

Basis Sets and Performance

(Lecture 3)

*NSF Computational Nanotechnology and Molecular Engineering
Pan-American Advanced Studies Institutes (PASI) Workshop
January 5-16, 2004*

California Institute of Technology, Pasadena, CA

Andrew S. Ichimura

Basis Set Approximation

MOs are expanded in terms of Atomic Orbitals

$$\phi_i = \sum_{\alpha}^M c_{\alpha i} \chi_{\alpha}$$

LCAO - MO representation

Coefficients are variational parameters

- ϕ_i (MO) is initially unknown; describing (expanding) the MO as a combination of known (χ) AO functions.
- As $M \rightarrow \infty$, reach the complete basis set limit; not an approximation.
- When M is finite, the representation is approximate.

Two criteria for selecting basis functions.

- i) They should be physically meaningful.
- ii) computation of the integrals should be tractable.

Slater Type Orbitals (STO)

$$\chi_{\zeta,n,l,m}(r,\theta,\phi) = NY_{l,m}(\theta,\phi)r^{n-1}e^{-\zeta r}$$

STO depends on quantum numbers n,l,m and **zeta, ζ** .

$Y_{l,m}(\theta,\phi)$ Spherical harmonics; N - normalization

Advantages:

1. Physically, the exponential dependence on distance from the nucleus is very close to the exact hydrogenic orbitals.
2. Ensures fairly rapid convergence with increasing number of functions.

Disadvantages:

1. Three and four center integrals cannot be performed analytically.
2. No radial nodes. These can be introduced by making linear combinations of STOs.

Practical Use:

1. Calculations of very high accuracy, atomic and diatomic systems.
2. Semi-empirical methods where 3- and 4-center integrals are neglected.

Gaussian Type Orbitals (GTO)

$$\chi_{\zeta,n,l,m}(r,\theta,\phi) = NY_{l,m}(\theta,\phi)r^{2n-2-l}e^{-\zeta r^2} \quad \text{Polar coordinates}$$

GTO depends on quantum numbers n, l, m and exponent zeta, ζ .

d -function has five components ($Y_{2,2}, Y_{2,1}, Y_{2,0}, Y_{2,-1}, Y_{2,-2}$).

$$\chi_{\zeta,l_x,l_y,l_z}(x,y,z) = Nx^{l_x}y^{l_y}z^{l_z}r^{2n-2-l}e^{-\zeta r^2} \quad \text{Cartesian coordinates}$$

In Cartesian coords., the angular dependence of the GTO is computed from the sum of l_x, l_y , and l_z ($l_x+l_y+l_z=1$, a p -orbital).

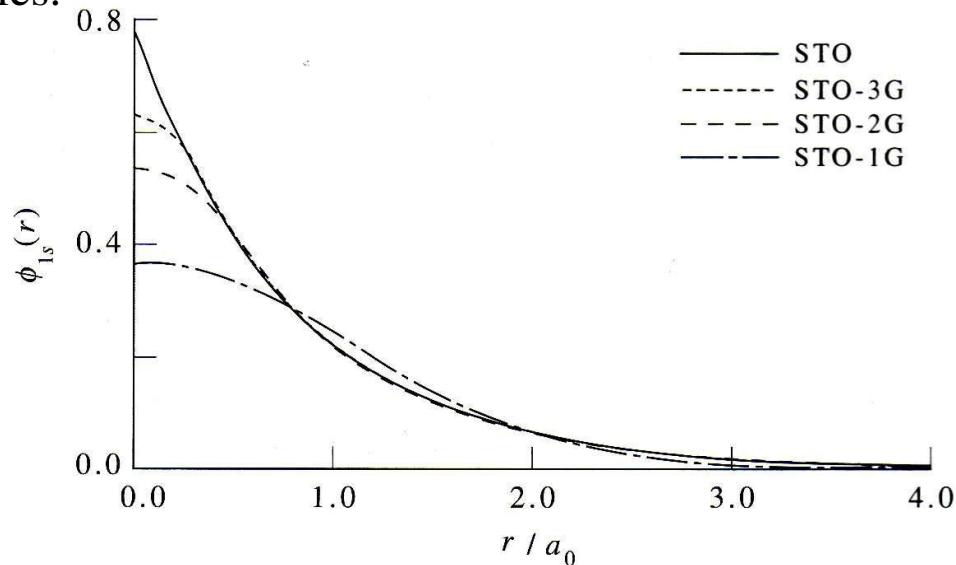
- d -function has six components ($x^2, y^2, z^2, xy, xz, yz$) in cartesian coord. These may be transformed to spherical functions plus one extra s -type function: $(x^2+y^2+z^2)$.
- f -orbitals have 10 components, which may be transformed to the 7-‘pure’ spherical ones plus 3 p -type functions.

GTOs are inferior to STOs in three ways:

1. At the nucleus, the GTO has zero slope; the STO has a cusp. Behavior near the nucleus is poorly represented.
2. GTOs diminish too rapidly with distance. The 'tail' behavior is poorly represented.
3. Extra d -, f -, g -, etc. functions (from Cart. rep.) may lead to linear dependence of the basis set. They are usually dropped when large basis sets are used.

Advantage:

GTOs have analytical solutions. Use a linear combination of GTOs to overcome these deficiencies.



Classification

Minimum basis: Only enough functions are used to contain the the electrons of the neutral atoms (usually core plus valence orbitals).

1st row: 1s, 2s, 2p 5-AOs

2nd row: 1s, 2s, 3s, 2p, 3p 9-AOs

Double Zeta (DZ) basis: Double the number of all basis functions.

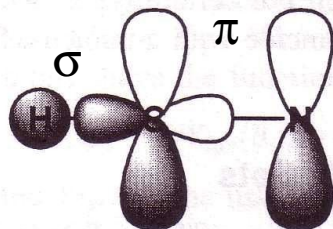
Hydrogen has two 1s-functions: 1s and 1s' 2-AOs

Li-Ne: 1s and 1s', 2s and 2s', 2p and 2p' 10-AOs

Think of 1s and 1s' as 'inner' and 'outer' functions. The inner function has larger ζ exponent and is \therefore tighter, outer 1s' has a smaller ζ , more diffuse.

DZ basis yields a better description of the charge distribution compared to a minimal basis.

Consider HCN,



Charge distributions are different in different parts of the molecule.

⇒ C-H σ -bond consists of the H 1s orbital and the C $2p_z$.
CN π -bond is made up of C and N $2p_x$ (and $2p_y$) AOs.

⇒ Because the π -bond is more diffuse, the optimal exponent ζ for p_x (p_y) should be smaller than that for the more localized p_z orbital.

DZ basis has the flexibility (while the minimal basis does not) to describe the charge distribution in both parts of the molecule. The optimized AO coefficient (in MO expansion) of the ‘tighter’ inner p_z function on carbon will be larger in the C-H bond. The more diffuse outer p_x and p_y functions will have larger AO coefficients in the π -bond.

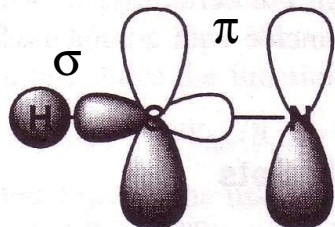
Split Valence Basis Sets

- Doubling the number of functions provides a much better description of bonding in the valence region.
- Doubling the number of functions in the core region improves the description of energetically important but chemically uninteresting core electrons.
- Split valence basis sets improve the flexibility of the valence region and use a single (contracted) set of functions for the core.

VDZ	double zeta	2x	number of basis functions in valence region
VTZ	triplet zeta	3x	“
VQZ	quadruple zeta	4x	“
V5Z	quintuple zeta	5x	“
V6Z	sextuple zeta	6x	“

Polarization Functions

Consider HCN,

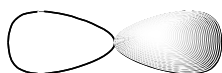


H-C σ -bond:

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



+



H 1s

H 2p_z

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

p-functions induce a **polarization** of s-orbitals.

d-function induce polarization of p-orbitals, etc.

For a single determinantal wavefunction, the 1st set of polarization functions is by far the most important and will describe most if not all of the important charge polarization effects.

Polarization Functions cont.

- To describe charge polarization effects at the SCF level, add
 - P-functions to H (one set)
 - D-functions to Li-Ne, Na-Ar (one set 1st row, 1-2 sets for 2nd row)
- To recover a larger fraction of the dynamical correlation energy, multiple functions of higher angular momentum (d, f, g, h, i...) are essential.

Electron correlation - energy is lowered by electrons avoiding each other.

Two types:

- 1) **Radial correlation** - two electrons, one close to the nucleus the other farther away. Need basis functions of the same type but different exponent. (tight and diffuse p-functions, for example)
- 2) **Angular correlation** - Two electrons on opposite sides of the nucleus. Basis set needs functions with the same exponent but different angular momentum. For s-functions, need p-functions (and d, f, g..) to account for angular correlation.

Radial \approx Angular in importance.

Diffuse Functions

Diffuse functions, **s-**, **p-**, and d-functions with small exponents are usually added for specific purposes.

- (1) Calculations on anions.
- (2) Dipole moment
- (3) Polarizability

Contracted Basis Sets

Energy optimized basis sets have a disadvantage. Many functions go toward representing the energetically important but chemical uninteresting core electrons.

Suppose 10s functions have been optimized for carbon.

Start with 10 primitive gaussians
PGTOs

End with 3 contracted gaussians
CGTOs

Inner 6 describe core 1s electrons
Next 4 describe valence electrons

Contract to one 1s function
contract to two 2s functions

$$\chi(\text{CGTO}) = \sum_i^k a_i \chi_i (\text{PGTO})$$

Energy always increases! Fewer variational parameters. But, less CPU time required.

Pople Style Basis Sets

STO-nG Minimal basis, n=# of gaussian primitives contracted to one STO.

k-nlmG

Split valence basis sets**

3-21G

Contraction scheme (6s3p/3s) -> [3s2p/2s]
(1st row elements /H)

3 PGTOs contracted to 1, forms core

2PGTOs contracted to 1, forms inner valence

1 PGTO, forms outer valence

After contraction of the PGTOs, C has 3s and 2p AOs.

6-31G

(10s4p/4s) -> [3s2p/2s] Valence double zeta basis

6-311G

(11s5p/4s) -> [4s3p/3s] Valence triple zeta basis

6-31+G*

Equivalent to 6-31+G(d). 6-31G basis augmented with diffuse sp-functions on heavy atoms, polarization function (d) on heavy atoms.

6-311++G(2df,2pd)

Triplet split valence; augmented with diffuse sp- on heavy atoms and diffuse s- on H's. Polarization functions 2d and 1f on heavy atoms; 2p and 1d on H's.

(**In the Pople scheme, s- and p-functions have the exponent. 6-31G(d,p) most common)

Huzinaga Basis Sets and Contractions

Huzinaga determined uncontracted energy optimized basis sets for the 1st row up to (10s,6p). First high quality basis sets published.

Later extended to,

(14s,9p) by van Duijneveldt

(18s,13p) by Partridge

Dunning used the Huzinaga basis sets in various contraction schemes:

(9s,5p/4s) -> [4s,2p/2p] (This DZ contraction, when used with polarization functions, is one of the first high quality contracted sets and perhaps the most highly cited.)
(6,1,1,1; 4,1 / 3,1)

(10s6p/5s) -> [5s,3p/3s] TZ valence
(5,3,1,1; 4,1,1 / 3,1,1)

McClellan and Chandler also used Huzinaga PGTOs

(12s,8p) -> [5s,3p] DZ

(13s,9p) -> [6s,4p] TZ

Tatewaki and Huzinaga developed minimum basis sets for most of the periodic table: MINI, MIDI, MAXI

Basis sets for Electron Correlation

Atomic Natural Orbital (ANO) basis; Amlof and Taylor

Correlation consistent (cc) basis sets; T.H. Dunning

Advantages:

- Recover a large fraction of the correlation energy.
- Provide systematic improvements that converge toward the complete basis set limit.
- Consistently reduce errors at both the HF and correlated levels with each step up in quality.

Disadvantage:

- The number of basis functions ~doubles with each increase in quality.

Basis	Primitive functions	Contracted functions	Valence E_{CE}
cc-pVDZ	9s,4p,1d/4s,1p	3s,2p,1d/2s,1p	~65%
cc-pVTZ	10s,5p,2d,1f/5s,2p,1d	4s,3p,2d,1f/3s,2p,1d	~85%
cc-pVQZ	12s,6p,3d,2f,1g/6s,3p,2d,1f	5s,4p,3d,2f,1g/4s,3p,2d,1f	~93%
cc-pV5Z	14s,9p,4d,3f,2g,1h/8s,4p,3d,2f,1g	6s,5p,4d,3f,2g,1h/5s,4p,3d,2f,1g	~95%
cc-pV6Z	16s,10p,5d,4f,3g,2h,1i/ 10s,5p,4d,3f,2g,1h	7s,6p,5d,4f,3g,2h,1i/ 6s,5p,4d,3f,2g,1h	~98%

Functions with high angular momentum (f,g,h,i) are necessary to recover E_{CE} .

Nomenclature

A compact notation suggested by Pople is commonly used to describe the hamiltonian plus basis set, which together represent the wavefunction.

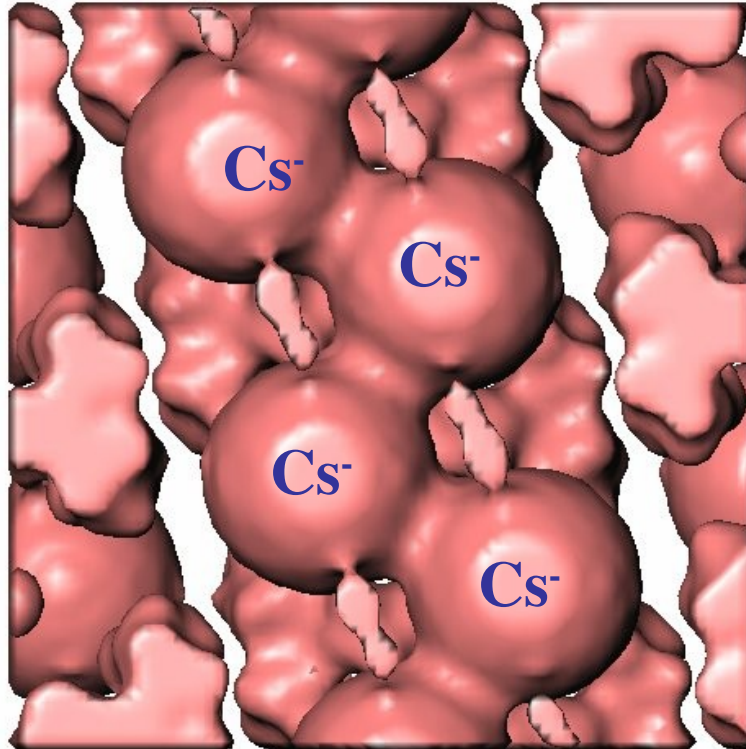
“model/basis set//model/basis set”

- **Model** refers to the type of Ψ (RHF, MP2, CCSD, B3LYP, etc.) and the **basis set** is written in the shorthand that designates the contracted set of AO functions.
- The **double slash ‘//’** separates the higher level *single point calculation* (usually to recover the correlation energy) from the level at which the geometry was optimized, respectively.

For example:

MP2/6-311g(2df,2pd)//RHF/6-31G(d,p)

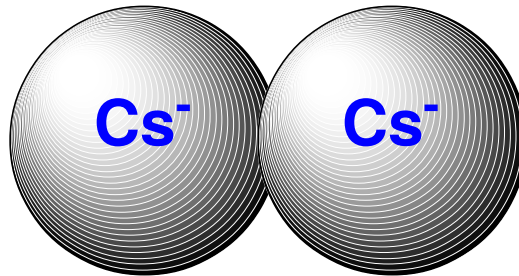
Denotes: The geometry of the molecule was first optimized with the Restricted HF method employing the standard Pople type basis, 6-31G with one set of ‘d-’polarization functions on heavy atoms (non-hydrogen) and one set of ‘p-’ functions on H’s. Then the correlation energy was computed with the MP2 method and a more extended basis set, the valence triple zeta 6-311G basis plus polarization functions on heavy atoms (two sets of ‘d’ and one set of ‘f’ functions) and hydrogens (two sets of p- and one set of d-functions).



Use *ab initio* quantum chemistry to gain insight into the nature of bonding between Cs^- in ceside chains.

Develop (find) a Basis Set for Cs⁻

Set out to describe
the Cs₂²⁻ dimer.



- Visit Pacific Northwest National Laboratory basis set website.
Use the 'Gaussian Basis Set Order Form': www.emsl.pnl.gov

Two options:

- (1) Full Atomic Orbital basis - all electron wavefunction.
Huzinaga well-tempered Cs-basis (30s, 23p, 17d) -> [6s, 5p, 2d]
- (2) Effective Core Potential (ECP) or Pseudopotential
LANL2DZ-ECP
Stuttgart-ECP

Effective Core Potential (ECP)

Two good reasons to use ECP:

- (1) A balanced basis requires a proper description of the core and valence regions. For Cs, there are a lot of core electrons (1st-5th periods)! Most of the computational effort is used to describe the energy but not the valence region.
- (2) For large Z , relativistic effects complicate matters.

Solve both by using an ECP: Core electrons are modeled by a suitable potential function, and only the valence electrons are treated explicitly.

LANL2DZ-ECP
Stuttgart-ECP



In the case of the Cs-ECPs, both also include the 5s and 5p filled shells explicitly. The rest are considered core electrons.

A slight problem arises...

The LANL2DZ-ECP and Stuttgart-ECP basis sets were parameterized (core electrons) for the atom, but valence Gaussian exponents were optimized for Cs^0 or Cs^+ . (Cs^- is the largest monatomic anion, $r=3.5\text{\AA}$)

Solution:

1. Create an even tempered set of diffuse functions (s, p, d) to model the anion.
2. Energy optimize polarization functions (f, g) at the CISD level.

LANL2DZ-ECP (8s,6p) \rightarrow [3s,3p] augmented by:

- Borrow 2 d-functions from Lanthanum
- Add up to 4 diffuse-s, 3 diffuse-p, 2 diffuse-d
- Add polarization functions: up to 2f, 1g

Largest basis set: (12s, 9p, 4d, 2f, 1g) \rightarrow [7s, 6p, 3d, 2f, 1g]

Compute the Electron Affinity of Gas Phase Cs⁰

CCSD(T) wavefunction

Electron Affinity (eV)

0.170

0.395

0.454

0.455

0.404

0.462

0.462

0.471

Additional diffuse 's'
functions necessary

Only 8 mH improvement
with extra d, f, g

Only 0.009 eV difference from expt. ~2 % error, < 1 kcal/mol

E.A. of Cs atom as Function of Model

Electron Affinity (eV)

-0.067

Unbound at HF
level

0.249

0.280

0.382

0.437

0.438

0.462

0.471

Best Variational result: within 7% (MR-CISD)

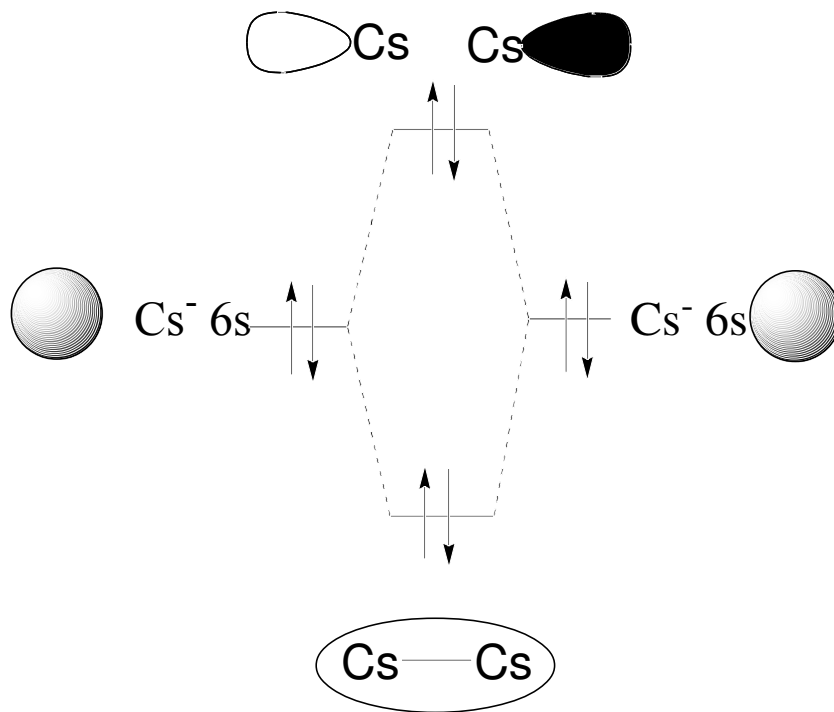
Best non-Variational result: within 2% CCSD(T)

Comparison between Expt. and Theory

Three points:

1. There is excellent agreement between experiment and theory (CCSD(T) or MR-CISD) for of the E.A. for the atom and molecule Cs_2 , the bond length and fundamental frequency of Cs_2 .
2. Therefore, this basis set is likely to be suitable to describe the bonding in Cs_2^{2-} dimers and higher order clusters.
3. Justifies the use of ECPs since the comparison between expt and theory is quite good and what we are interested in is the charge distribution in the valence region.

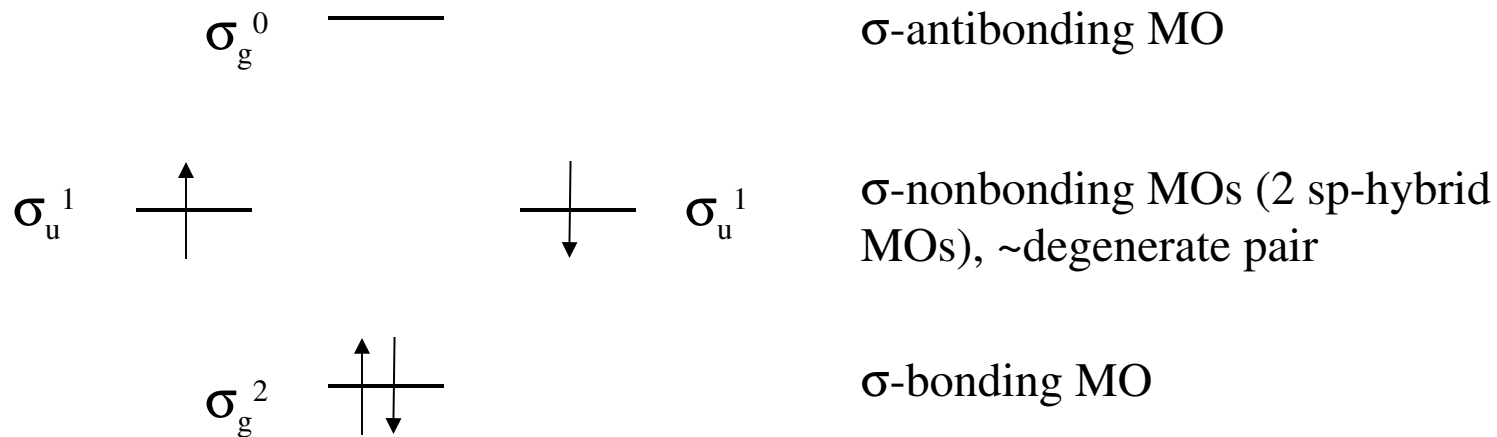
Qualitative MO Picture of Cs_2^{2-}



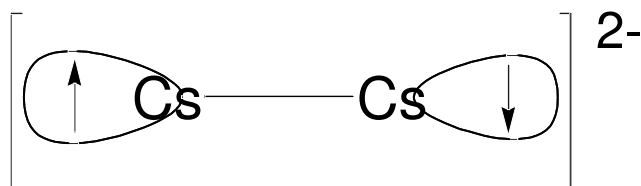
Should not be bound!!

To summarize:

MO Picture of Cs_2^{2-}



VB Picture of Cs_2^{2-}



Electrons from the two extra charges are localized to opposite ends of the molecule to reduce Coulomb repulsion.

Systematic Comparisons: H₂O Geometry

SCF			MP2		
Basis	$R_{\text{OH}} (\text{Å})$	α_{HOH}	Basis	$R_{\text{OH}} (\text{Å})$	α_{HOH}
cc-pVDZ	0.9463	104.61	cc-pVDZ	0.9649	101.90
cc-pVTZ	0.9406	106.00	cc-pVTZ	0.9591	103.59
cc-pVQZ	0.9396	106.22	cc-pVQZ	0.9577	104.02
cc-pV5Z	0.9396	106.33	cc-pV5Z	0.9579	104.29
cc-pV6Z	0.9396	106.33	cc-pV6Z	0.9581	104.36

CCSD(T)				
Basis	$R_{\text{OH}} (\text{Å})$	α_{HOH}	$\Delta R_{\text{OH}} (\text{Å})$	$\Delta \alpha_{\text{HOH}}$
cc-pVDZ	0.9663	101.91	0.0014	0.01
cc-pVTZ	0.9594	103.58	0.0003	0.06
cc-pVQZ	0.9579	104.12	0.0002	0.10
cc-pV5Z	0.9580	104.38	0.0001	0.09

Expt. $r(\text{OH}) = 0.9578 \text{ Å}$; $\theta = 104.48^\circ$

Convergence of Correlation Energy (H₂O)

Table 11.7 % electron correlation recovered by different methods in the cc-pVDZ basis

Method	% EC
MP2	94.0
MP3	97.0
MP4	99.5
MP5	99.8
CCSD	98.3
CCSD(T)	99.7
CISD	94.5
CISDT	95.8
CISDTQ	99.9

Table 11.8 Total energy (+76 a.u.) as a function of basis set and electron correlation (valence only)

Method	cc-pVDZ	cc-pVTZ	cc-pVQZ	cc-pV5Z	cc-pV6Z	cc-pV ∞ Z
HF	-0.02677	-0.05713	-0.06479	-0.06704	-0.06735	-0.0676
MP2	-0.22844	-0.31863	-0.34763	-0.35860	-0.36264	-0.368
MP3	-0.23544	-0.32275	-0.34939	-0.35815	-0.36094	-0.364
MP4	-0.24067	-0.33302	-0.36104	-0.37051	-0.37357	-0.377
MP5	-0.24120	-0.33159				
CCSD	-0.23801	-0.32455	-0.35080	-0.35952		-0.366
CCSD(T)	-0.24104	-0.33219	-0.35979	-0.36904		-0.376
CISD	-0.22997	-0.31384	-0.33922	-0.34765		-0.354

H₂O Dipole Moment

Table 11.10 H₂O dipole moment (Debye) as a function of theory (valence correlation only), experimental value is 1.847 D

Basis	HF	MP2	CCSD(T)	
cc-pVDZ	2.057	1.964	1.936	
cc-pVTZ	2.026	1.922	1.903	
cc-pVQZ	2.008	1.904	1.890	
cc-pV5Z	2.003	1.895		Diffuse functions essential
cc-pV6Z	1.990			
aug-cc-pVDZ	2.000	1.867	1.848	
aug-cc-pVTZ	1.984	1.852	1.839	
aug-cc-pVQZ	1.982	1.858	1.848	
aug-cc-pV5Z	1.982	1.861		

H₂O Harmonic Frequencies

		Basis	ω_1	ω_2	ω_3	
HF		cc-pVDZ	4212	4114	1776	~ 6% too high
		cc-pVTZ	4227	4127	1753	
		cc-pVQZ	4229	4130	1751	
		cc-pV5Z	4231	4131	1748	
<hr/>						
		Basis	ω_1	ω_2	ω_3	
MP2		cc-pVDZ	3971	3852	1678	Excellent agreement
		cc-pVTZ	3976	3855	1651	
		cc-pVQZ	3978	3855	1643	
		cc-pV5Z	3974	3849	1636	
		Expt.	3943 cm ⁻¹ , 3832 cm ⁻¹ , 1649 cm ⁻¹			'f'-function improves bending freq.
		Basis	ω_1	ω_2	ω_3	
CCSD(T)		cc-pVDZ	3928	3822	1690	Excellent agreement
		cc-pVTZ	3946	3841	1669	
		cc-pVQZ	3952	3845	1659	

Ozone, a problematic system...

Ozone Harmonic Frequencies

Table 11.23 Harmonic frequencies for O₃ with the cc-pVTZ basis

Method	ω_1	ω_2	ω_3
HF	1537	1418	867
MP2	1166	2241	743
MP3	1364	1713	798
MP4	1106	1592	695
CCSD	1278	1267	762
CCSD(T)	1154	1067	717
CISD	1407	1535	816
[2,2]-CASSCF	1189	1497	799
SVWN	1249	1148	744
BLYP	1130	980	683
BPW91	1177	1047	706
B3LYP	1252	1194	746
B3PW91	1288	1244	762
Experimental	1135	1089	716

← Excellent agreement

Summary:

- The term ‘chemical accuracy’ is used when a calculation has an error of ~ 1 kcal/mol. Chemical accuracy for almost any property of interest is best achieved with highly correlated wavefunctions (e.g. CCSD(T) or MR-CISD) and large basis sets (cc-pVTZ and higher), and is only practical for small molecules.
- MP2 methods perform well for many properties of interest (geometry prediction, frequencies, dipole moment, ...), is size extensive, recovers a good fraction of the correlation energy (80-90%), and is applicable to modest sized systems that contain 20 carbon atoms or more even with triple zeta basis sets, (cc-pVTZ). MP2 does not perform well when the unperturbed state is multi-reference in nature (e.g. O_3). Multi-reference MP2 methods may be used in this case.
- Consult references at end for extended discussion.

References:

Much of this presentation was drawn from:

Jensen, F.; *Introduction to Computational Chemistry*, John Wiley & Sons: Chichester, 1999.

Another suitable reference for electronic structure theory:

Szabo, A.; Ostlung, N.A. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, McGraw-Hill: New York, 1989.

Quantum Chemistry texts:

McQuarrie, D. A.; Simon, J. A. *Physical Chemistry, A Molecular Approach*; University Science Books: Sausalito, 1997.

Lowe, J.P. *Quantum Chemistry*, Academic Press Inc.: New York, 1978.

Basis Sets:

Feller, D.; Davidson, E.R. *Rev. Comput. Chem*, **1990**, 1 .

Davidson, E.R.; Feller, D. *Chem. Rev.* **1986**, 86, 681-696.

Helgaker, T.; Taylor, P.R. *Modern Electronic Structure Theory, Part II*, ed. D.Yarkony, World Scientific: 1995.

Gaussian Basis Sets on the web: <http://www.emsl.pnl.gov/forms/basisform.html>

Acknowledgements

National Science Foundation
California Institute of Technology

Dr. Mario Blanco

Dr. William Goddard, III