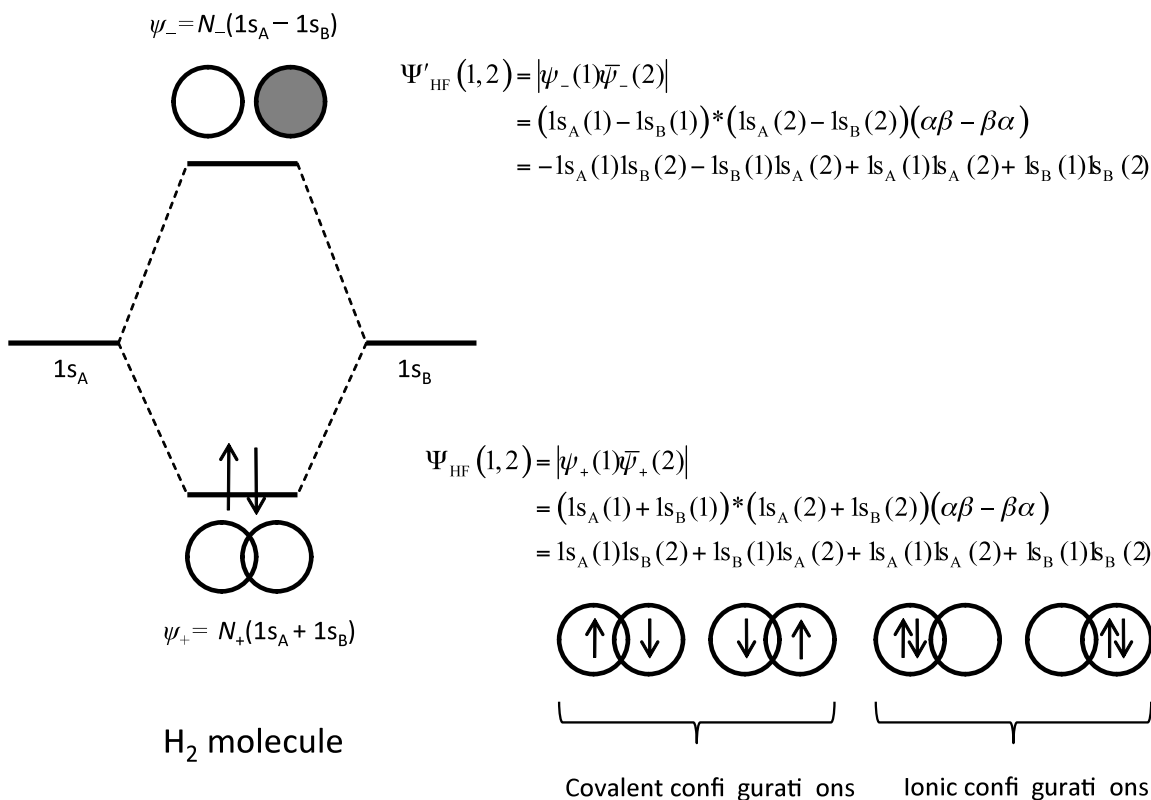


## 1. Electron Correlation

Hartree-Fock treats electron-electron repulsion in an average way; exchange ensures spin-aligned electrons generally avoid one another, but the negative cusp when two electrons are in the same place ( $\mathbf{r}_1 = \mathbf{r}_2 \Rightarrow \frac{1}{r_{12}} \rightarrow \infty$ ) is not captured correctly. The difference between HF and the “truth” is called “electron correlation.” Correlation lowers the total energy of a system by about 1 eV/electron pair in a well-behaved molecule. To move beyond HF, need to introduce “many body” character into wavefunction.

## 2. Static Correlation

H<sub>2</sub> molecule molecular orbital diagram



Hartree-Fock ground-state description necessarily contains an equal mixture of covalent and ionic terms. At equilibrium this might be a reasonable description of the H<sub>2</sub> molecule.

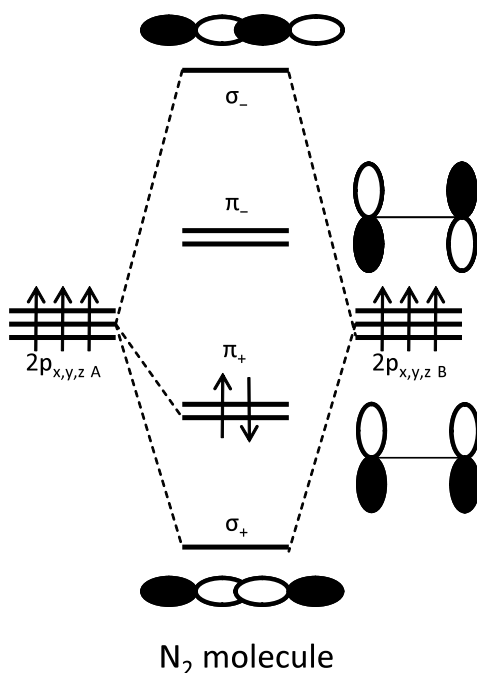
As the H – H distance is increased, though, this model becomes physically unreasonable—the electrons always have equal likelihood of being on either atom, when in reality we know at some point each electron localizes to one atom. With increasing separation, the H-F energy becomes much higher than the true (H<sub>2</sub> bond energy) energy.

This is a breakdown of the H-F MO model due to the limitation of a single determinant. UHF helps somewhat, but doesn't really fix the underlying problem. Example of a *static correlation* problem.

At large separation we want  $\Psi \rightarrow 1s_A(1)1s_B(2) + 1s_B(1)1s_A(2)$ . How to get that from  $\Psi_{\text{HF}}$ ? One way would be to construct a *two-configuration* (or *two-determinant*) wavefunction  $\Psi(1,2) = t\Psi_{\text{HF}}(1,2) + t'\Psi'_{\text{HF}}(1,2)$ . Here  $t$  are new (beyond H-F) variational parameters. Near equilibrium,  $t \rightarrow 1$  and  $t' \rightarrow 0$ , while at large separation,  $t = -t'$ , canceling out the ionic terms.

In this example, the forms of the MOs  $\psi_+$  and  $\psi_-$  are fixed by symmetry, but in general (say in a heteronuclear diatomic) that won't be the case. In *multiconfiguration self-consistent field* (MCSCF) calculations, a trial wavefunction is constructed from a few determinants added together, and both the molecular orbital coefficients and weights on the determinants are simultaneously optimized. Much trickier optimization than straight H-F.

Consider orbitals of  $\text{N}_2$  molecule. Try to choose configurations that are important for the problem at hand:



**Complete active space (CAS)** consider all excitations within some space of "active orbitals"

$$\Psi' = \dots t_n \left| \sigma_-(1)\bar{\sigma}_-(2)\pi_x^-(3)\bar{\pi}_x^-(4)\pi_y^-(5)\bar{\pi}_y^-(6) \right|$$

**MCSCF:** selectively include a few excited determinants

$$\begin{aligned} \Psi' = & t_1 \left| \sigma_-(1)\bar{\sigma}_-(2)\pi_x^-(3)\bar{\pi}_x^-(4)\pi_y^-(5)\bar{\pi}_y^-(6) \right| \\ & + t_2 \left| \sigma_+(1)\bar{\sigma}_-(2)\pi_x^-(3)\bar{\pi}_x^-(4)\pi_y^-(5)\bar{\pi}_y^-(6) \right| \\ & + t_3 \left| \sigma_+(1)\bar{\sigma}_-(2)\pi_x^-(3)\bar{\pi}_x^-(4)\pi_y^-(5)\bar{\pi}_y^-(6) \right| \end{aligned}$$

**Hartree-Fock:** single determinant

$$\Psi_{\text{HF}} = \left| \sigma_+(1)\bar{\sigma}_-(2)\pi_x^-(3)\bar{\pi}_x^-(4)\pi_y^-(5)\bar{\pi}_y^-(6) \right|$$

CASSCF on the p-based orbitals of  $\text{N}_2$  would involve 6 electrons distributed (in all ways consistent with spin and symmetry) in 12 orbitals. More general than MCSCF, but also much more expensive. Size of the calculation grows very rapidly with number of orbitals and electrons correlated, and there is no unambiguous way to choose configurations. Not for the faint of heart, and used only in very exacting cases today.

Illustrates general approach for dealing with correlation, of introducing "many-body" character into  $\Psi$  by adding together configurations.

### 3. Dynamic Correlation

MCSCF and CASSCF treat the single determinant deficiencies of Hartree-Fock, but they don't do a very good job of handling *dynamic correlation* that comes from instantaneous repulsion of electrons at all separations.

Example of higher excited states of  $\text{H}_2$ .  $1\sigma_g-2\sigma_g$  for in-out correlation,  $1\sigma_g-2\sigma_u$  for left-right correlation,  $1\sigma_g-2\pi_u$  for angular correlation.

## 4. Configuration Interaction

Most straight-forward approach to dynamic correlation is to combine a whole bunch of determinants variationally:

$$\Psi = \Psi_{\text{HF}} + \sum_{ia} t_i^a \Psi_i^a + \sum_{ijab} t_{ij}^{ab} \Psi_{ij}^{ab} + \sum_{ijkabc} t_{ijk}^{abc} \Psi_{ijk}^{abc} + \dots$$

$\Psi_{\text{HF}} = |\psi_1 \cdots \psi_n|$  reference determinant

$\Psi_i^a = |\psi_1 \cdots \psi_{i-1} \psi_a \psi_{i+1} \cdots \psi_n|$  singly excited determinant

(excite electron from occupied orbital  $\psi_i$  to unoccupied orbital  $\psi_a$ )

$\Psi_{ij}^{ab} = |\psi_1 \cdots \psi_{i-1} \psi_a \psi_{i+1} \cdots \psi_{j-1} \psi_b \psi_{j+1} \cdots \psi_n|$  doubly excited determinant

( $\psi_i \rightarrow \psi_a, \psi_j \rightarrow \psi_b$ )

(adopted from Schlegel)

If carried out to all possible excitations to all possible orbitals, called a *full configuration interaction* (full CI) model. This wavefunction would be exact within a given basis.

To find  $t$ , use variational principle and minimize  $E_{\text{FCI}} = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}$ . Yields a secular matrix that has to be diagonalized (adopted from Cramer).

	$\Psi_{\text{HF}}$	$\Psi_i^a$	$\Psi_{ij}^{ab}$	$\Psi_{ijk}^{abc}$
$\Psi_{\text{HF}}$	$E_{\text{HF}}$	0	dense	0
$\Psi_i^a$	0	dense	sparse	very sparse
$\Psi_{ij}^{ab}$	d e n s e	sparse	sparse	really sparse
$\Psi_{ijk}^{abc}$	0	very sparse	really sparse	...

Eigenvalues are the electronic energy states including correlation. Result is a *variational bound* on the ground-state energy.

*Condon-Slater* rules and *Brillouin's theorem* guarantee that many of the matrix elements are zero (any that differ by three or more orbitals, and any between H-F configuration and single excitations) and also make it pretty easy to evaluate any given matrix element. Further, matrix is block diagonalized by the electronic state symmetries. On the other hand, CI matrix can be of order millions  $\times$  millions, so special iterative diagonalization schemes that just find the lowest or a few eigenfunctions have to be used.

Full CI is way too expensive for anything but benchmarks. For example, a DZP basis for water would give 10 electrons in 50 spin-orbitals, and the number of possible singlet configurations in a full CI would be

$$\frac{25!}{5!20!} \frac{25!}{5!20!} = 2822796900$$

(half the electrons in spin-up orbitals and half in spin-down). Gives enormous matrix to diagonalize.

CI singles (CIS) – relatively inexpensive way to calculate energies of electronic excited states from an H-F reference configuration. By orthogonality of determinants, CIS does not change or improve the ground-state energy.

CI singles and doubles (CISD) – double excitations connect most directly with reference determinant and contain most of the correlation. Singles contribute indirectly and are inexpensive to include. Most general purpose method.

CI singles doubles and triples (CISDT) – super-expensive

Multi-reference CI (MRCI) – Replace H-F reference configuration with an MCSCF or CASSCF reference, and then CI on top of that.

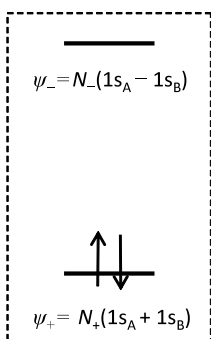
In all of these have to make a decision about which occupied and virtual orbitals to include in the active space. These methods definitely *not* for the casual user.

## 5. Size consistency

In calculating chemical things, would like *size consistency*, so that it is true that  $E(A+B) = E(A) + E(B)$  when A and B are too far apart to affect one another. Consider two  $H_2$  molecules like the one at the beginning of this lecture, with two molecular orbitals, and calculate CID energy either as that of a single molecule multiplied times two, or literally as two molecules with very large separations. We find that the energies are *not* equal.

$$\Psi_{\text{CID}}(H_2) = \Psi_{\text{HF}} + t\Psi_{++}^{--}$$

$$\Psi_{\text{HF}}(1,2) = |\psi_+(1)\bar{\psi}_+(2)|$$



$H_2$  molecule

$$E = 2E_{\text{CID}}(H_2)$$

$$\Psi_{\text{CID}}(2H_2) = (\Psi_{\text{HF}} + t\Psi_{++}^{--})^*$$

$$(\Psi_{\text{HF}} + t\Psi_{++}^{--})$$

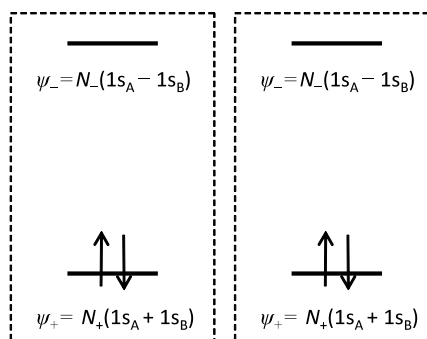
includes  
“quadruple”  
excitation

Two  $H_2$   
calculated  
separately

$$\Psi_{\text{CID}}(2H_2) = \Psi_{\text{HF}}^2 + t_A\Psi_{++}^{--} + t_B\Psi_{++}^{--}$$

excludes “quadruple” excitation

$$E \neq 2E_{\text{CID}}(H_2)$$



Two  $H_2$  calculated at  
very large separation

Any truncated CI is *not* size extensive. Can try to correct, but unwieldy.

## 6. Perturbation methods

Perturbation theory is approach whereby a Hamiltonian  $H$  that is difficult to calculate is separated into a part  $H_0$  (the zero-order part) that has known eigenvalues  $E_j^{(0)}$  and eigenfunctions  $\Psi_j^{(0)}$ , and a correction part  $H_1$ . Write  $\hat{H} = \hat{H}_0 + \lambda \hat{H}_1$  and Taylor-expand true eigenfunctions and eigenvalues in powers of the perturbation parameter  $\lambda$ , which goes from 0 for the reference system to 1 for the fully-interacting system. Collecting terms in  $\lambda$  gives a sequence of higher-order and (hopefully) more accurate estimates of  $\Psi$  and  $E$ . For instance, first-order correction is  $E_1 = \langle \Psi_0^{(0)} | \hat{H}_1 | \Psi_0^{(0)} \rangle$  and second-order one is

$$E_2 = \langle \Psi_0^{(0)} | \hat{H}_1 | \Psi_0^{(1)} \rangle = \frac{\sum_{j>0} \left| \langle \Psi_j^{(0)} | \hat{H}_1 | \Psi_0^{(0)} \rangle \right|^2}{E_j^{(0)} - E_0^{(0)}}.$$

Applied to dynamical correlation problem in molecular systems containing many electrons, is called *Møller-Plesset perturbation theory* (MPn) or *many-body perturbation theory* (MBPTn). The sum of all the one-electron Fock operators is taken as  $H_0$ , which includes the mean-field repulsion of every electron with every other and the difference between this and the true electron-electron interaction is the “perturbation” (although not a small one!). The zero-order wavefunctions are H-F plus all possible excitations.

$$\begin{aligned} \hat{H}_0 &= \sum_i^{N/2} \hat{f}_i & \Psi_0^{(0)} &= |\psi_1(1) \dots \psi_n(n)| = \Psi_{\text{HF}} & E_0^{(0)} &= 2 \sum_i^{N/2} \epsilon_i = E_{\text{HF}} + 2 \sum_{i=1}^{N/2} \sum_{j=1}^{N/2} (2J_{ij} - K_{ij}) \\ \hat{H}_1 &= \sum_i \sum_{j>i} \left( \frac{1}{r_{ij}} - \left( 2 \sum_{i=1}^{N/2} \sum_{j=1}^{N/2} (2J_{ij} - K_{ij}) \right) \right) \end{aligned}$$

In this model, the first-order energy (“MP1”) is identical to Hartree-Fock.

Second-order energy (“MP2”) given by

$$E_{\text{MP2}} = E_{\text{HF}} + E_2 = E_{\text{HF}} + \int \Psi_{\text{HF}} \hat{H}_1 \Psi_1 d\tau = E_{\text{HF}} - \sum_{i>j} \sum_{a>b} \frac{\left| \langle \Psi_{\text{HF}} | \hat{H}_1 | \Psi_{ij}^{ab} \rangle \right|^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}$$

MP2 includes only double excitations, and the necessary integrals are just the two-electron integrals which can be evaluated relatively inexpensively (scales as fifth power of system size, vs. fourth power for H-F). Implemented for both closed (restricted MP2) and open-shell (unrestricted MP2) systems, including gradients of the energy. 2nd order recovers a large fraction of the correlation energy when Hartree-Fock is a good starting point. Gives very reliable ground-state structures and improves energies relative to H-F.

Approach can be extended to higher orders, including MP3, MP4, .... all in closed form, non-iterative. Sometimes selectively include only some excitations (MP4(SDQ)). Because the perturbation series include sums over all possible excitations, they are all *size-consistent*, but the trade-off is that they are *not variational*. MP2 energy could be below true energy. Computational expensive increases rapidly with perturbation order.

Practical up to fourth order (includes single, doubly, triple and quadruple excitations). MP3 at best a marginal improvement over MP2, MP4 recovers most of the rest of the correlation energy, but series tends to oscillate (even orders lower) and does not necessarily converge, especially if there is serious spin contamination or if Hartree-Fock is not a good starting point.

Very user-friendly: choose a basis, choose a perturbation order, possibly exclude some orbitals, away you go!

For difficult correlation cases, can be combined with a multi-determinant reference (CASPT2), now more difficult to use.

## 7. Coupled-cluster methods

Probably most general purpose reliable method available today. Define excitation operators:

$$\hat{T}_1 = \sum_i \sum_a t_i^a \Psi_i^a \quad \hat{T}_2 = \sum_{i>j} \sum_{a>b} t_{ij}^{ab} \Psi_{ij}^{ab} \quad \hat{T}_3 = \dots \quad \hat{T} = \sum_n \hat{T}_n$$

$$\Psi_{\text{CISD}} = (1 + \hat{T}_1 + \hat{T}_2) \Psi_{\text{HF}} \quad \Psi_{\text{FCI}} = e^{\hat{T}} \Psi_{\text{HF}} = (1 + \hat{T} + \frac{\hat{T}^2}{2} + \dots) \Psi_{\text{HF}}$$

This is a compact way to write things, and computational problem is to find the “amplitudes”  $t$ .

Becomes useful when the excitation operators are truncated.

*Coupled-cluster doubles* (CCD) would have

$$\hat{T} = \hat{T}_2 \quad \Psi_{\text{CCD}} = e^{\hat{T}_2} \Psi_{\text{HF}} = (1 + \hat{T}_2 + \frac{\hat{T}_2^2}{2} + \dots) \Psi_{\text{HF}}$$

This includes all double excitations, plus some quadruples (through squared  $T_2$  term), and hexuples, and .... By construction this is size-consistent, and CCD goes beyond MP2 in including the contributions of double-excitations to infinite order. Have to find the amplitudes  $t$ , which are determined by a set of non-linear equations that have to be solved iteratively. More expensive than MP2.

CC equations quadratic in the amplitudes and must be solve iteratively, cost of each iteration similar to MP4SDQ.

Coupled-cluster singles and doubles (CCSD) – single and double excitations correct to infinite order, quadruple excitations correct to 4th order.

CCSD(T) – add triple excitations by perturbation theory. This is the highest-quality correlation method applied routinely. Works extremely well in capturing the chemically significant components of correlation for ground-state, equilibrium molecules. Biggest problem is with static correlation (multireference character) in the reference configuration. Diagnostics designed to detect this problem. Also can be extended to treat excited states.

QCISD – configuration interaction with enough quadratic terms added to make it size extensive. Developed by Pople, later shown to be equivalent to a truncated form of CCSD. Can also add triples by perturbation theory, QCISD(T). CC and QCI are very similar in performance, choice is largely a matter of what is implemented in a particular code.

Heirarchy of “quality”:

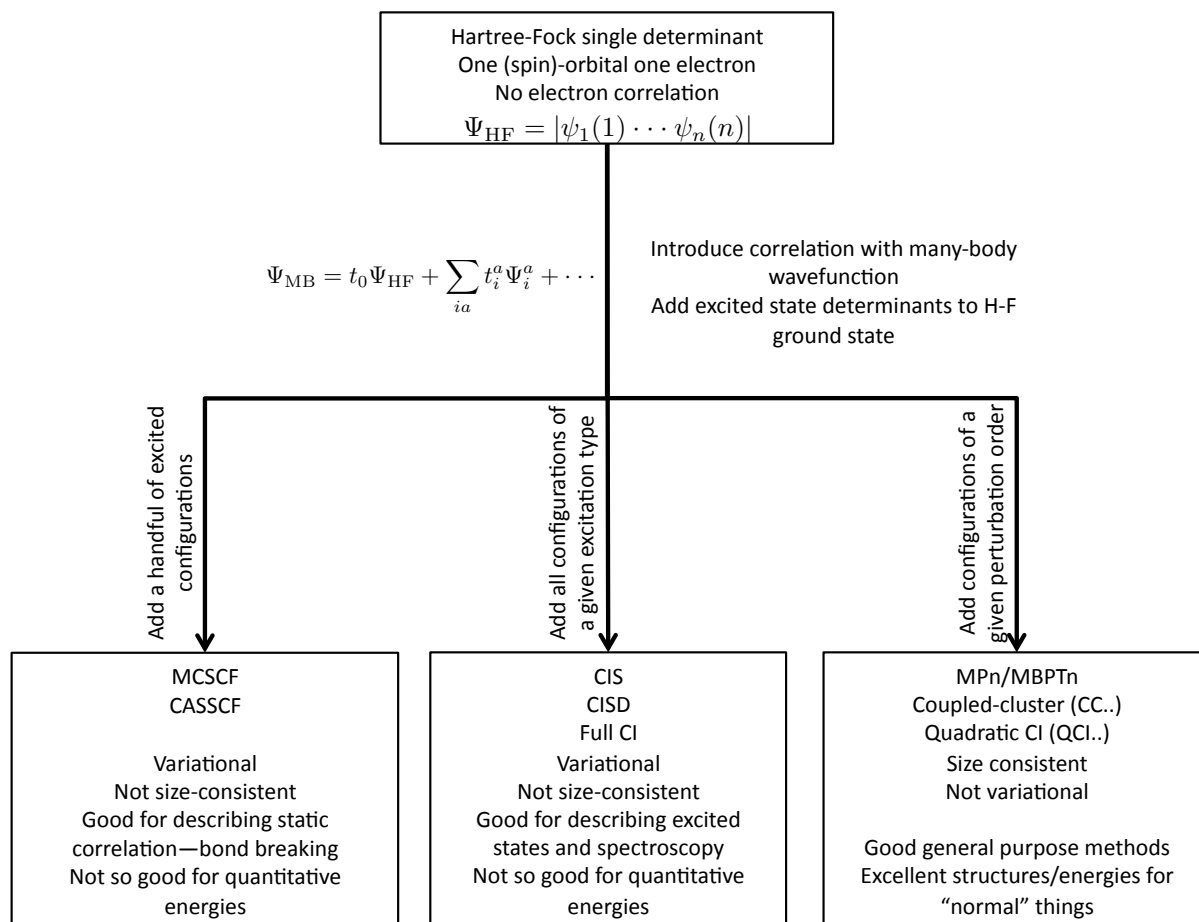
$$\text{HF} < \text{MP2} \sim \text{MP3} \sim \text{CCD} < \text{CISD} < \text{QCISD} \sim \text{CCSD} < \text{MP4} < \text{QCISD(T)} \sim \text{CCSD(T)}$$

Not surprising, also the hierarchy of computational expense! This expense includes not only CPU cycles, but also computer memory and disk space. As with SCF, various approaches can trade off these

requirements. A significant limitation is that many of these methods are not parallelized and do not take good advantage of the “Linux cluster” type of computer architecture that is so common today.

## 8. *Ab initio* solution of the Schrödinger equation

### Electron Correlation Summary



We have now laid out an array of approaches of varying accuracy for solving the molecular Schrödinger equation in the Born-Oppenheimer approximation. If we limit ourselves to single-reference correlation methods, to completely define any particular approach, we just have to specify the atomic basis set and the electron correlation method. As these methods are all analytically defined, this is enough to exactly specify a model. For example, CCSD(T)/6-311+G(2d,p) exactly defines one model, with a well-defined energy and wavefunction for any given configuration of atoms comprising a molecule. These are frequently termed *ab initio*, or “from the beginning,” methods to distinguish from semi-empirical and DFT.

We thus have two axes to march along, and the *exact* solution of the Schrödinger equation can be gotten but taken both to their limits, i.e. FCI/complete basis. Such a calculation is of course impossible for any real system.

		Correlation →					
		HF	MP2	CCSD	CCSD(T)	...	FCI
Basis ↓	SZ						FCI/SZ
	DZ						
	DZP						
	TZ2P				CCSD(T)/TZ2P		FCI/TZ2P
	...						
	Complete basis	HF limit	MP2 limit				exact

The complete basis set correlation energy can be approximated by extrapolation. The two axes are coupled through the fact that systematically more correlation is captured as the radial and angular degrees of freedom of the basis set are increased. Using a series of basis sets constructed to systematically increase both degrees of freedom, like the “correlation consistent” ones, it is possible to get the infinite basis correlation energy by extrapolation. Following from Thom Dunning:

Feature Article

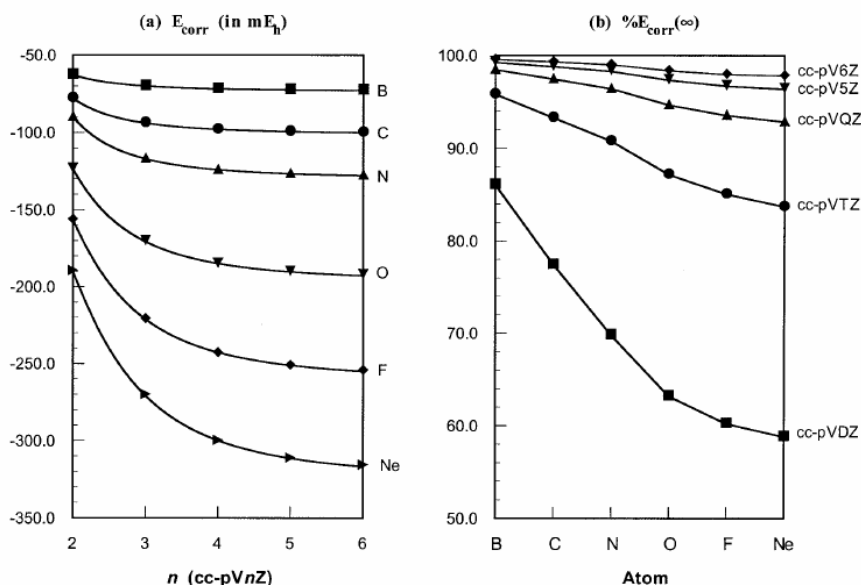
*J. Phys. Chem. A, Vol. 104, No. 40, 2000 9067*

Figure 3. Calculation of the valence-electron correlation energy of the first-row atoms, B–Ne, using the CCSD(T) or RCCSD(T) method and the correlation-consistent basis sets. (a) Total valence-electron correlation energy (in millihartrees). (b) Fraction of the correlation energy recovered for each atom and basis set, i.e.,  $E_{\text{corr}}(n)/E_{\text{corr}}(\infty)$  (in percent).

$$E_{\text{corr}}(n) = E_a + E_b/n^3 + E_c/n^5 \quad (12)$$

Correlation energy fits to

Ability to calculate energy and PES accurately depend a lot on the system. Normal closed-shell molecules near their equilibrium geometry are described more readily than molecules with “exotic” electronic structure (diradicals) or at regions far from equilibrium. No method or algorithm is universally right and tractable. Cramer has a nice table, taken from R. J. Bartlett, in *Modern Electronic Structure*



*Theory*, Part 2, Chap. 6, D. R. Yarkony, ed., World Scientific, New York (1995), which I shamelessly reproduce in part here.

Table. Average errors in correlation energies ( $\text{kcal mol}^{-1}$ ) compared to full CI for various methods applied to HB,  $\text{H}_2\text{O}$ , and HF at two geometries with a DZP basis.

<i>Level of theory</i>	<i>Equilibrium geometry</i>	<i>Equilibrium and stretched geometry</i>
MP2	10.4	17.4
MP3	5.0	14.4
CISD	5.8	13.8
CCSD	1.9	4.5
QCISD	1.7	4.0
MP4	1.3	3.7
CCSD(T)	0.3	0.6
QCISD(T)	0.3	0.5
CCSDT	0.2	0.5
CCSDTQ	0.01	0.02


## 9. Model chemistries

Fortunately, for most of chemistry we do not really need to know the *exact* energy or wavefunction of a particular system. Many properties of interest, like molecular structures or vibrational spectra or charge distributions, are well-described by relatively low-level treatments, like Hartree-Fock or MP2 (or, as we'll see shortly, DFT). Reaction or activation energies are actually energy *differences*, and error cancellation generally means these differences converge much more rapidly than do the absolute molecular energy.

**NOTE:** Because of the errors associated with basis set incompleteness and incomplete treatment of correlation, it is only meaningful to compare energies between molecules calculated using exactly the same method.

From Schlegel and Gaussian Inc.: (note fortuitous error cancellation at some levels)

Model Chemistry Calculations Using Gaussian 98



# RMS errors in calculated equilibrium bond length, Re, ( Å ), of diatomics.

Basis Set	Level of Theory					
	HF	B3LYP	MP2	MP3	QCISD	QCISD(T)
3-21G	0.0187 <sup>a</sup>	0.0442	0.0664	0.0420	0.0528	0.0581
6-31G(d)	0.0260 <sup>a</sup>	0.0108	0.0190	0.0082	0.0148	0.0192
6-311G(d,p)	0.0327	0.0049	0.0091	0.0075	0.0052	0.0069
6-311+G(2df,2pd)	0.0344	0.0062	0.0088	0.0112	0.0055	0.0037

<sup>a</sup> Accurate bond angles require polarization functions.

9/26/01

Copyright © 1990-2001, Gaussian, Inc.

The Pople community talks about *theoretical model chemistries*: a completely determined, size-consistent algorithm for the calculation of the energy of any molecular system. A *simple model chemistry* employs a single theoretical method and basis set and corresponds to one spot on the matrix above; a *compound model chemistry* combines several theoretical methods and basis sets to achieve higher accuracy at lower computational cost. Advantage of model chemistry concept is that it allows systematic comparison and validation against experiment.

Simple model chemistries often represented as Method1/Basis1//Method2/Basis2, where Method2/Basis2 describes the model for calculating the geometry and Method1/Basis1 for subsequently calculating the energy.

Compound model chemistries built around idea that corrections for increasing basis set size and increasing correlation treatment are additive. For instance, in my old COF<sub>2</sub> work, used

$$E[\text{QCISD(T)/aug-cc-pVTZ}] \approx E[\text{QCISD(T)/cc-pVTZ}] + E[\text{MP2/aug-cc-pVTZ}] - E[\text{MP2/cc-pVTZ}]$$

Assumes affects of augmented/diffuse functions at QCISD(T) level can be approximated by contribution at MP2 level.

Most venerable of these additive models is the “Gaussian 2” (or G2) model (Curtis, Raghavachari, Trucks, and Pople, *J. Chem. Phys.* **1991**, 94, 7221, intended to reproduce QCISD(T)/6-311+G(3df,p) energies. Designed around G2 test-set, a collection of well-established thermochemical data for 147 atoms, molecules, and ions, that provide well-established values for 125 atomization, ionization, electron affinity, and proton affinity energies. Includes an empirical “higher level correction” to account for e.g. core-valence correlation, and anything else that is missing.

G2 algorithm:

	<i>Hartree-Fock</i>	<i>MP2</i>	<i>MP4</i>	<i>QCISD(T)</i>
6-31G(d)	<i>ZPE</i>	Structure		
6-311G(d)		1	2	3
G-311+G(d,p)		4	5	
6-311G(2df,p)		6	7	
6-311+G(3df,p)		8		<i>E</i>

$$\text{QCISD(T)/6-311+G(3df,p)} \approx 2 + (3 - 2) + (5 - 2) + (7 - 2) + (8 - 1) - (4 - 1) - (6 - 1)$$

$$\Delta E(\text{HLC}) = -2.50 \text{ mHa} * (\# \text{ electron pairs}) - 0.19 \text{ mHa} * (\# \text{ electron pairs})$$

$$E(\text{G2}) = \text{QCISD(T)/6-311+G(3df,p)} + 0.8929 * \text{ZPE} + \Delta E(\text{HLC})$$

Performance often measured in “MAD,” or mean-absolute deviation. G2 appears to do pretty well, as shown in table below from Curtiss, Raghavachari, Redfern, Rassolov, and Pople *J. Chem. Phys.* **1998**, 109, 7764.

TABLE VI. Comparison of average absolute deviations and root-mean-square deviations for G2 and G3 theory.<sup>a</sup>

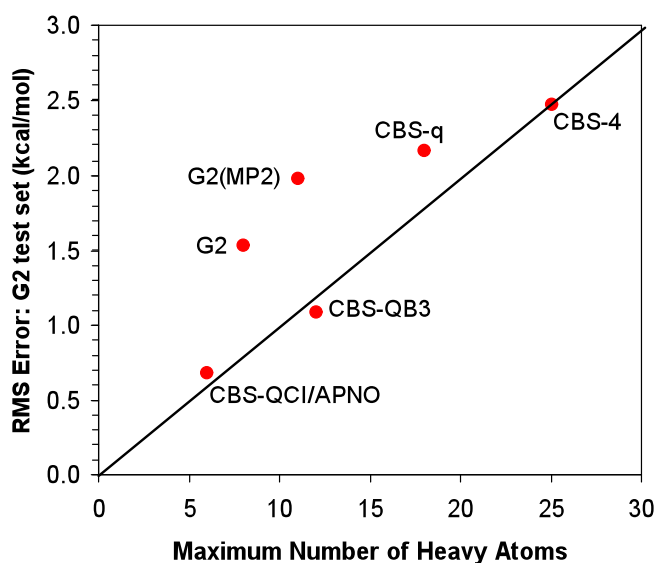
Type	Average absolute deviation, kcal/mol		Root-mean-square deviation, kcal/mol	
	G2	G3	G2	G3
Enthalpies of formation (148)	1.56	0.94	2.10	1.35
Nonhydrogen (35)	2.44	1.72	3.14	2.19
Hydrocarbons (22)	1.29	0.68	1.66	1.00
Subst. hydrocarbons (47)	1.48	0.56	1.81	0.85
Inorganic hydrides (15)	0.95	0.87	1.32	1.09
Radical (29)	1.16	0.84	2.25	0.98
Ionization energies (85)	1.41	1.13	1.84	1.65
Electron affinities (58)	1.41	1.00	1.68	1.38
Proton affinities (8)	1.08	1.34	1.22	1.44
All (299)	1.48	1.02	1.93	1.45

<sup>a</sup>HLC parameters for G3 theory:  $A = 6.386$  mhartrees,  $B = 2.977$  mhartrees,  $C = 6.219$  mhartrees,  $D = 1.185$  mhartrees. The average absolute deviations for G2 theory are slightly different from those reported in Refs. 5 and 6 due to the use of a new value for the enthalpy of formation of  $\text{COF}_2$  (see Table II) and ionization potential of  $\text{C}_2\text{H}_5\text{OH}$  (see Table III).

Remember though that this is within a fairly narrow space of chemical species, so no guarantee that it will perform as well for your compound. Also, MAD doesn't necessarily capture effects of outliers.

Many, many variants. For example, G2(MP2) replaces MP4 calculations with MP2, increases MAD to  $1.8 \text{ kcal mol}^{-1}$ .

Other model chemistries have been developed. G3 is faster and more accurate, plus it adds support for second row elements. CBS (or "complete basis set") methods by Petersson use basis extrapolations to improve accuracy.  $W_n$  (or "Weizmann- $n$ ") methods really juiced for accuracy.



Copied from Schlegel

Compound Model Chemistry Calculations Using Gaussian 98

Worst Case Errors							
Errors in calculated atomization energies (kcal/mol) for each of the six models							
Exp.-Calc.							
Molecule	CBS-4	CBS-q	CBS-QB3	G2(MP2)	G2	CBS-QCI APNO	Exp.
O <sub>3</sub>	11.9	-2.2	0.4	0.0	-1.1	0.8	142.2±0.5
CH <sub>3</sub> CH <sub>2</sub> CH <sub>3</sub>	-2.9	-0.9	1.0	0.8	-0.2	-2.1	942.9±0.2
CF <sub>4</sub>	-0.6	4.0	-2.1	-8.2	-5.6	-2.9	465.5±0.2
SiF <sub>4</sub>	5.1	10.6	-0.7	6.6	8.4		566.6±0.3
CCl <sub>4</sub>	-11.5	-1.3	-6.0	-6.5	-2.8		307.3±0.7
Naphthalene	-18.3	3.6	1.6	7.2	5.9		2094±1
Max. Error	7.4	4.9	5.5	4.1	5.5	4.3	
RMS Error							

9/26/01 Copyright © 1990-2001, Gaussian, Inc. 27

Others developed to reproduce other quantities, like activation energies. Truhlar has a small business in generating new models fitted to various data and intended to reproduce various properties. Philosophical debate between this approach and the purist all *ab initio* methods.

	HF/3-21G		G2	
	F-C-C-F angle (deg)	<i>E</i> <sub>rel</sub> (kJ/mol)	F-C-C-F angle (deg)	<i>E</i> <sub>rel</sub> (kJ/mol)
<i>trans</i>	180.0	0	180.0	0
<i>gauche</i>	74.97	3.5	68.94	3.3
H-F eclipsed drag	120	11.97		
H-F eclipsed TS	121.3	11.99		
F-F eclipsed	0	38	0	26.7

## 10. References

Electron correlation: Raghavachari and Anderson, *J. Phys. Chem.* **1996**, *100*, 12960-12973.

Model Chemistries: Head-Gordon, *J. Phys. Chem.* **1996**, *100*, 13213-13225.

Basis extrapolation: T. Dunning, *J. Phys. Chem. A* **2000**, *104*, 9062-9080.

Cramer, Chapter 7.

Foresman and Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian, Inc.