

Basis Sets and Extrapolation

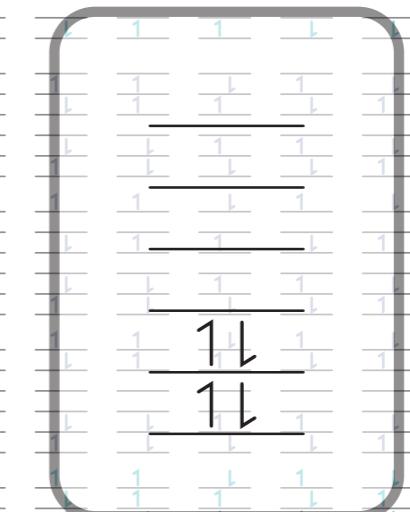
5 May 2016

Two limitations to solving the Schrödinger Equation

Two limitations to solving the Schrödinger Equation

Arrange Four Electrons in Six Orbitals

1 Ground Configuration HF

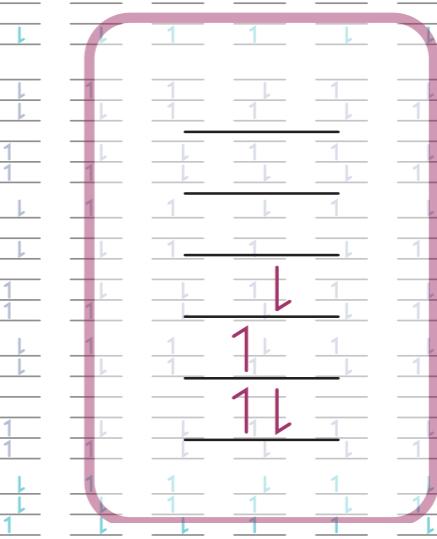


Two limitations to solving the Schrödinger Equation

Arrange Four Electrons in Six Orbitals

1 Ground Configuration HF

32 Single Excitations S



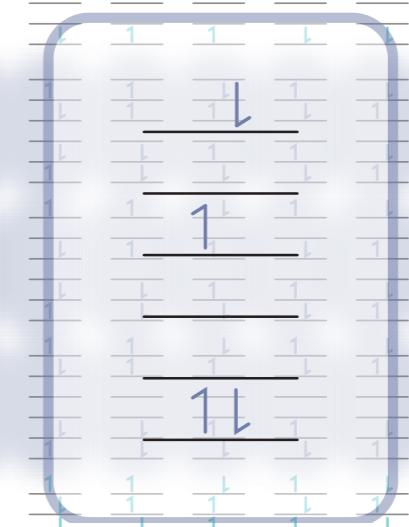
Two limitations to solving the Schrödinger Equation

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

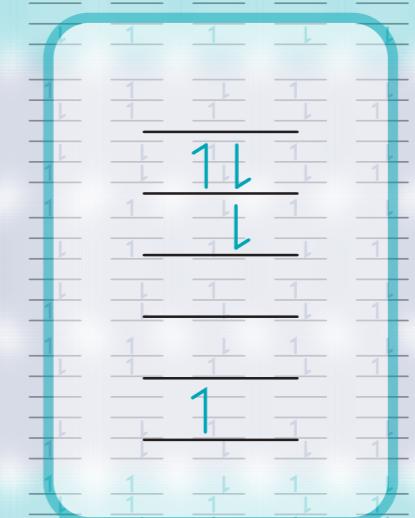
Arrange Four Electrons in Six Orbitals

1 Ground Configuration HF

32 Single Excitations S

168 Double Excitations D

224 Triple Excitations T



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

224 Triple Excitations T

70 Quadruple Excitations Q

495 Slater Determinants



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

224 Triple Excitations T

70 Quadruple Excitations Q

Full CI

495 Slater Determinants

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

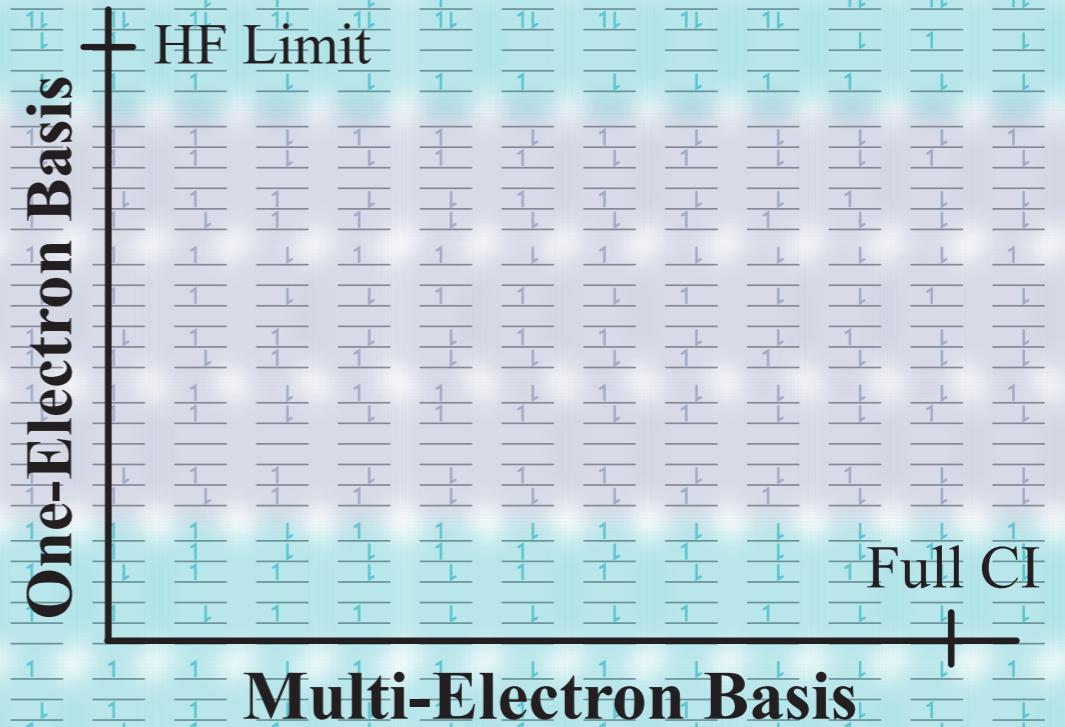


Full multi-electron basis recovers all correlation within a given one-electron basis



Full one-electron basis recovers the SCF energy without limitation on the shape of MO

Two limitations to solving the Schrödinger Equation



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

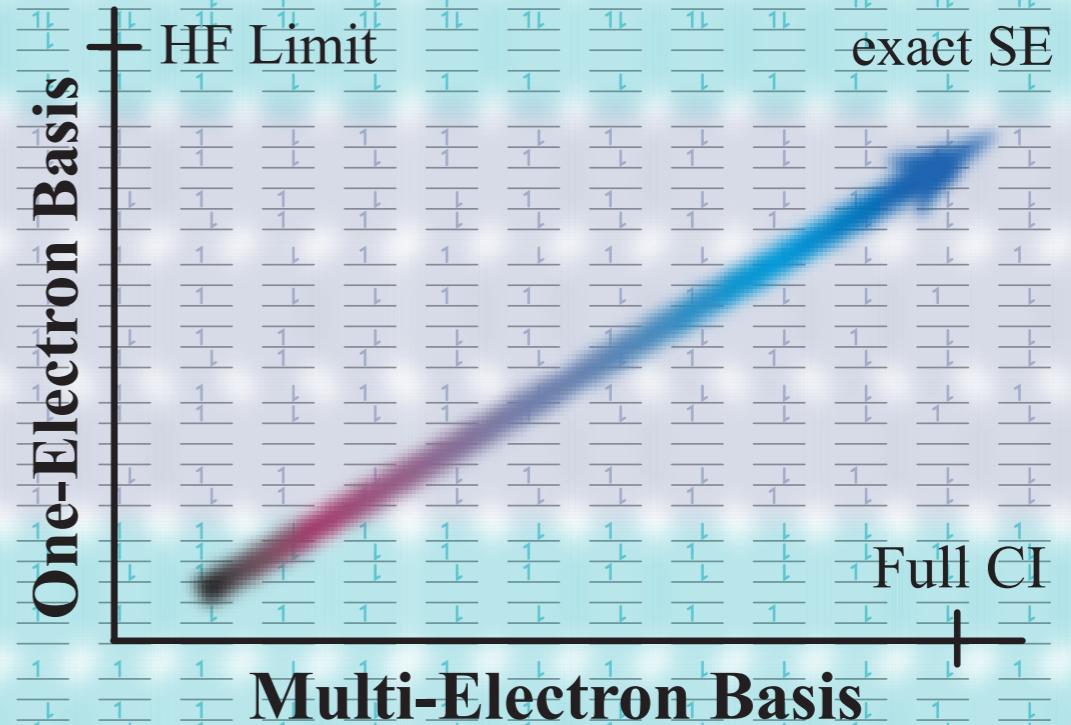


Full multi-electron basis recovers all correlation within a given one-electron basis



Full one-electron basis recovers the SCF energy without limitation on the shape of MO

Two limitations to solving the Schrödinger Equation



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



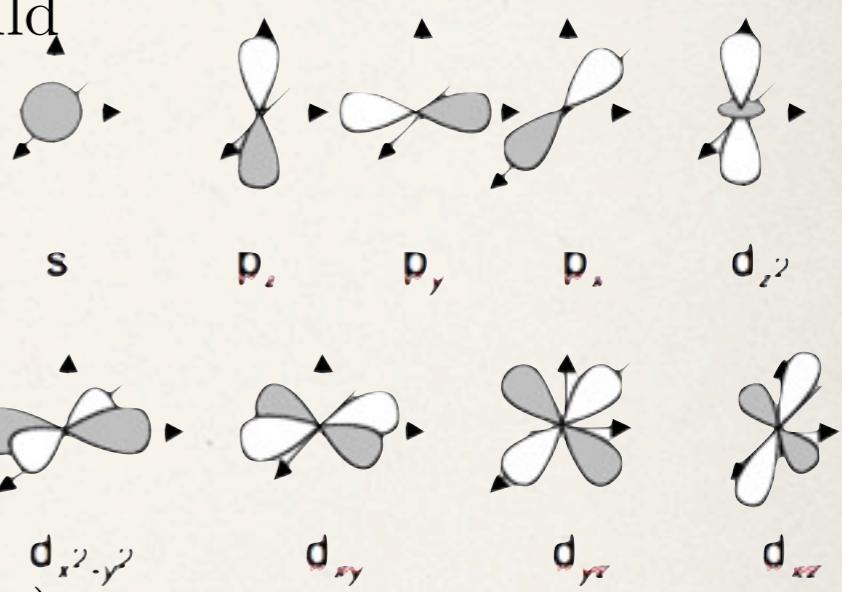
Full multi-electron basis recovers all correlation within a given one-electron basis



Full one-electron basis recovers the SCF energy without limitation on the shape of MO

Basis set in quantum chemistry

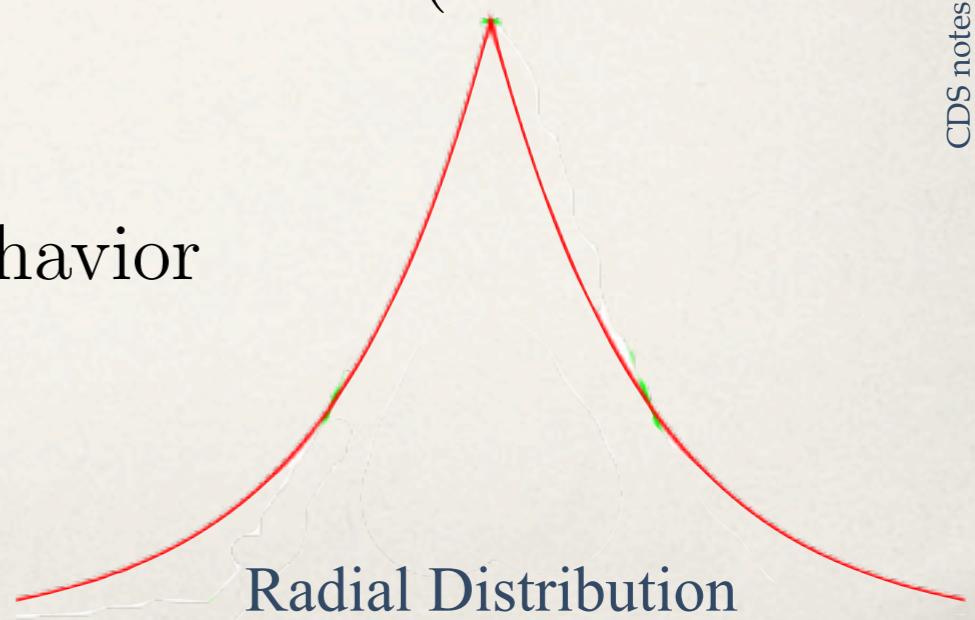
- Generically, a basis set is a collection of vectors which spans (defines) a space in which a problem is solved
- $\hat{i}, \hat{j}, \hat{k}$ define a Cartesian, 3D linear vector space
- In quantum chemistry, the “basis set” usually refers to the set of (nonorthogonal) one-particle functions used to 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)

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

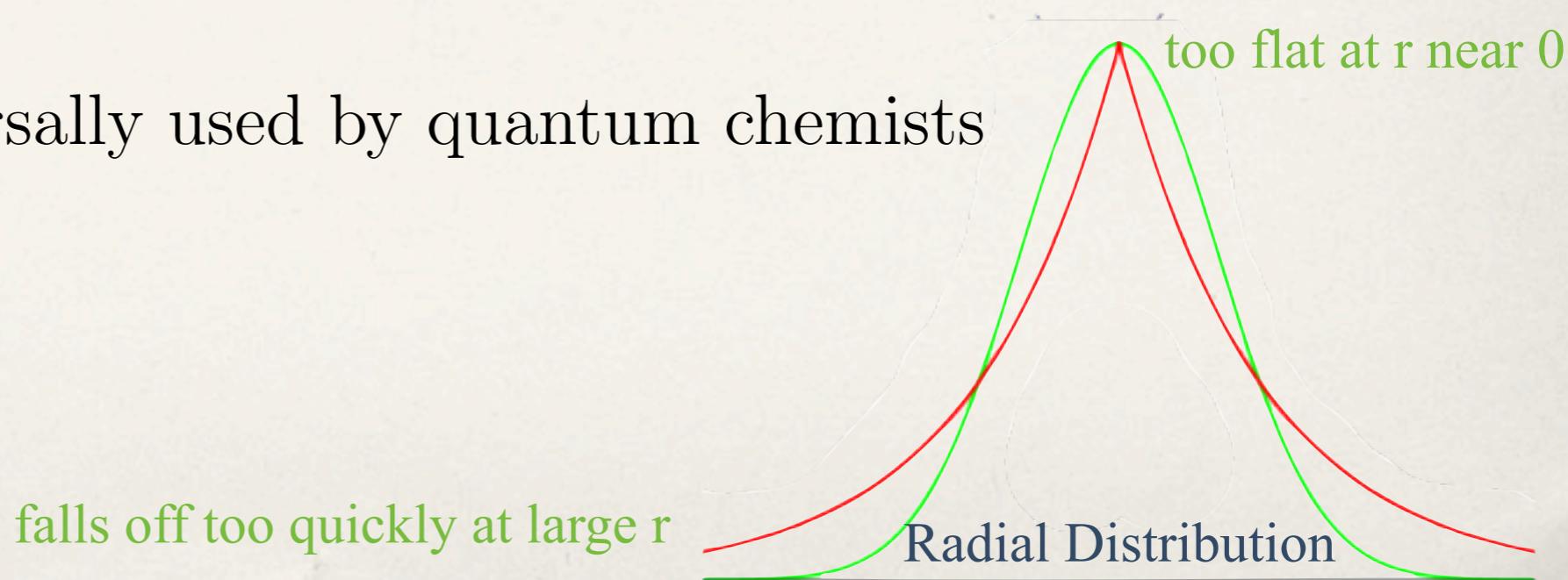
- N is a normalization constant
- a, b, c control angular momentum, $L = a + b + c$
- ζ (zeta) controls the width of the orbital (large ζ gives tight function, small ζ gives diffuse function)
- These are H-atom-like, at least for 1s; however, they lack radial nodes and are not pure spherical harmonics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



Gaussian-type orbitals (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, 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 (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...

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

Contracted GTO Primitive GTO

Contracted Gaussian-type orbitals (CGTOs), cont.

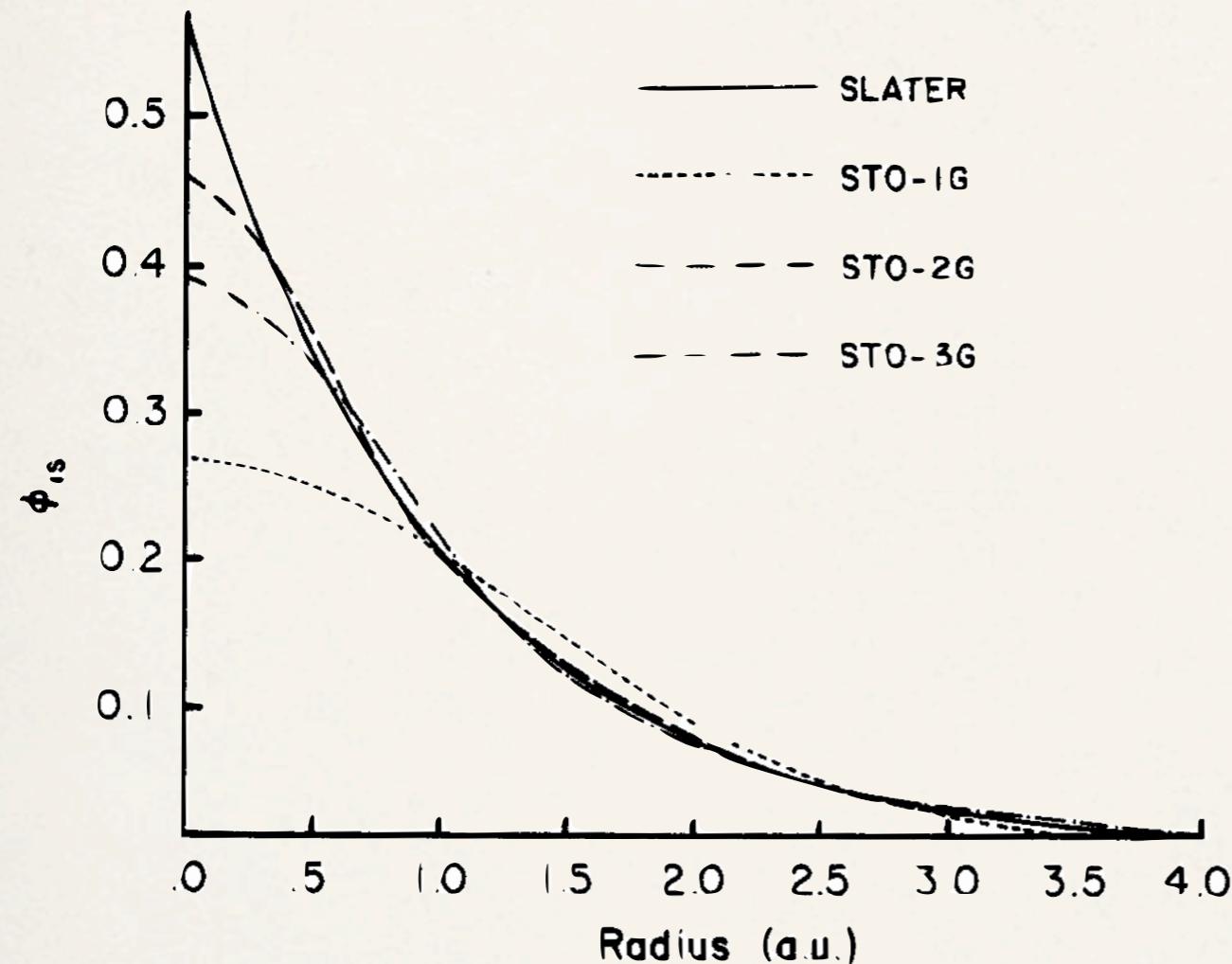


Figure 3.3 Comparison of the quality of the least-squares fit of a 1s Slater function ($\zeta = 1.0$) obtained at the STO-1G, STO-2G, and STO-3G levels.

- 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 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

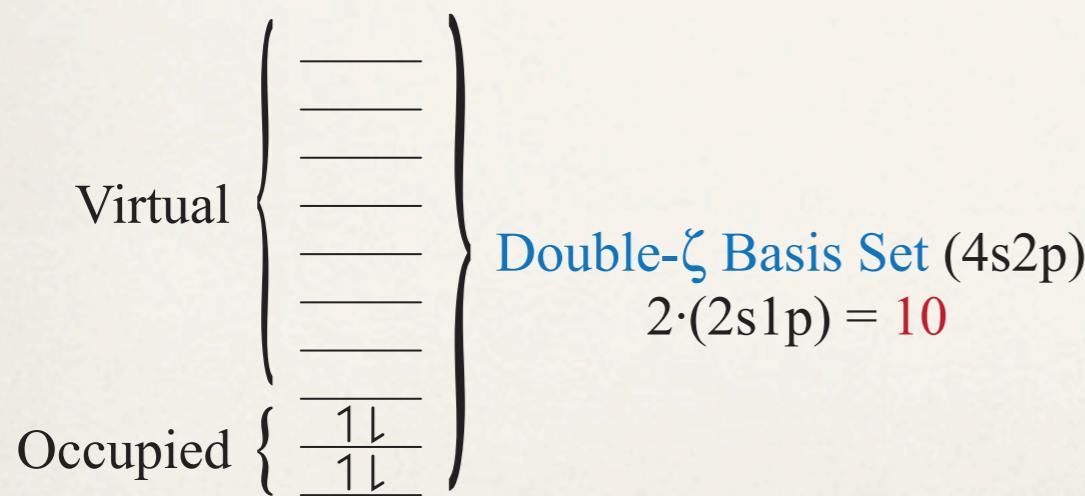
- Minimal: one basis function (STO, GTO, or CGTO) for each atomic orbital in each atom
- Double-zeta (DZ): two basis functions for each AO
- Triple-zeta (TZ): three basis functions for each AO
- etc., for quadruple-zeta (QZ), 5Z, 6Z, ...
- The presence of different-sized functions allows the orbital to get bigger or smaller when other atoms approach it, adds flexibility to adequately describe anisotropic electron distribution in molecules



Adding: add'l fns of varying sizes with valence ang. mom.

Designations of basis set size

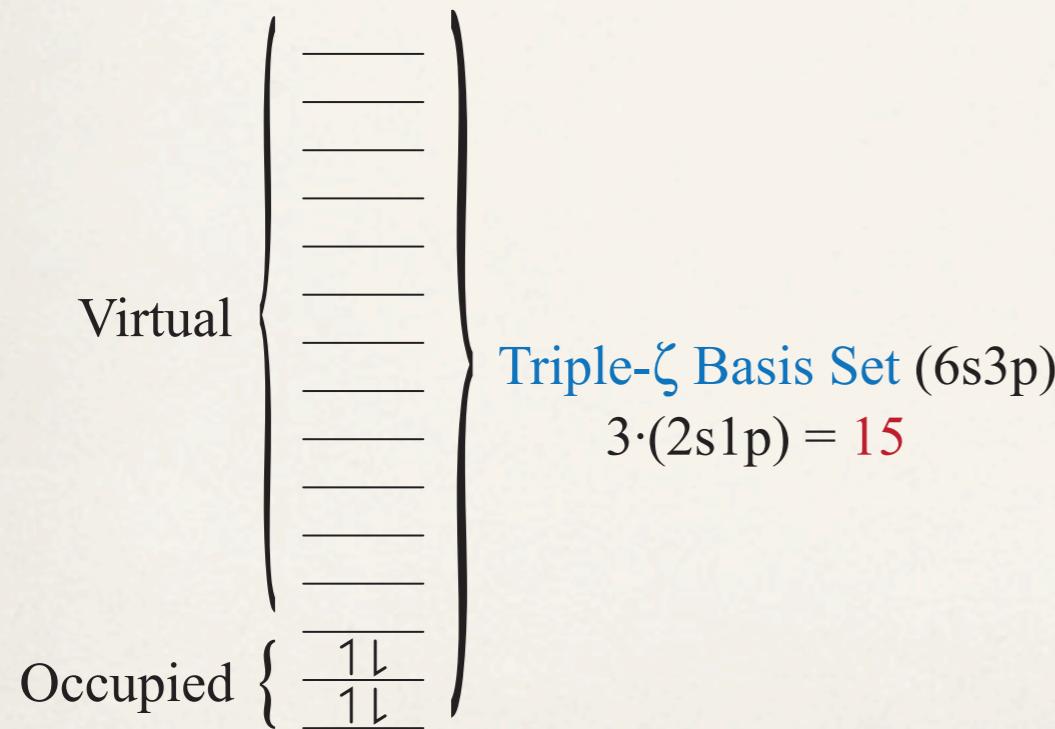
- Minimal: one basis function (STO, GTO, or CGTO) for each atomic orbital in each atom
- Double-zeta (DZ): two basis functions for each AO
- Triple-zeta (TZ): three basis functions for each AO
- etc., for quadruple-zeta (QZ), 5Z, 6Z, ...
- The presence of different-sized functions allows the orbital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

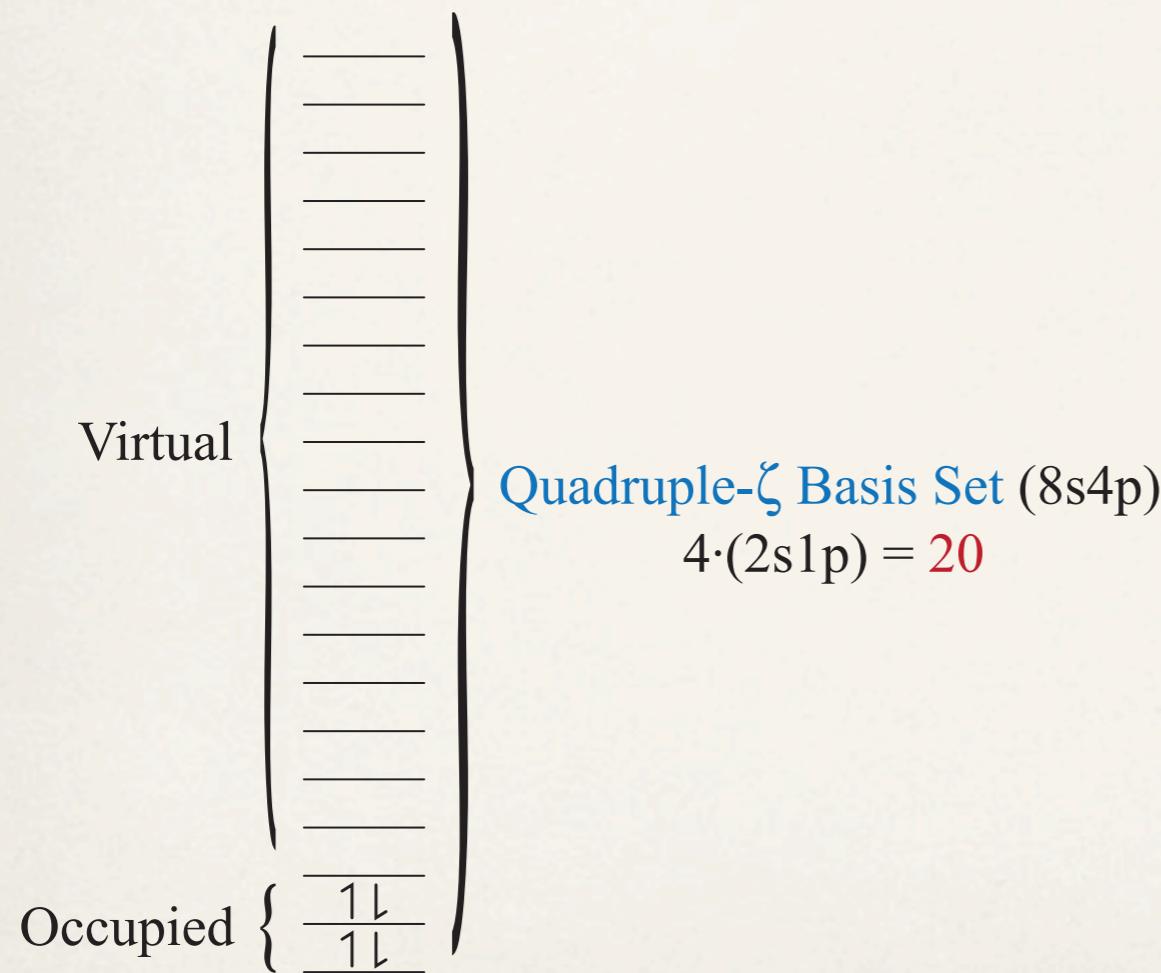
- Minimal: one basis function (STO, GTO, or CGTO) for each atomic orbital in each atom
- Double-zeta (DZ): two basis functions for each AO
- Triple-zeta (TZ): three basis functions for each AO
- etc., for quadruple-zeta (QZ), 5Z, 6Z, ...
- The presence of different-sized functions allows the orbital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

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

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

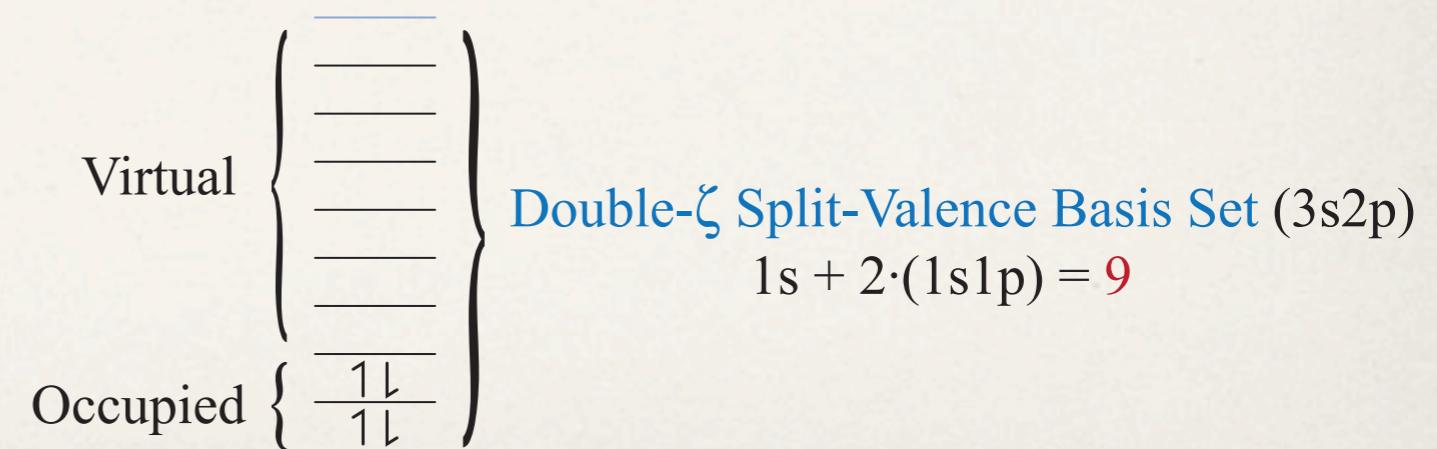
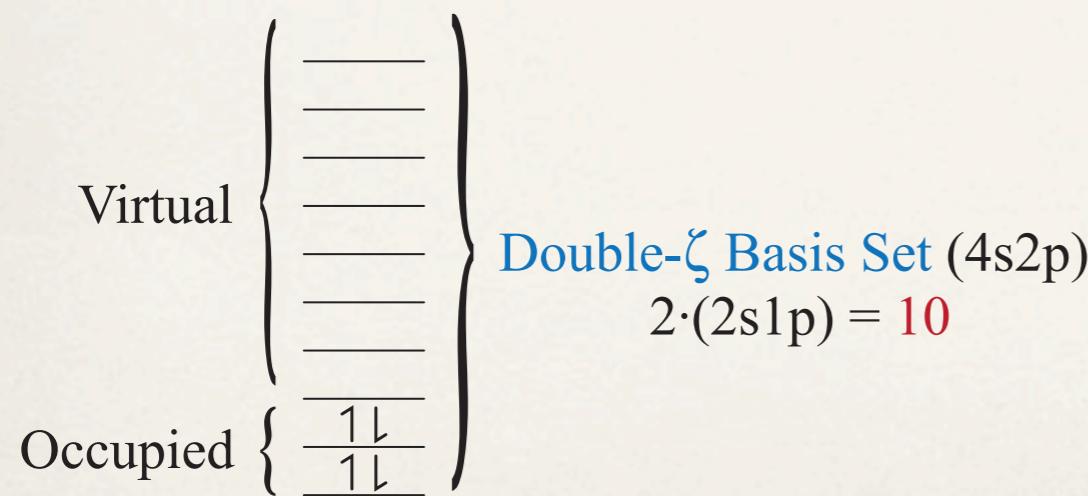
Split-valence basis sets

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



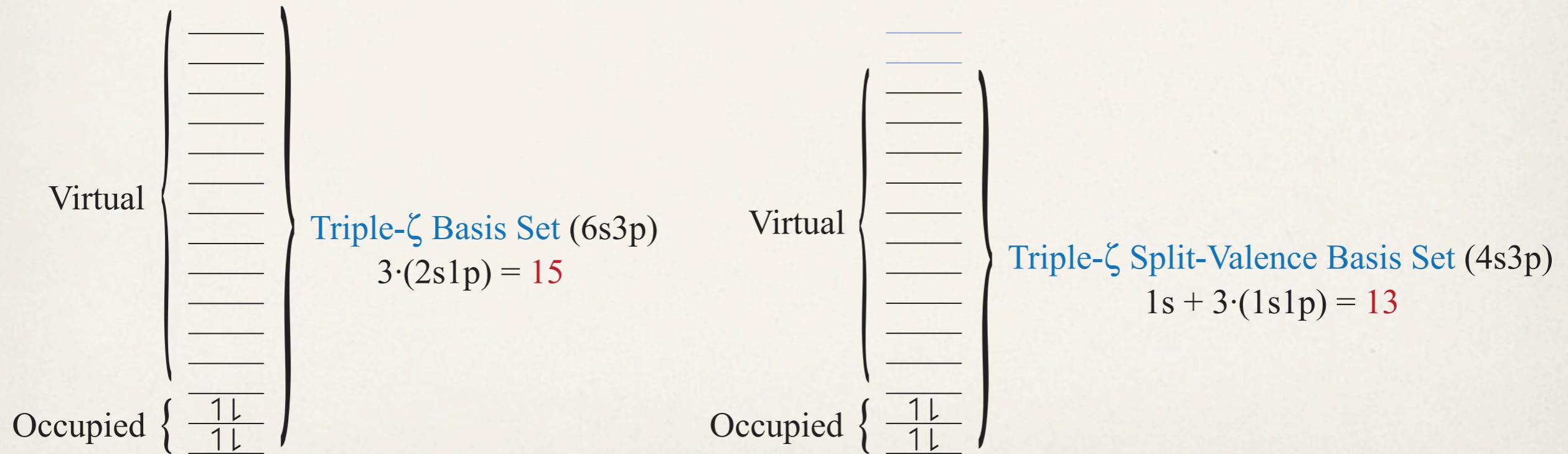
Split-valence basis sets

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



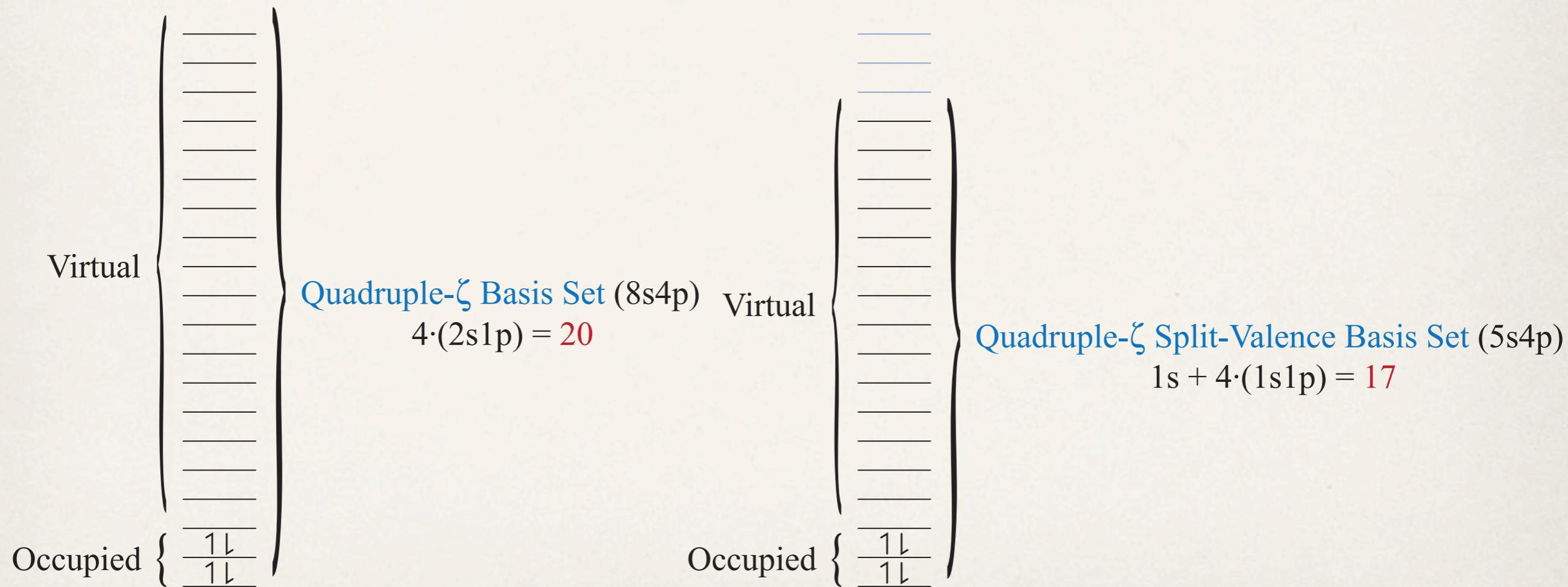
Split-valence basis sets

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



Split-valence basis sets

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

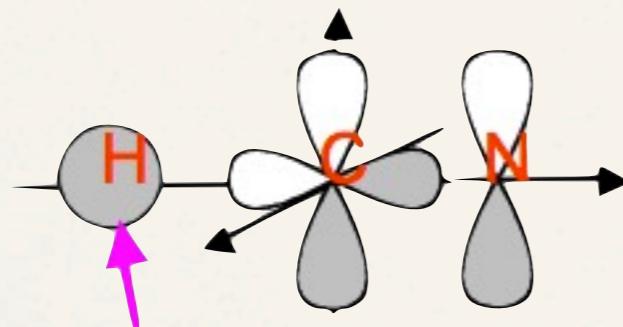


Pitfall: Occupancy

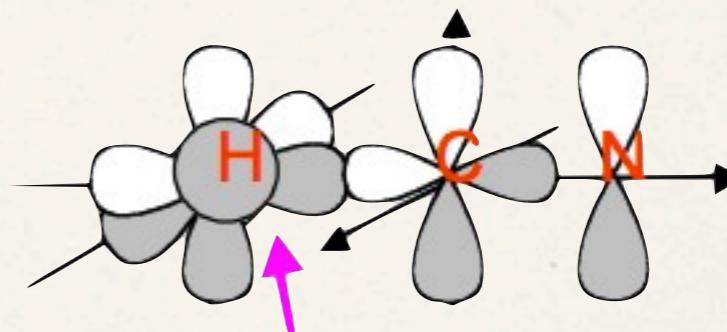
- ➊ QC programs which use symmetry can yield valuable efficiency gains for symmetry molecules (prefactor 1/point group order)
- ➋ Since there is no coupling between each subgroup (*e.g.*, a_1 , b_2 , b_1 , a_2) of orbitals, the program must guess an occupancy
 - ➌ If guess is incorrect, SCF energies 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
 - ➍ This pitfall is more likely with larger extended basis sets
 - ➎ Beware different ordering of symmetry subgroups between QC programs, esp. C_{2v}
- ➏ **Pitfall:** for extended basis sets, when using symmetry, a QC program may allocate electrons to a non-ground state configuration of orbitals.
 - ➐ **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



Anisotropic distribution
that can not be
described by *s* orbitals



Anisotropic distribution
that can now be
described by *p* orbitals

Pitfall: Counting polarization functions

- ➊ 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)

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

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

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

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

- ➌ Similar answers are obtained using 5 or 6 d functions

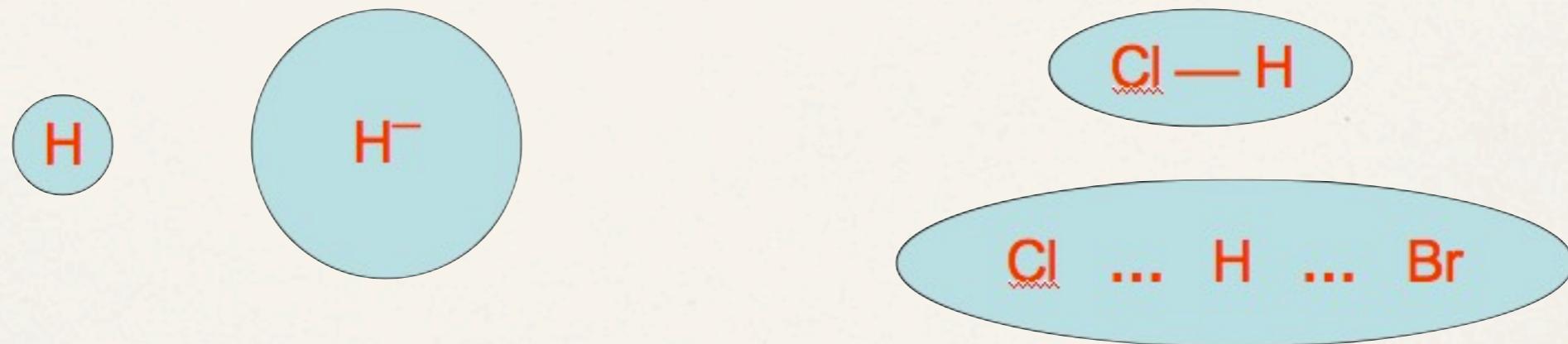
- ➌ For f functions, it’s 7 versus 10 f functions

- ➍ **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

- ➎ **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!



Pitfall: Linear Dependency

- ➊ 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 over-complete
 - ➋ With a loss in uniqueness in the molecular orbital coefficients, SCF may be slow to converge or behave erratically
 - ➌ Usually diagnosed from eigenvalues of the overlap matrix; thereafter near-degeneracies can be projected out according to a cutoff value
-
- ➍ **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.
 - ➎ **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

STO-2G, STO-3G, STO-6G, STO-3G*, 3-21G, 3-21++G, 3-21G*, 3-21GSP, 4-31G, 4-22GSP, 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), 6-311G, 6-311G*, 6-311G**, 6-311+G*, 6-311++G**, 6-311++G(2d,2p), 6-311G(2df,2pd), 6-311++G(3df,3pd), MINI (Huzinaga), MINI (Scaled), MIDI (Huzinaga), MIDI!, 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, pV6Z, pV7Z, cc-pVDZ(seg-opt), cc-pVTZ(seg-opt), cc-PVQZ(seg-opt), cc-pCVDZ, cc-pCVTZ, cc-pCVQZ, cc-pCV5Z, aug-cc-pVDZ, aug-cc-pVTZ, aug-cc-pVQZ, aug-cc-pV5Z, aug-cc-pV6Z, aug-pV7Z, aug-cc-pCVDZ, aug-cc-pCVTZ, aug-cc-pCVQZ, aug-cc-pCV5Z, d-aug-cc-pVDZ, d-aug-cc-pVTZ, d-aug-cc-pVQZ, d-aug-cc-pV5Z, d-aug-cc-pV6Z, Feller Misc. CVDZ, Feller Misc cVTZ, Feller Misc. CVQZ, NASA Ames ANO, Roos Augmented Double Zeta ANO, Roos Augmented Triple Zeta ANO, WTBS, GAMESS VTZ, GAMESS PVTZ, Partridge Uncontr. 1, Partridge Uncontr. 2, Partridge Uncontr. 3, Ahlrichs VDZ, Ahlrichs, pVDZ, Ahlrichs VTZ, Ahlrichs TZV, Binning/Curtiss SV, Binning/Curtiss VTZ, Binning/Curtiss SVP, Binning-Curtiss VTZP, Mclean/Chandler VTZ, SV+Rydberg (Dunning-Hay), SVP+Rudberg (Dunning-Hay), SVP+Diffuse+Rydberg, DZ+Rydberg (Dunning), DZP+Rydberg (Dunning), DZ+Double Rydberg (Dunning-Hay), SV+Double Rydberg (Dunning-Hay), Wachters+f, Bauschlicher ANO, Sadlej pVTZ, Hay-Wadt MB(n+1)ECP, Hay-Wadt VDZ(n+1)ECP, LANL2DZ ECP, SBKJC VDZ ECP, CRENBL ECP, CRENBS ECP, Stuttgart RLC ECP, Stuttgart RSC ECP, DZVP (DFT Orbital), DZVP2 (DFT Orbital), TZP (DFT Orbital), DeMon Coulomb Fitting, DGauss A1 DFT Coulomb Fitting, DGauss A1 DFT Exchange Fitting, DGauss A2 DFT Coulomb Fitting, DGauss A2 DFT Exchange Fitting, Ahlrichs Coulomb Fitting, cc-pVDZ-fit2-1, cc-pVTZ-fit2-1, cc-pVDZ_DK, cc-pVTZ_DK, cc-pVQZ_DK, cc-pV5Z_DK, cc-pVDZ(pt/sf/fw), cc-PVTZ(pt/sf/fw), cc-pVQZ(pt/sf/fw), cc-pV5Z(pt/sf/fw), cc-pVDZ(fi/sf/fw), cc-pVTZ(fi/sf/fw), cc-pVQZ(fi/sf/fw), cc-pV5Z(fi/sf/fw), cc-pVDZ(pt/sf/sc), cc-pVDZ(pt/sf/lc), cc-pVTZ(pt/sf/sc), cc-PVTZ(pt/sf/lc), ccp-PVQZ(pt/sf/sc), cc-pVQZ(pt/sf/lc), cc-PV5Z(pt/sf/sc), cc-PV5Z(pt/sf/lc), cc-pVDZ(fi/sf/sc), cc-PVDZ(fi/sf/lc), cc-PVTZ(fi/sf/sc), cc-PVTZ(fi/sf/lc), cc-PVQZ(fi/sf/sc), cc-PVQZ(fi/sf/lc), cc-PV5Z(fi/sf/sc), cc-pV5Z(fi/sf/lc), Pople-Style Diffuse, STO-3G* Polarization, 3-21G* Polarization, 6-31G* Polarization, 6-31G** Polarization, 6-311G* Polarization, 6-311G** Polarization, Pople (2d/2p) Polarization, Pople (3df,3pd) Polarization), HONDO7 Polarization, Huzinaga Polarization, Dunning-Hay Diffuse, aug-cc-pVDZ Diffuse, aug-cc-pVTZ Diffuse, aug-cc-pVQZ Diffuse, aug-cc-pV5Z Diffuse, aug-cc-pV6Z Diffuse, aug-pV7Z Diffuse, d-aug-cc-pVDZ Diffuse, d-aug-cc-pVTZ Diffuse, d-aug-cc-pVQZ Diffuse, d-aug-cc-pV5Z Diffuse, d-aug-cc-pV6Z Diffuse, DHMS Polarization, Dunning-Hay Rydberg, Dunning-Hay Double Rydberg, Binning-Curtiss (1d Polarization), Binning-Curtiss (df) Polarization, Ahlrichs Polarization, Glendenning Polarization, Blaudeau Polarization, Core/val. Functions (cc-pCVDZ), Core/val. Functions (cc-pCVTZ), Core/val. Functions (cc-pCVQZ), Core/val. Functions (cc-pCV5Z).



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 *e.g., 3-21G, 6-31G, 6-311G*
 - 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 *e.g., 6-31G*, 6-311G(d,p)*
- 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 *e.g., 6-311++G*

Meet the basis sets: Dunning-style bases

- ➊ Pople-style basis sets generally optimized from HF calculations on atoms and small molecules
- ➋ Thom Dunning (UT,ORNL) utilized higher order correlated methods, CISD, hence “correlation-consistent”
- ➌ Designed to systematically converge the correlation energy to the basis set limit
- ➍ Basis set notation looks like: (aug)-cc-p(C)VXZ, X=D,T,Q,5,6,7
- ➎ Valence & Polarization functions
 - ➏ cc-pVXZ is a polarized valence X-zeta basis *e.g., cc-pVTZ*
 - ➏ 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
- ➏ Diffuse functions
 - ➏ 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* *e.g., aug-cc-pVTZ*
 - ➏ extra diffuse functions available with “d-aug”
- ➏ Core functions
 - ➏ letter “C” indicates presence of core correlation functions *e.g., cc-pCVTZ*

Meet the basis sets: Others

Atomic Natural Orbitals (ANO)

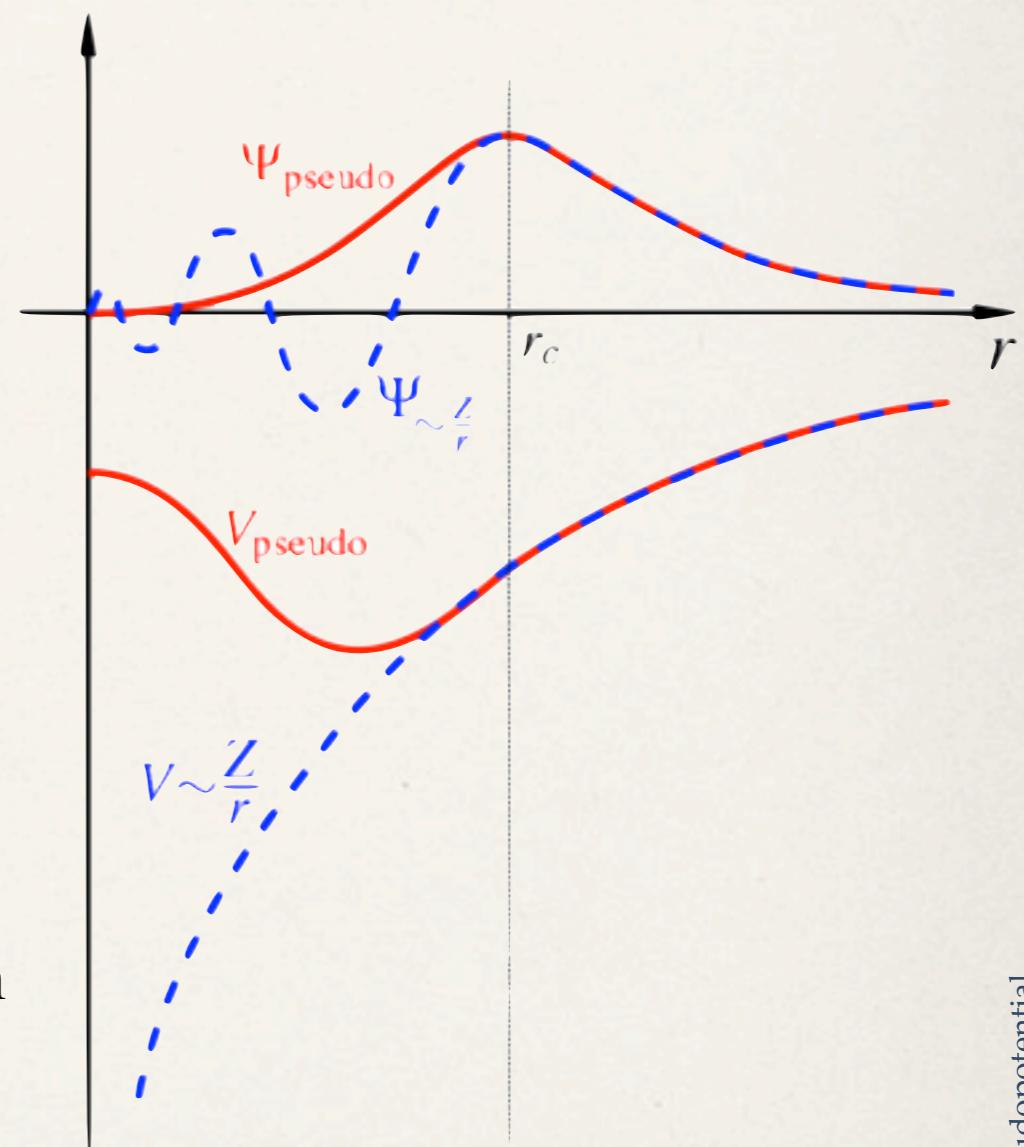
- J. Amlof and P. R. Taylor at NASA Ames
- ANO designed to reproduce the natural orbitals for correlated (CISD) calculations on atoms
- Generally contracted- more complicated scheme to implement
- well-suited for harmonic and anharmonic force fields of correlated methods
- Very large basis sets (consequently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

The screenshot shows the EMSL Basis Set Exchange interface. At the top, there's a navigation bar with links like PDF Database, TV Planner 2, Dessins DHC...ns and kits, CCL Jobs, Apple (212), News (1265), Chemistry, TV Planner, and Amazon Wish List. Below the navigation bar is the EMSL logo and the title "BASIS SET EXCHANGE" with a molecular icon. On the right side, there are fields for "Username" and "Password" with "Login" and "Become a Contributor" buttons. A green banner at the top says "Basis Set Exchange: v1.2.2". Below it are links for Feedback, About, ReleaseNotes, and Help. The main content area has a sidebar on the left with a list of basis set names, a search bar, and a dropdown for "Format" set to "NWChem". The main area displays a grid of elements from H to Lu. Below the grid, there's a "Get Basis Set" button. At the bottom, there's detailed information about the "aug-cc-pVDZ" basis set, including its summary, developer (N/A), last modified date (Mon, 15 Jan 2007 23:46:11 GMT), contributor (Dr. David Feller), curation status (published), and links to more information and user annotations.

"aug-cc-pVDZ" Basis Set Information

Summary: VDZ2PD Valence Double Zeta + Polarization + Diffuse
Primary Developer: N/A
Last Modified: Mon, 15 Jan 2007 23:46:11 GMT

Contributor: Dr. David Feller
Curation Status: published

[More information...](#)
[User annotations...](#)

When publishing results obtained from use of the Basis Set Exchange (BSE) software and the EMSL Basis Set Library, please cite:

The Role of Databases in Support of Computational Chemistry Calculations
Feller, D., J. Comp. Chem., 17(13), 1571-1586, 1996.

Basis Set Exchange: A Community Database for Computational Sciences
Schuchardt, K.L., Didier, B.T., Elsethagen, T., Sun, L., Gurumoorthi, V., Chase, J., Li, J., and Windus, T.L.
J. Chem. Inf. Model., 47(3), 1045-1052, 2007, doi:10.1021/ci600510j.

[Security and Privacy](#) | [Disclaimer](#)

Acquiring basis sets- the EMSL library

Basis Set: cc-pVDZ

BASIS SET EXCHANGE v1.2.2

```
! cc-pVDZ EMSL Basis Set Exchange Library 7/6/10 3:44 AM
! Elements References
! -----
! H      : T.H. Dunning, Jr. J. Chem. Phys. 90, 1007 (1989).
! He     : D.E. Woon and T.H. Dunning, Jr. J. Chem. Phys. 100, 2975 (1994).
! Li - Ne: T.H. Dunning, Jr. J. Chem. Phys. 90, 1007 (1989).
! Na - Mg: D.E. Woon and T.H. Dunning, Jr. (to be published)
! Al - Ar: D.E. Woon and T.H. Dunning, Jr. J. Chem. Phys. 98, 1358 (1993).
! Ca      : J. Koput and K.A. Peterson, J. Phys. Chem. A, 106, 9595 (2002).
!

*****
H      0
S      3   1.00
        13.0100000    0.0196850
        1.9620000    0.1379770
        0.4446000    0.4781480
S      1   1.00
        0.1220000    1.0000000
P      1   1.00
        0.7270000    1.0000000
*****
```

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)

Basis Set: cc-pVDZ

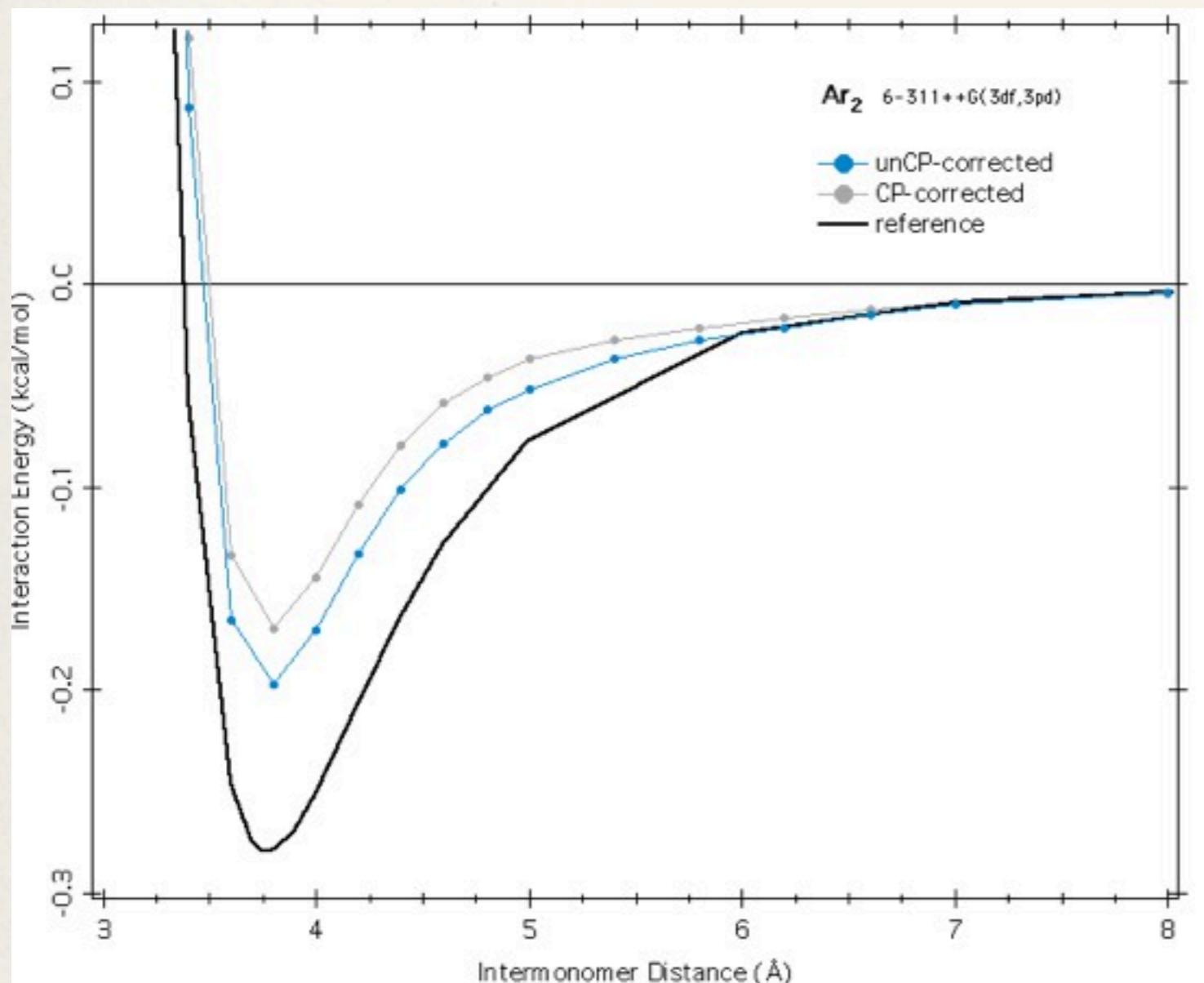
BASIS SET EXCHANGE v1.2.2

```
! cc-pVDZ EMSL Basis Set Exchange Library 7/6/10 10:54 AM
! Elements References
! -----
! H      : T.H. Dunning, Jr. J. Chem. Phys. 90, 1007 (1989).
! He     : D.E. Woon and T.H. Dunning, Jr. J. Chem. Phys. 100, 2975 (1994).
! Li - Ne: T.H. Dunning, Jr. J. Chem. Phys. 90, 1007 (1989).
! Na - Mg: D.E. Woon and T.H. Dunning, Jr. (to be published)
! Al - Ar: D.E. Woon and T.H. Dunning, Jr. J. Chem. Phys. 98, 1358 (1993).
! Ca     : J. Koput and K.A. Peterson, J. Phys. Chem. A, 106, 9595 (2002).
!

*****
C    0
S    8   1.00
       6665.0000000      0.0006920
       1000.0000000      0.0053290
       228.0000000      0.0270770
       64.7100000      0.1017180
       21.0600000      0.2747400
       7.4950000      0.4485640
       2.7970000      0.2850740
       0.5215000      0.0152040
S    8   1.00
       6665.0000000     -0.0001460
       1000.0000000     -0.0011540
       228.0000000     -0.0057250
       64.7100000     -0.0233120
       21.0600000     -0.0639550
       7.4950000     -0.1499810
       2.7970000     -0.1272620
       0.5215000      0.5445290
S    1   1.00
       0.1596000      1.0000000
P    3   1.00
       9.4390000      0.0381090
       2.0020000      0.2094800
       0.5456000      0.5085570
P    1   1.00
       0.1517000      1.0000000
D    1   1.00
       0.5500000      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)

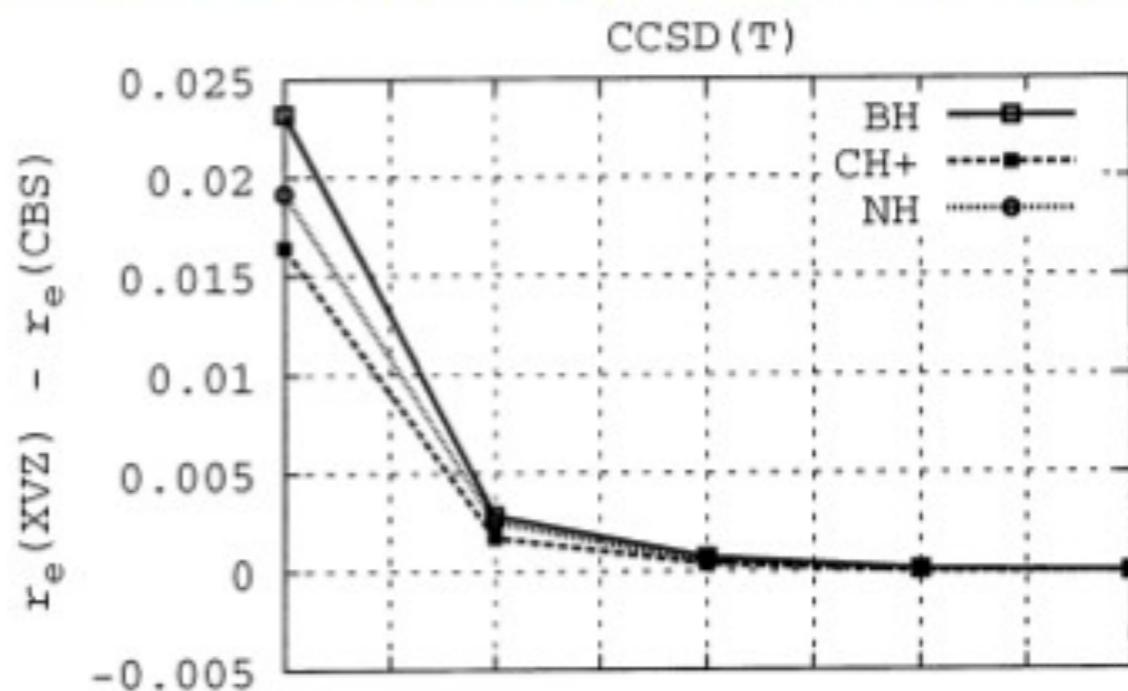


$$E_{unCP} = E_{\text{dimer}}^{AB} - E_{\text{mono A}}^A - E_{\text{mono B}}^B$$

$$E_{CP} = E_{\text{dimer}}^{AB} - E_{\text{mono A}}^{AB} - E_{\text{mono B}}^{AB}$$

Properties and basis sets

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



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

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

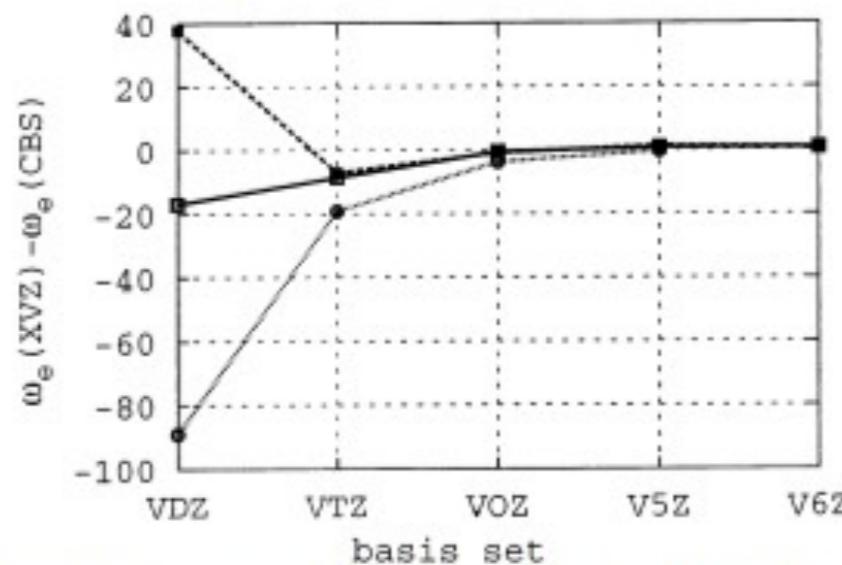


Figure 1. Convergence of CCSD(T) r_e and ω_e toward the CBS limit derived for valence-only (cc-pVNZ) basis sets.

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

Balancing methods and basis sets

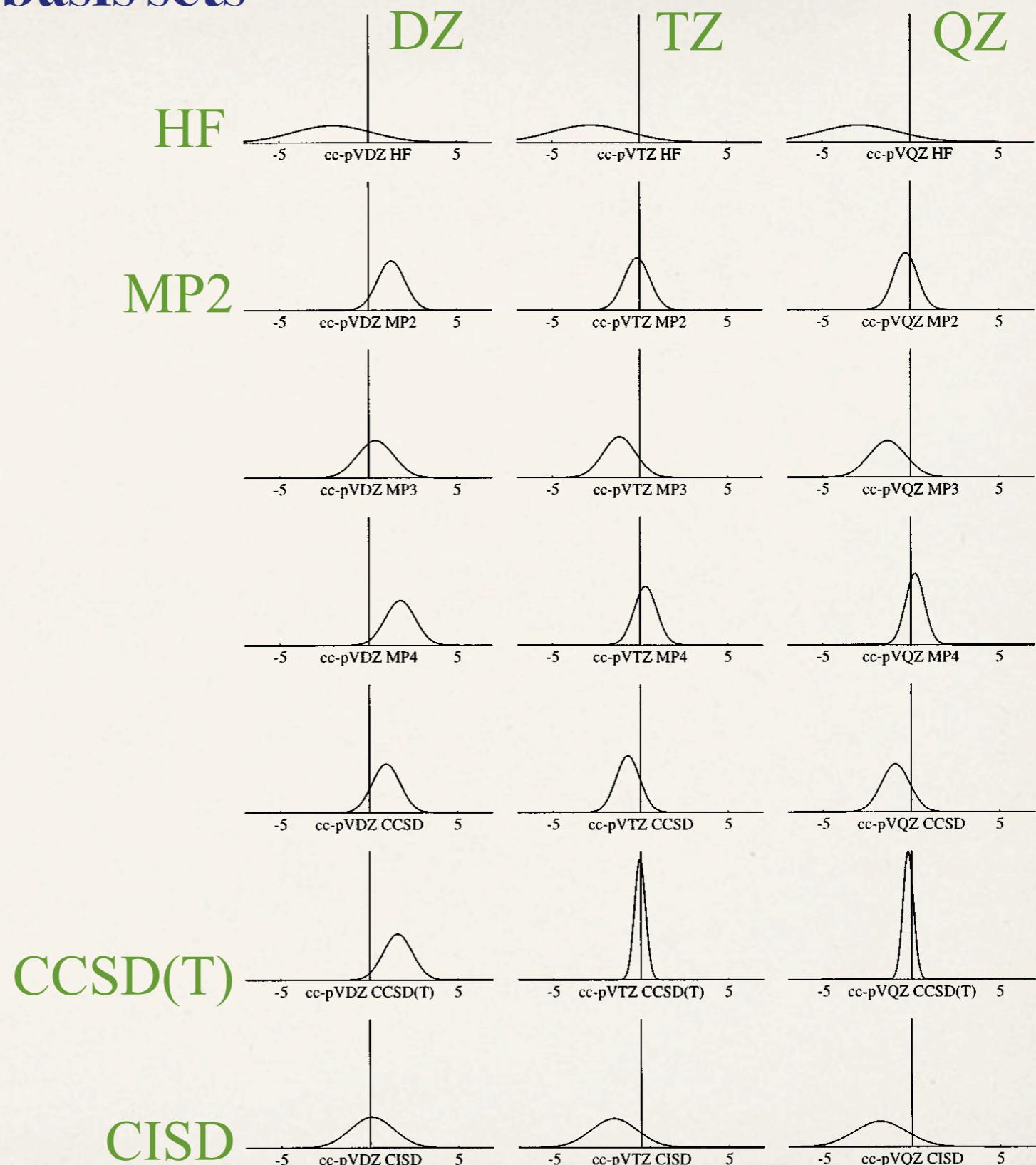
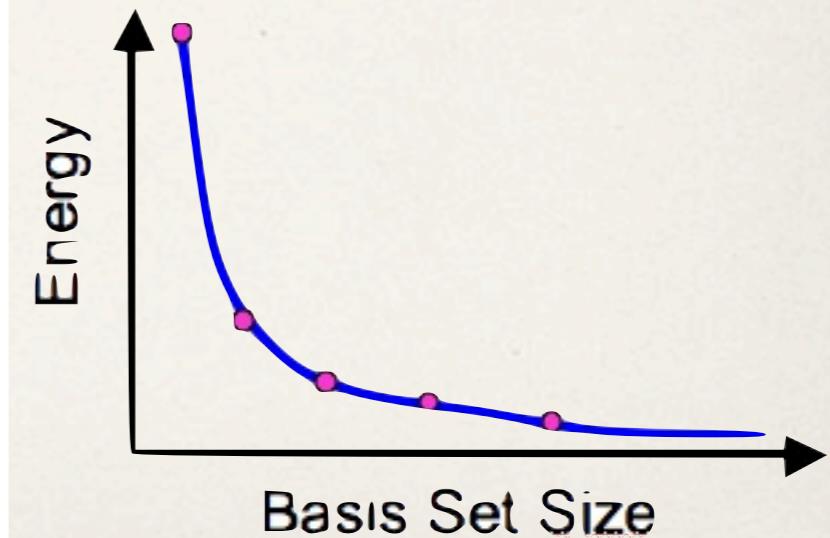
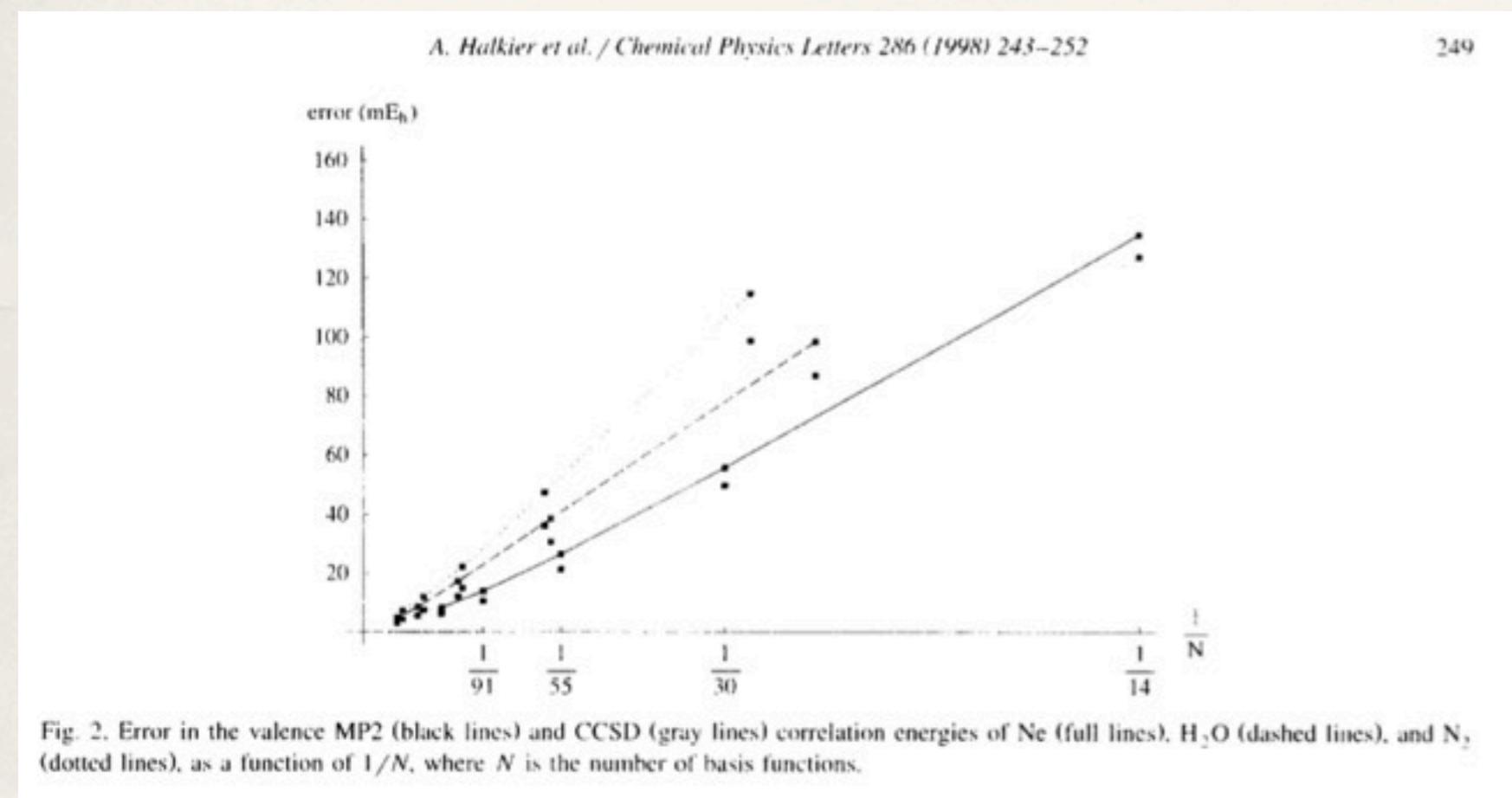


FIG. 3. Normal distributions $\rho(R)$ for the errors in the calculated bond distances. The distributions have been calculated from the mean errors and standard deviations in Table IV (pm). For easy comparison, all distributions have been normalized to one and plotted on the same horizontal scale.

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



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 $E_{\text{CCSD(T)}}^{\text{CBS}} = E_{\text{HF}}^{\text{CBS}} + E_{\text{corr CCSD(T)}}^{\text{CBS}}$
- Two schemes
 - Exponential Form $E_X^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + A \exp(-\alpha X)$ and $E_X^{\text{corr}} = E_{\text{CBS}}^{\text{corr}} + BX^{-3}$
 - Power Form $E_X^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + BX^{-\alpha}$ and $E_X^{\text{corr}} = E_{\text{CBS}}^{\text{corr}} + BX^{-\beta}$.
- 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 $E_{\text{CCSD(T)}}^{\text{CBS}} \approx E_{\text{MP2}}^{\text{CBS}} + \Delta_{\text{CCSD(T)}} = E_{\text{HF}}^{\text{CBS}} + E_{\text{corr MP2}}^{\text{CBS}} + (E_{\text{corr CCSD(T)}}^{\text{small}} - E_{\text{corr MP2}}^{\text{small}})$
 - 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 $E_{\text{CCSD(T)}}^{\text{CBS}} = E_{\text{HF}}^{\text{CBS}} + E_{\text{corr CCSD(T)}}^{\text{CBS}}$
- Two schemes
 - Exponential Form $E_X^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + A \exp(-\alpha X)$ and $E_X^{\text{corr}} = E_{\text{CBS}}^{\text{corr}} + BX^{-3}$
 - Power Form $E_X^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + BX^{-\alpha}$ and $E_X^{\text{corr}} = E_{\text{CBS}}^{\text{corr}} + BX^{-\beta}$.
- X^{-3} term remedies DZ error but is not recommended
 - DZ is erratic so avoid
 - Two-step extrapolation is affordable at TZ & QZ but extrapolation at the MP2 level is not the best
 - Use three-step extrapolation HF and MP2 correction terms $\delta\text{CCSD(T)}$ between MP2 and CBS
 - With large enough basis sets to recover correlation energy of the CBS limit

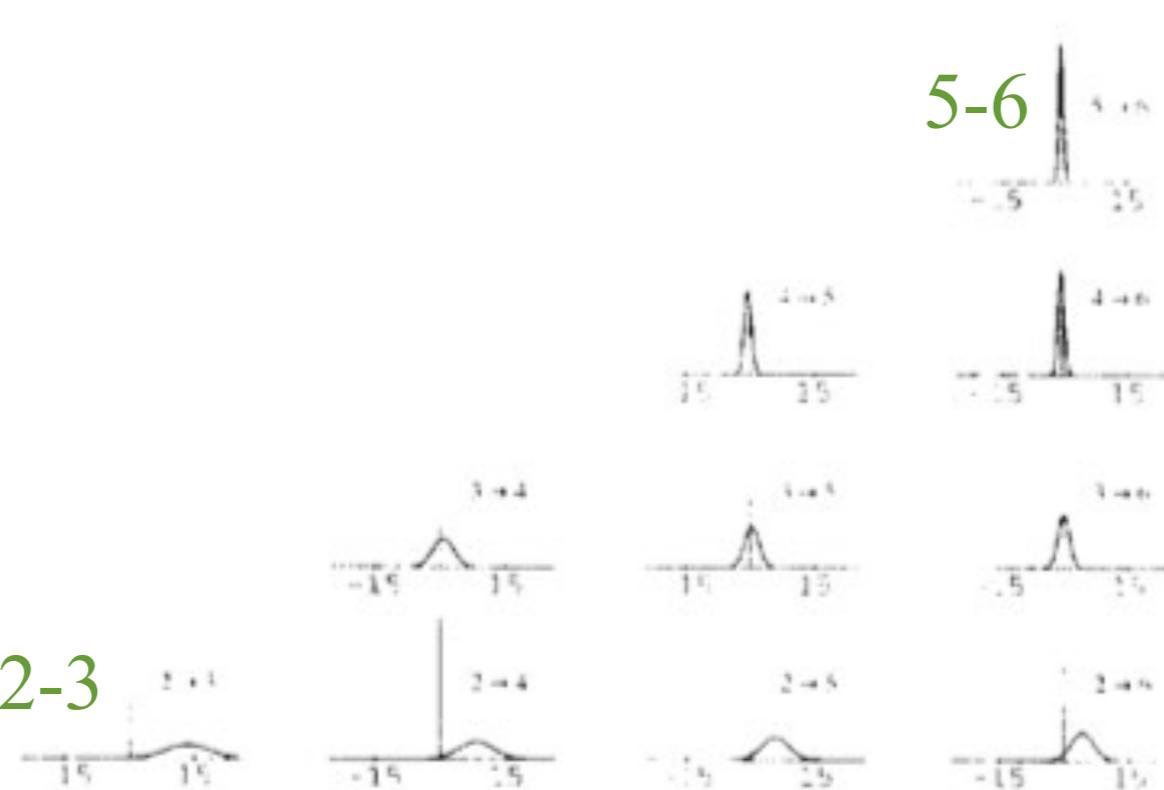


Fig. 3. Normal distributions of the two-parameter fits $E_X^{\text{corr}} = E_{\text{corr}}^{\text{small}} + AX^{-3}$ as a function of the data fitted (from X_{min} to X_{max}).

● CP-correction

● CCSD(T) is extrapolation at the MP2 level

$S_{\text{MP2}} + (E_{\text{corr CCSD(T)}}^{\text{small}} - E_{\text{corr MP2}}^{\text{small}})$

● to recover correlation

● 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 $E_{\text{CCSD(T)}}^{\text{CBS}} = E_{\text{HF}}^{\text{CBS}} + E_{\text{corr CCSD(T)}}^{\text{CBS}}$
- Two schemes
 - Exponential Form $E_X^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + A \exp(-\alpha X)$ and $E_X^{\text{corr}} = E_{\text{CBS}}^{\text{corr}} + BX^{-3}$
 - Power Form $E_X^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + BX^{-\alpha}$ and $E_X^{\text{corr}} = E_{\text{CBS}}^{\text{corr}} + BX^{-\beta}$.
- 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 $E_{\text{CCSD(T)}}^{\text{CBS}} \approx E_{\text{MP2}}^{\text{CBS}} + \Delta_{\text{CCSD(T)}} = E_{\text{HF}}^{\text{CBS}} + E_{\text{corr MP2}}^{\text{CBS}} + (E_{\text{corr CCSD(T)}}^{\text{small}} - E_{\text{corr MP2}}^{\text{small}})$
 - 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)