry

Georgia Institute of Technology

Basis Sets

- 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

- Sometimes, theorists might also refer to N-electron basis sets, which is something else entirely — sets of Slater determinants

Basis Sets in Quantum Chemistry

- 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 (STO's)

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

- N is a normalization constant

- a, b, c control angular momentum, $L = a + b + c$

- ζ (zeta) controls the width of the orbital (large ζ gives tight function, small ζ 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 (GTO's)

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

- Again, a, b, c control angular momentum, $L = a + b + c$

- Again, ζ 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

Contracted Gaussian-Type Orbitals (CGTO's)

- *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...

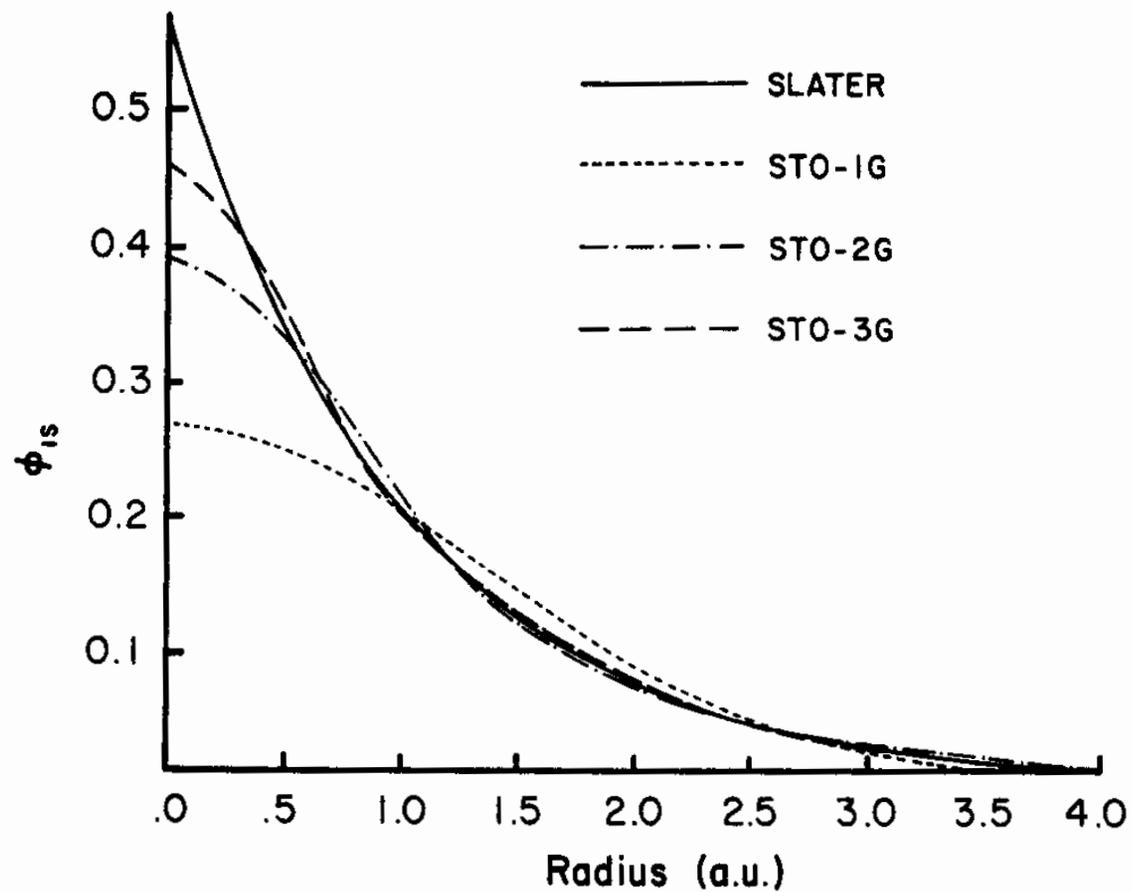$$\phi_{abc}^{CGTO}(x, y, z) = N \sum_{i=1}^{n} c_i x^a y^b z^c e^{-\zeta_i r^2} \tag{1}$$

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.

Types of Basis Sets

Minimal: One basis function (STO, GTO, or CGTO) for each atomic orbital in the atom

Double-zeta: Two basis functions for each AO

Triple-zeta: Three basis functions for each AO

... and etc. for quadruple-zeta (QZ), 5Z, 6Z, ... Having different-sized functions allows the orbital to get bigger or smaller when other atoms approach it

A split-valence basis uses only one basis function for each core AO, and a larger basis for the valence AO's

Examples

H atom, minimal basis: One 1s AO, one (STO, GTO, or CGTO) basis function

C atom, minimal basis: 1s, 2s, $2p_x$, $2p_y$, $2p_z$ AO's (5), so 5 basis functions

C atom, double-zeta basis: Two basis functions per AO, so 10 basis functions

C atom, split-valence double-zeta basis: 9 basis functions (why?)

Polarization Functions

- 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, mix it with basis functions of angular momentum $l + 1$

- This gives "polarized double-zeta", or "double-zeta plus polarization" basis sets, etc

Counting Polarization Functions

- We know there should be 5 d functions (usually chosen as $d_{x^2-y^2}$, d_{z^2}, d_{xy}, d_{xz}, and d_{yz}); these are called "pure angular momentum" functions (even though they aren't, really)

- Computers would prefer to work with 6 d functions (d_{x^2}, d_{y^2}, d_{z^2}, d_{xy}, d_{xz}, and d_{yz}); these are called "6 Cartesian d functions"

- $d_{x^2} + d_{y^2} + d_{z^2}$ looks like an s orbital

- Similar answers are obtained using 5 or 6 d functions

- For f functions, it's 7 versus 10 f functions

- Common reason for disagreement between calculated

results and literature values — different number of polarization functions

- Some basis sets were developed using 5d's [cc-pVXZ, newer Pople basis sets like 6-311G(3df)], some using 6 d's [older Pople basis sets like 6-31G(d)]; results don't change much, but it is more consistent to use the same number of polarization functions as the basis set designer

- Different programs have different default values for number of polarization functions — can make it hard to compare across programs if you don't notice it

Diffuse Functions

- Diffuse functions have small ζ exponents; this means 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!

Anatomy of a Basis Set: H atom, STO-3G

A basis set of Contracted Gaussian-Type Orbitals (CGTO's) needs to specify the exponents (ζ_i's) and the contraction coefficients (c_i's). These are given below in the format used by the Gaussian program (exponents first)

```
 H    0
 S    3  1.00
        3.42525091              0.154328970
        0.623913730             0.535328140
        0.168855400             0.444634540
 ****
```

Anatomy of a Basis Set: C atom, STO-3G

```
C    0
S    3   1.00
         71.6168370              0.154328970
         13.0450960              0.535328140
         3.53051220              0.444634540
SP   3   1.00
         2.94124940             -0.999672300E-01        0.155916
         0.683483100             0.399512830            0.607683
         0.222289900             0.700115470            0.391957
     ****
```

STO-3G and other basis sets by John Pople use "SP" shells,
which share exponents for s and p functions

Pople Basis Sets

- Developed by the late Nobel Laureate, John Pople, and popularized by the Gaussian set of programs

- STO-3G is a minimal basis set in which each AO is represented by 3 Gaussians (3G), chosen to mimic the behavior of a STO

- 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

Anatomy of a Basis Set: C atom, 6-31G

```
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
        1.88128850             -0.160854200              0.316424000
        0.544249300             1.14345640               0.744308300
SP   1   1.00
        0.168714400             1.00000000               1.00000000
```

More on 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

Designations of Basis Set Size

It is useful to refer to how large a basis set is by listing how many sets of functions it has for each angular momentum type. Examples:

H atom, minimal basis: 1s (1 function)

C atom, minimal basis: 2s1p (5 functions)

C atom, double-zeta basis: 4s2p (10 functions)

C atom, split-valence double-zeta basis: 3s2p (9 functions)

Dunning's Correlation-Consistent Basis Sets

- Thom Dunning pointed out that basis sets optimized at the Hartree-Fock level might not be ideal for correlated computations

- The "correlation consistent" basis sets are optimized using correlated (CISD) wavefunctions

- cc-pVXZ means a Dunning correlation-consistent, polarized valence, X-zeta basis; X=D,T,Q,5,6,7

- Functions are added in shells. cc-pVDZ for C atom consists of 3s2p1d. cc-pVTZ would be 4s3p2d1f. cc-pVQZ would be 5s4p3d2f1g.

- The Dunning basis sets are designed to converge smoothly

toward the complete (infinite) basis set limit

- A prefix "aug" means one set of diffuse functions is added for every angular momentum present in the basis; aug-cc-pVDZ for C atom has diffuse s,p,d

- Somewhat cheaper versions of aug-cc-pVXZ are the "minimally augmented" maug-cc-pVXZ basis sets of Truhlar, which add only diffuse s and p functions on heavy atoms; these are the most important of the diffuse functions, and these basis sets will be smaller (faster computations) hopefully without adding much error

- Functions describing core correlation are denoted by the letter "C" in the cc-pCVXZ or aug-cc-pCVXZ basis sets; you should use basis sets like this if you do not freeze core

electrons in post-Hartree-Fock computations like MP2, CISD, CCSD(T), etc.

- For 2nd row elements Al-Ar, the original d polarization functions were not tight enough; revised in cc-pV(X+d)Z basis sets

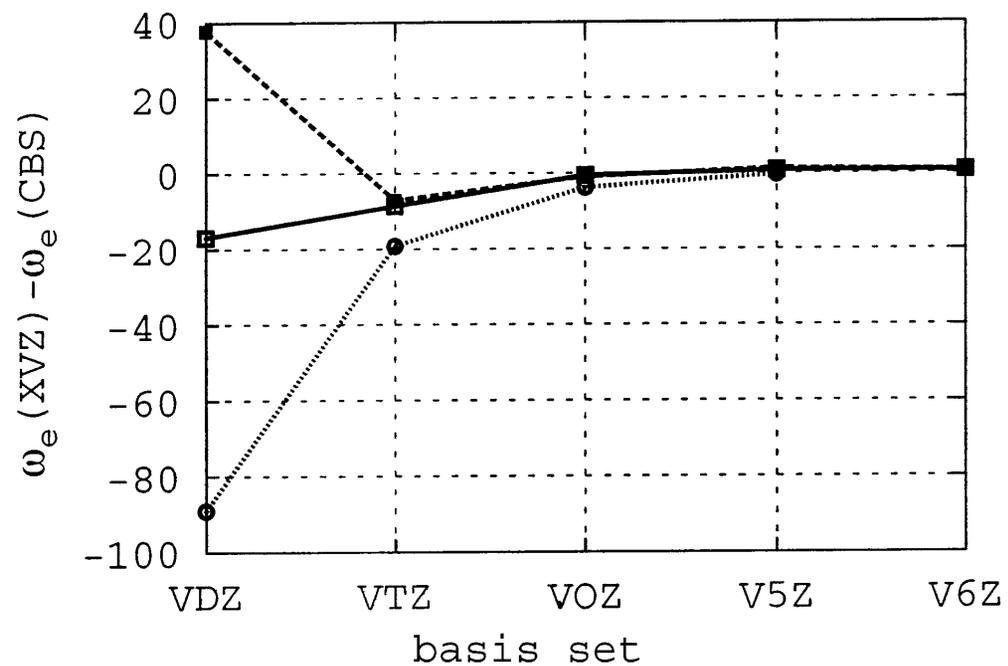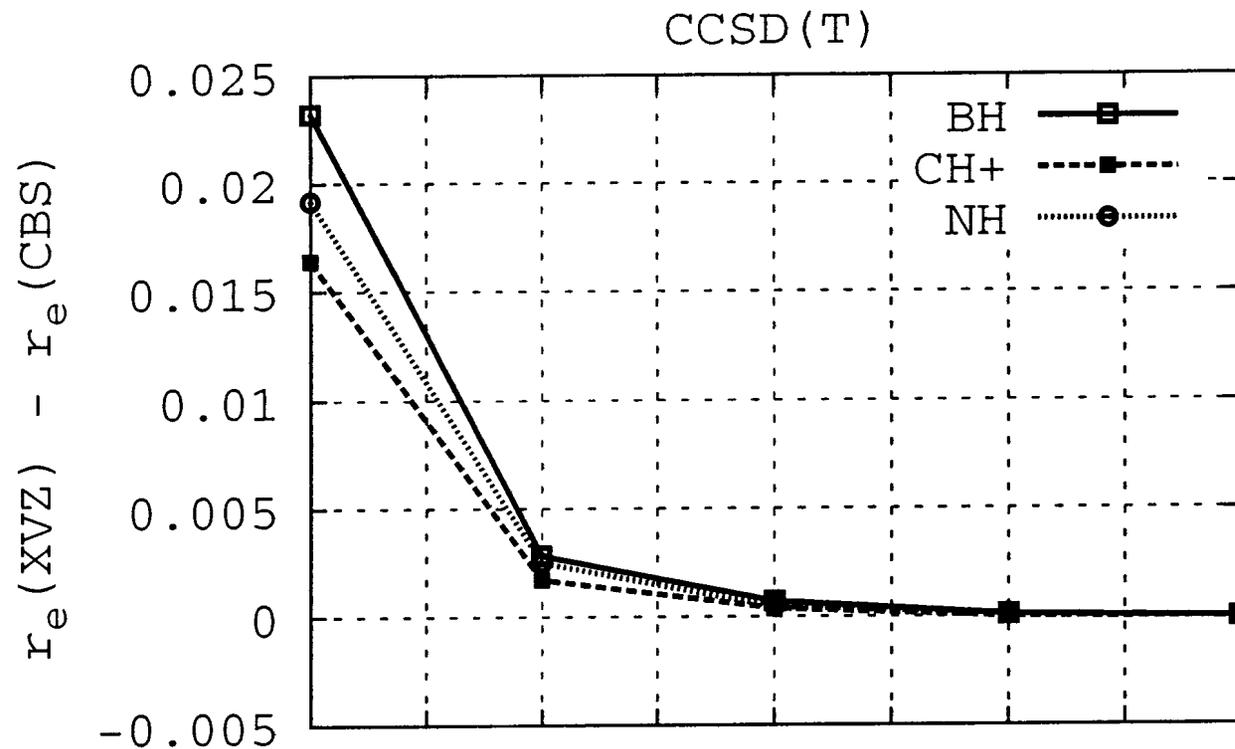# Convergence of Vibrational Frequencies for BH, CH$^+$, and NH

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

Temelso, Valeev, and Sherrill, *J. Phys. Chem. A* **108**, 3068 (2004).

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

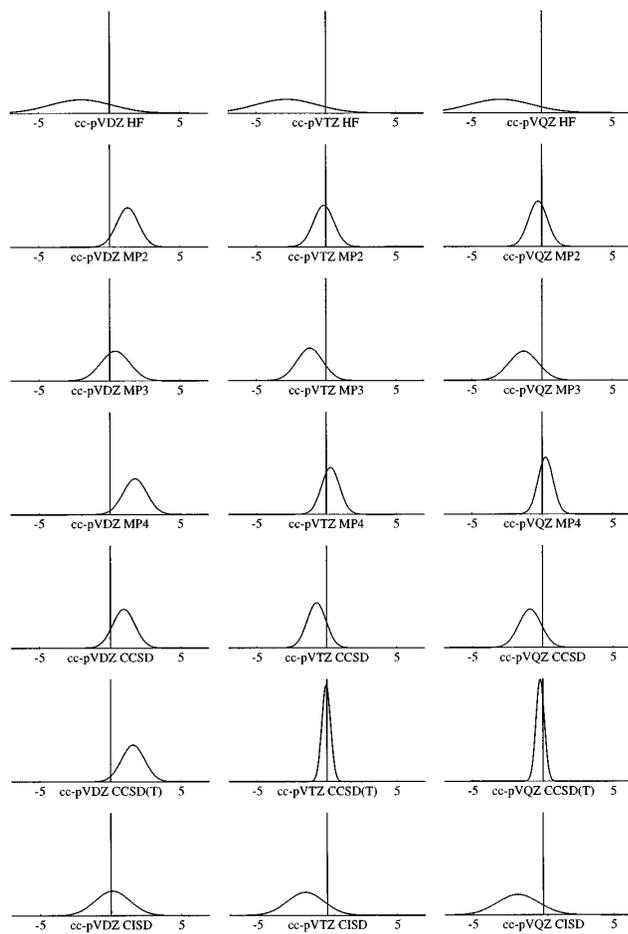Temelso, Valeev, and Sherrill, *J. Phys. Chem. A* **108**, 3068 (2004).

FIG. 3. Normal distributions $\rho(R)$ for the errors in the calculated bond distances. The distributions have been calculated from the mean errors in Table III 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.

Helgaker et al, *J. Chem. Phys.* **106**, 6430 (1997).

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)

- Best resource for getting basis sets: https://bse.pnl.gov/bse/portal

- I couldn't mention all the important basis sets — others are out there!