MODERN ELECTRONIC STRUCTURE THEORY:
Electron Correlation

In the previous lecture, we covered all the ingredients necessary to choose a good atomic orbital basis set. In the present lecture, we will discuss the other half of accurate electronic structure calculations: how we compute the energy. For simple MO theory, we used the non-interacting (NI) electron model for the energy:

$$E_{NI} = \sum_{\mu=1}^{N} E_\mu = \sum_{\mu=1}^{N} \int \psi_\mu(1) \hat{H}_0 \psi_\mu(1) d\tau$$

Where, on the right hand side we have noted that we can write the NI energy as a sum of integrals involving the orbitals. We already know from looking at atoms that this isn't going to be good enough to get us really accurate answers; the electron-electron interaction is just too important. In real calculations, one must choose a method for computing the energy from among several choices, and the accuracy of each method basically boils down to how accurately it treats electron correlation.

Self Consistent Fields

The Hartree Fock (HF) Approximation

The Hartree-Fock method uses the IPM energy expression we've already encountered:

$$E_{IPM} = \sum_{\mu=1}^{N} E_\mu + \sum_{\mu<\nu}^{N} \tilde{J}_{\mu\nu} - \tilde{K}_{\mu\nu}$$

$$E_\mu = \int \psi_\mu(1) \hat{H}_0 \psi_\mu(1) d\tau$$

$$\tilde{J}_{\mu\nu} \equiv \iint \psi_\mu^*(1) \psi_\nu^*(2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \psi_\mu(1) \psi_\nu(2) \, d\mathbf{r}_1 \, d\mathbf{r}_2 \, d\sigma_1 \, d\sigma_2$$

$$\tilde{K}_{\mu\nu} \equiv \iint \psi_\mu^*(1) \psi_\nu^*(2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \psi_\mu(2) \psi_\nu(1) \, d\mathbf{r}_1 \, d\mathbf{r}_2 \, d\sigma_1 \, d\sigma_2$$

Since the energy contains the average repulsion, we expect our results will be more accurate. However, there is an ambiguity in this expression. The IPM energy above is correct for a determinant constructed out of any set of orbitals $\{\psi_\mu\}$ and the energy will be different depending on the orbitals we choose. For example, we could have chosen a different set of orbitals, $\{\psi'_\mu\}$, and gotten a different energy:

$$E'_{NI} = \sum_{\mu=1}^{N} E'_i + \sum_{\mu\nu}^{N} \tilde{J}'_{\mu\nu} - \tilde{K}'_{\mu\nu}$$

How do we choose the best set of orbitals then? Hartree-Fock uses the variational principle to determine the optimal orbitals. That is, in HF **we find the set of orbitals that minimize the independent particle energy**. These orbitals will be different from the non-interacting orbitals because they will take into account the average electron-electron repulsion terms in the Hamiltonian. Thus, effects like shielding that we have discussed qualitatively will be incorporated into the shapes of the orbitals. This will tend to lead to slightly more spread out orbitals and will also occasionally change the ordering of different orbitals (e.g. σ might shift below π once interactions are included).

Now, the molecular orbitals (and hence the energy) are determined by their coefficients. Finding the best orbitals is thus equivalent to finding the best coefficients. Mathematically, then, we want to find the coefficients that make the derivative of the IPM energy zero:

$$\frac{\partial E_{IPM}}{\partial \mathbf{c}^{\mu}} = \frac{\partial}{\partial \mathbf{c}^{\mu}} \sum_{\mu=1}^{N} E_{\mu} + \sum_{\mu\nu}^{N} \tilde{J}_{\mu\nu} - \tilde{K}_{\mu\nu} = 0$$

After some algebra, this condition can be re-written to look like an eigenvalue problem:

$$\mathbf{F}[\mathbf{c}]\mathbf{c}^{\mu} = E_{\mu}\,\mathbf{S}\,\mathbf{c}^{\mu} \qquad\qquad \textbf{Eq. 1}$$
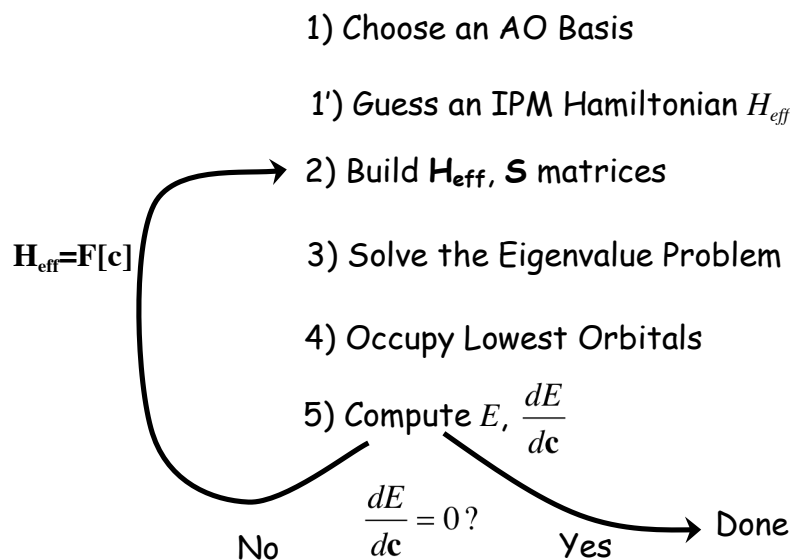
The matrix $\mathbf{F}$ is called the Fock matrix and it depends on the MO coefficients:

$$F_{ij} \equiv \int \phi_i^{AO} \hat{H}_0 \phi_j^{AO}\, d\tau$$

$$+ \sum_{\mu=1}^{N} \int \phi_i^{AO}(1)\psi_{\mu}(2)\frac{1}{r_{12}}\phi_j^{AO}(1)\psi_{\mu}(2)d\tau_1 d\tau_2 - \int \phi_i^{AO}(1)\psi_{\mu}(2)\frac{1}{r_{12}}\phi_j^{AO}(2)\psi_{\mu}(1)d\tau_1 d\tau_2$$

The first term is just the independent particle Hamiltonian that we've been dealing with in cruder forms of MO theory. The second and third terms are new contributions that arise from the average electron-electron repulsion present in our system. These terms reflect the fact that, on average, **each electron will see not only the nuclear attraction (first term) but also the average Coulomb (second) and Exchange (third) interactions with all the remaining electrons**. For this reason, HF is also known as "mean field" theory, because it includes both the interaction of each electron with the

nuclei (through $\int \phi_i^{AO} \hat{H}_0 \phi_j^{AO} \, d\tau$) and the average, or mean, repulsion of the μth electron with all the other electrons (via the Coulomb and exchange terms). Just as was the case for crude MO theory, the eigenvalue, E_μ, of the μth orbital can be interpreted as an **ionization potential** if the orbital is occupied, or as an **electron affinity** if the orbital is not occupied. The latter result is known as Koopmans' theorem and it is a rather remarkable feature of HF, given that it includes the average repulsion between electrons and the IP and EA refer to systems with *different numbers* of electrons.

Note the dependence of the Coulomb and exchange terms on the molecular orbitals ψ_μ. If **F** did not depend on the MO coefficients, we could solve Eq.2 by just finding the eigenvalues and eigenvectors of **F**. However, this would be inconsistent because we would be using one set of coefficients (**c**) to define **F** and we would get a *different set of coefficients* (**c'**) as the eigenvectors of **F**. We need MO coefficients that are **self-consistent** with the mean field they generate. Thus, in order to satisfy the Hartree-Fock equations one typically resorts to an iterative procedure, where steps 2-5 of our MO procedure are performed repeatedly:

1) Choose an AO Basis

1') Guess an IPM Hamiltonian H_{eff}

2) Build **H_{eff}**, **S** matrices

$\mathbf{H_{eff}=F[c]}$

3) Solve the Eigenvalue Problem

4) Occupy Lowest Orbitals

5) Compute E, $\dfrac{dE}{d\mathbf{c}}$

$\dfrac{dE}{d\mathbf{c}} = 0$?   No   Yes → Done

Here, HF makes use of the fact that defining an IPM Hamiltonian, H_{eff}, completely determines the molecular orbital coefficients, **c**. Thus, the most convenient way to change the orbitals is actually to change the Hamiltonian that generates the orbitals, and the form of the Hartree Fock Equations (Eq. 2) suggest a natural way to improve the effective Hamiltonian. The calculation converges when we find the molecular orbitals that give us the

lowest possible energy, because then $\frac{dE}{d\mathbf{c}} = 0$. At this point, the Hamiltonian will be self-consistent: the orbital coefficients we used to build $\mathbf{F}$ and define $\mathbf{H}_{eff}$ will be the same as the ones we get from the eigenvectors of $\mathbf{H}_{eff}$. For this reason, these iterations are called self-consistent field (SCF) iterations. Physically, these iterations locate a set of MOs that are eigenstates of the averaged electrostatic repulsion that they, themselves, generate.

One can go on to define atomic charges, bond orders and other qualitative metrics within the HF approximation. Before talking about the practical accuracy of HF, though, we first discuss a closely related method.

Density Functional Theory (DFT)
Here, we still use a Slater determinant to describe the electrons. Hence, the things we want to optimize are still the MO coefficients $\mathbf{c}^{\mu}$. However, we use a different prescription for the energy – one that is entirely based on the electron density. For a single determinant, the electron density, $\rho(\mathbf{r})$ is just the probability of finding an electron at the point $\mathbf{r}$. In terms of the occupied orbitals, the electron density for a Slater Determinant is:

$$\rho(\mathbf{r}) = \sum_{\mu=1}^{N} |\psi_{\mu}(\mathbf{r})|^2 \qquad \text{Eq. 2}$$

This has a nice interpretation: $|\psi_{\mu}(\mathbf{r})|^2$ is the probability of finding an electron in orbital μ at a point $\mathbf{r}$. So the formula above tells us that for a determinant the probability of finding an electron at a point $\mathbf{r}$ is just the sum of the probabilities of finding it in one of the orbitals at that point.

There is a deep theorem (the Hohenberg-Kohn Theorem) that states:

> There exists a functional $E_v[\rho]$ such that, given the ground state density, ρ_0, $E_v[\rho_0]=E_0$ where E_0 is the **exact ground state energy**. Further, for any density, ρ', that is not the ground state density, $E_v[\rho']>E_0$.

This result is rather remarkable. While solving the Schrödinger Equation required a very complicated 3N dimensional wave function $\Psi_{el}(\mathbf{r}_1, \mathbf{r}_2, ... \mathbf{r}_N)$, this

theorem tells us we only need to know the density - which is a 3D function! – and we can get the exact ground state energy. Further, if we don't know the density, the second part of this theorem gives us a simple way to find it: just look for the density that minimizes the functional E_v.

The unfortunate point is that we don't know the form of the functional E_v. We can prove it exists, but we can't construct it. However, from a pragmatic point of view, we do have very good approximations to E_v, and the basic idea is to choose an approximate (but perhaps very, very good) form for E_v and then minimize the energy as a function of the density. That is, we look for the point where $\frac{dE_v}{d\rho} = 0$. Based on Eq. 2 above, we see that ρ just depends on the MOs and hence on the MO coefficients, so once again we are looking for the set of MO coefficients such that $\frac{dE_v}{d\mathbf{c}} = 0$. Given the similarity between DFT and HF, it is not surprising that DFT is also solved by self consistent field iterations. In fact, in a standard electronic structure code, DFT and HF are performed in exactly the same manner (see flow chart above). The only change is the way one computes the energy and $\frac{dE}{d\mathbf{c}}$. Note that the condition $\frac{dE_v}{d\mathbf{c}} = 0$ can also be cast as a generalized eigenvalue-like equation:

$$\mathbf{H_{KS}}\left[\mathbf{c}\right]\mathbf{c}^\mu = E_\mu\,\mathbf{S}\,\mathbf{c}^\mu$$

The effective Hamiltonian appearing in this equation is usually called the Kohn-Sham Hamiltonian.

Now, as alluded to above, there exist good approximation**s** (note the plural) to E_v. Just as was the case with approximate AO basis sets, these approximate energy expressions have strange abbreviations. We won't go into the fine differences between different DFT energy expressions here. I'll simply note that roughly, the quality of the different functionals is expected to follow:

LSDA ‹ PBE $\approx$ BLYP ‹ PBE0 $\approx$ B3LYP

Thus, LSDA is typically the worst DFT approximation and PBE0 and B3LYP are typically among the best. I should mention that this is just a rule of thumb; unlike the case for basis sets where we were approaching a well-defined limit, here we are trying various uncontrolled approximations to an

unknown functional. Experience shows us that B3LYP is usually the best, but this need not always be the case.

How big a molecule can I treat with SCF theory?

Now we come to the practical aspect of SCF theory: What is it good for? Given the speed of modern computers, can I do an SCF calculation on (fill in your favorite molecule)? How accurate will my answers be? Because SCF has been around for 75 years, all of these empirical questions have been answered by computational experiments. Here, we will merely summarize some of the most important facts

The question of what molecules are "too big" for HF is, in some sense, a poorly defined question. Clearly, bigger molecules will mean longer calculations, but if I am very patient or have a very fast computer, what is "too big" for you might be just right for me. Further, since computers are getting exponentially faster due to Moore's law, any limits we place on HF now will seem terribly outdated in 10 years.

Thus, rather than state a rule for the maximum system size we can treat in SCF calculations, it is better to learn how to calculate a ballpark figure for how long a calculation will take. From that we can back out the current limits on molecule size, but we can also extrapolate what size molecules will be available in the future.

Toward this end, we note that in order to do a HF calculation, we must compute the eigenvalues of the Fock matrix. As a first step, we will make the gross approximation that this is the **only** step of the calculation that requires significant computational effort. In practice, there are other steps that take time (computing integrals, for example) but to get a rough estimate, we can neglect these parts. The Fock matrix has dimension (# of AOs)x(# of AOs). Assume every atom is a first row atom (BCNOF) and that we are working in a DZP basis set. Then each atom will need a basis of size (3s2p1d), for a total of 14 orbitals per atom. For simplicity, we will round this down to 10, since multiplying by 14 is inconvenient. Thus, the Fock Matrix will have dimension ~(10A)x(10A). This is somewhat of an underestimate, but it will suffice for the time being.

Now, there are two ways the calculation can become too big for the computer. First, it could run out of memory (RAM). Second, we could run

out of time. In the first case, the primary challenge is finding enough memory for the $(10A)x(10A)=100A^2$ numbers that comprise the Fock matrix. As a rule of thumb, each number takes up 8 bytes so that the Fock matrix requires $800A^2$ bytes of storage. A typical computer these days has on the order of 2 Gigabytes ($2x10^9$ bytes) of storage, which will get maxed out if $800A^2=2x10^9$, which happens when A=2500 or so. So, based on storage, we conclude HF calculations are possible on systems up to a couple thousand atoms in size. Looking at the time the calculation takes, the key step is finding the eigenvectors of the Fock matrix. As a rule of thumb, finding the eigenvectors of an N-by-N matrix takes about $5N^3$ operations (add, subtract, multiply, divide). Thus, finding the eigenvectors of $\mathbf{F}$ will require about $5x(10A)^3=5000A^3$ operations. A modern CPU can do about 3 billion operations in a second (the clock speed on processors is on the order of 3 GHz). Thus, in an hour a good CPU can do (3600 s/hr)x($3x10^9$ ops/s)$\approx1x10^{13}$ operations. Thus, the calculation will take an hour when $5000A^3\approx1x10^{13}$, which again occurs when A$\approx$1200. This estimate is a little bit less accurate than the storage estimate because we've neglected all the other steps in the SCF procedure, and we've also ignored the fact that in the SCF iterations will require us to find the eigenvectors of $\mathbf{F}$ many, many times. Thus, a 1200 atom calculation will actually take quite a bit more than an hour, but that's OK because in practice even waiting 1-2 days for an answer isn't so bad, because the computer can work while we're off doing something else.

In any case, whether we look at time or storage, we arrive at the same conclusion: SCF calculations can be run on molecules with up to a couple thousand atoms. This opens up a huge variety of chemical systems to *ab initio* computation and has led to MO theory calculations becoming par for the course in many research publications. The limiting factor in expanding the size of systems one can study with HF or DFT is that the cost grows non-linearly with the size of the system. If your present computer can handle A atoms, then it has something like A^2 memory and A^3 CPU speed. In order to handle 2A atoms, you'll need $(2A)^2=4A^2$ memory and $(2A)^3=8A^3$ CPU. Thus, the computational demands grow faster than the system size – if you double the system size, you need four times the memory and 8 times the CPU power. This makes the calculation increasingly difficult for large systems and one inevitably hits a wall beyond which the calculation is just too difficult. To circumvent this, many groups are investing time in developing *linear scaling* algorithms for MO simulations that hinge on two

points: First, for very large systems, many of the terms in the RH equations are very, very small (e.g. the interaction between two very distant atoms). Avoiding computing these irrelevant terms can save an immense amount of time and storage. Second, by reformulating the SCF iterations, one can avoid computationally intensive steps like finding eigenvectors in favor of many less expensive operations (like matrix-vector multiplication). At the moment, these algorithms are still work in progress, but they will be required if one wants to be able to simulate truly large molecules and aggregates in the 100 kDa range.

In any case, HF and DFT are both limited by the same computational bottlenecks and thus, in practice, the two methods have very similar computational cost. The same can **not** be said about the methods' accuracy, however, as we now discuss.

Accuracy of SCF Calculations

Even though Hartree Fock and DFT are much more accurate than qualitative MO theory, they are still an approximation. The accuracy varies depending on the property one is interested in, and the reliability of HF/DFT for various predictions has been extensively tested empirically. We now summarize a few key rules of thumb that come out of those computer experiments. Note that, as discussed above, the quality of DFT results depends on the functional used. The figures given below are characteristic of the B3LYP functional (one of the better forms available). In the future, as new and better functionals are developed, these error bars may improve.

Ionization Potentials and Electron Affinities: Just as for qualitative MO theory, we can use SCF calculations to compute the IP and EA of a molecule, M. There are two ways we can do this, which will give somewhat different answers. The first is to use Koopman's theorem. Here we do just one calculation on the neutral molecule and then estimate the IP (EA) from ε_{HOMO} (ε_{LUMO}). In the second route, we perform two separate calculations on the neutral and cation (or anion) to compute the IP (or EA) directly from IP=E[M$^+$]-E[M] (or EA=E[M]-E[M$^-$]). In practice the first route is somewhat faster, while the second is somewhat more accurate. Taking the more accurate route, HF predicts IPs and EAs to within about 0.5 eV. This isn't perfect, but is close enough to sometimes be a useful guide. DFT on the

other hand predicts IPs and EAs to within about 0.2 eV. Thus, DFT is not perfect, either, but a clear improvement.

Bond Lengths: Given the positions of the nuclei, any electronic structure technique gives us a prescription for the energy. Thus, we have an energy that depends on the positions of all the atoms – E(R_1,R_2,R_3,…R_M). This energy is what have called the **potential energy surface** (PES), and it contains a wealth of information, about stable conformations of the molecule – local minima on the PES. To find these, we are interested in points where $\nabla E(\mathbf{R}_I^{eq}) = 0$. It turns out that HF is fairly reliable for predicting these stable structures – the resulting bond lengths are typically about 1% too short. DFT is even a little more accurate still, giving bond lengths correct to within about 1 pm (0.01 Å) As SCF calculations are often a very useful guide to structural information.

Vibrational Frequencies One can also obtain the energies of vibrational transitions from the PES. The most accurate way to do this would be to solve the Schrödinger equation for the nuclei moving on the PES

$$\left(\hat{T} + E_{PES}(\mathbf{R}_I)\right)\Psi_{nuc}(\mathbf{R}) = E_v \Psi_{nuc}(\mathbf{R})$$

where $\hat{T}$ is the nuclear kinetic energy operator. Often, a much simpler approximation suffices. Around one of the stable minima, one can approximate the PES using a power series expansion:

$$E_{PES}(\mathbf{R}_I) \approx E_{PES}(\mathbf{R}_I^{eq}) + \tfrac{1}{2}(\mathbf{R}_I - \mathbf{R}_I^{eq}) \cdot E_{PES}''(\mathbf{R}_I^{eq}) \cdot (\mathbf{R}_I - \mathbf{R}_I^{eq})$$

Thus we see that the energy looks like a Harmonic oscillator with a force constant $E_{PES}''(\mathbf{R}_I^{eq})$. Thus, if one computes the second derivatives of the PES around the minimum, one can directly predict the harmonic vibrational frequencies of the molecule. HF does a passable job of predicting these, typically overestimating the correct values by 10%. DFT is, once again, quite a bit better – it underestimates by about 3% on average.

Reaction Barrier Heights: In predicting the rates of chemical reactions, one is very interested in the energetic barrier between reactants and products. One can approximate this in a calculation by looking for saddle points on the PES. Here, one still looks for configurations where $\nabla E_{PES} = 0$, but at least one of the second derivatives is negative, indicating that the energy is maximized (rather than minimized) along at least one coordinate. Finding saddle points is much more difficult than finding minima, because they are

inherently unstable. However, once these points have been found, one can predict the barrier height by computing energy difference between the minimum and the saddle point. HF is not terribly good at predicting these barrier heights, typically overestimating them by 30-50%. Unfortunately, current density functionals are not very good at this, either, typically underestimating barrier heights by ~25%. The challenge of improving barrier heights in DFT (without making other properties worse!) is an active area of research.

<u>Binding Energies</u> One of the most important properties we can predict is the binding energy of a complex. We can obtain the binding energy of AX by doing three separate calculations (on AX, A and X). The binding energy is then $\Delta E = E(AX) - E(A) - E(X)$. Unfortunately, HF is quite bad at predicting binding energies in this fashion, underestimating them by 50% in practice. DFT on the other hand is very successful at predicting binding energies, with a typical error of about 0.1 eV

Putting all this data together, we can make a table of expected accuracy for HF and DFT:

Property	HF Accuracy	DFT Accuracy
IPs and EAs	±0.5 eV	±0.2 eV
Bond Lengths	-1%	±1 pm
Vibrational Frequencies	+10%	-3%
Barrier Heights	+30-50%	-25%
Bond Energies	-50%	±0.1 eV

Because DFT is essentially universally more accurate than HF but comes at the same computational cost, DFT calculations have taken over the quatum chemistry literature.

Correlated Wave Function Approximations
While DFT gives us terrific bang for our buck, the one thing it does not provide is a convergent hierarchy for doing calculations. We have some empirical rules of thumb that one functional should be better than another, but no guarantees. And further, if the best functional (e.g. B3LYP) gives the

wrong answer, we have no avenue by which we can improve it. To obtain something that is more controlled, we use methods that are commonly referred to as correlated wave function methods. Here, the idea is to employ wave functions that are more flexible than a Slater determinant. This can be done by adding up various combinations of Slater determinants, by adding terms that explicitly correlate pairs of electrons (e.g. functions that depend on $\mathbf{r}_1$ and $\mathbf{r}_2$ simultaneously) and a variety of other creative techniques. These approaches are all aimed at incorporating the **correlation** between electrons – i.e. the fact that electrons tend to spend more time far apart from one another as opposed to close together. This correlation reduces the average repulsion employed in HF and brings us closer to the true ground state energy.

First, we must properly frame the task. The best ground state wave function we have obtained so far comes from HF (in technical terms, DFT only gives us a good *density*, not a good wave function). Of course, the HF wave function is *not* the true ground state wave function and the HF energy is not the true ground state energy

$$\Psi_{HF} \neq \Psi_{exact} \qquad\qquad E_{HF} \neq E_{exact}$$

In particular, we know from the variational principle that the HF energy is above the ground state, so we define the **correlation** energy as

$$E_{corr} = E_{exact} - E_{HF} \ .$$

This energy captures the energy difference between our HF estimate and the correct ground state energy. It is negative (stabilizing) by definition and physically arises from the fact that electrons are not independent (as in HF) but move in a correlated fashion so that they avoid one another and thus minimize the electron-electron repulsion. As a rule of thumb, E_{HF} is about 99% of the total energy, so that E_{corr} is only about 1% of the total energy. However, this 1% determines a whole range of chemical properties (recall how bad HF is for predicting, say, bond energies) and so it is crucial that we describe it accurately.

Hence, our goal is to derive the wave function corrections necessary to turn E_{HF} into E_{corr} – or at least get a better approximation. There are generally two routes to this: the first treats the missing correlation as a perturbation, the second seeks to directly expand the wave function and use the tools of matrix mechanics.

Perturbation Theory

Because the correlation energy is usually small, perhaps the first impulse is to treat it as a perturbation. To this end, we look at the many electron Hamiltonian

$$\hat{H}_{el} = Kinetic\ Energy + Electron\,/\,Nuclear\ Attraction + Electron\ Electron\ Repulsion$$

and immediately recognize that we can split this into a reference (the first two terms) and a perturbation (the last term). The reference can then be treated using MO theory and we include the effects of the electron-electron interaction as a perturbation to first, second … order in the hopes of describing E_{corr} accurately. This is the general idea of many-body perturbation theory. In practice, the naïve choice is sub-optimal, because it uses the *non-interacting* MOs (i.e. without the mean field of electron repulsion) as a reference. It is much better to use the HF MOs as a reference. We can accomplish this by defining the Fock operator $\hat{F}$ (whose matrix representation is the Fock matrix) and writing

$$\hat{H} = \hat{F} + \left(\hat{H} - \hat{F}\right) \equiv \hat{H}^{(0)} + \hat{H}^{(1)}$$

Now, the reference ground sate will be HF and the perturbation will be $\hat{H} - \hat{F}$. A little bit of work shows that physically this perturbation is the electron-electron interaction minus the mean field predicted by HF. Thus, the perturbation is smaller – since we've subtracted the average, we only have to worry about *fluctuations* from the average – and so we expect perturbation theory to work a little better with this reference. This particular choice of reference was first proposed by Møller and Plesset (MP) and the nth order in the expansion is known as MPn theory.

After some significant algebra, one can show that the MP1 energy (i.e including the interaction to first order) is identical to Hartree-Fock. Thus, the first order correlation energy is zero. This fact is one of the strongest arguments for using the HF reference: the interaction is, in some sense, as weak as possible because it has no effect on the energy to first order. Thus, the second order correction (MP2) is the lowest non-vanishing term and it is this level of MP theory that is most commonly used. The empirical accuracy of MP2 theory is detailed in the table below. As you can see, MP2 is sometimes better than something like B3LYP, sometimes worse. But MP2 has the benefit that, if we are unsure of our answer we can in principle go to MP3, MP4,… to verify that the calculation is correct.

| Property	HF	DFT	MP2
IPs and EAs	±0.5 eV	±0.2 eV	±0.2 eV
Bond Lengths	-1%	±1 pm	±1 pm
Vibrational Frequencies	+10%	+3%	+3%
Barrier Heights	+30-50%	-25%	+10%
Bond Energies	-50%	±3 kcal/mol	±10 kcal/mol

In practice, one very rarely sees MP3 or MP4 calculations in the literature for two reasons. The first is cost. If we compile the computational cost of MP2 as compared to HF or DFT we find a somewhat sobering situation:

| Resource	HF/DFT	MP2
Storage	$O(N^2)$	$O(N^3)$
CPU Time	$O(N^3)$	$O(N^{4-5})$
Maximum Feasible N	≈10,000	≈1,000
Maximum Feasible Atoms	≈1,000	≈50

Here we define the number of basis functions to be N, and keep track of the highest power of N required by the method with the notation $O(N^x)$. Thus, MP2 requires exponentially more time and space than the corresponding SCF calculation. This is compounded by the fact that, as discussed at the end of the basis set discussion, an accurate MP2 calculation requires a more accurate AO basis (e.g. QZ2P) to get the type of precision that HF or DFT can get in a smaller basis (e.g. DZP). Thus, with an MP2 calculation, one is limited to much smaller molecules than with DFT or HF. Going to MP3 or

even MP4 requires more storage and more CPU time, making them almost prohibitively expensive.

The other problem is that the perturbative expansion will only converge if the perturbation is, in some sense, "small". A precise characterization of when the MP series does and does not converge is difficult to come by. By its nature peturbation theory is what is called an asymptotic expansion, and it is very difficult to guarantee convergence of these types of expansions. In practice, one can simply calculate very high orders of the MPn hierarchy and look for convergence. In a very good case, we might see something like the figure below, which shows the MPn energies for the BH molecule as a function of n:
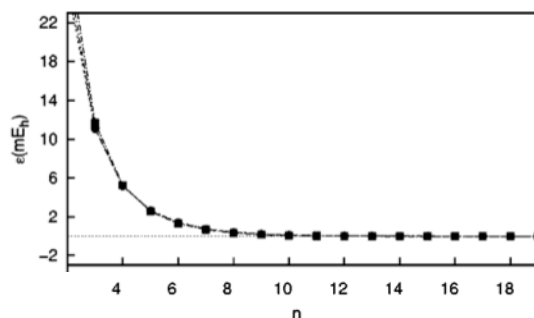
Here, we see that the energy converges monotonically and rapidly with n: MP2 is within 0.012 au (0.3 eV) of the exact total energy (which is the dotted line), MP4 is within .005 au (0.1 eV), MP5 is within .002 au (0.03 eV) …. Meanwhile for N_2 we see
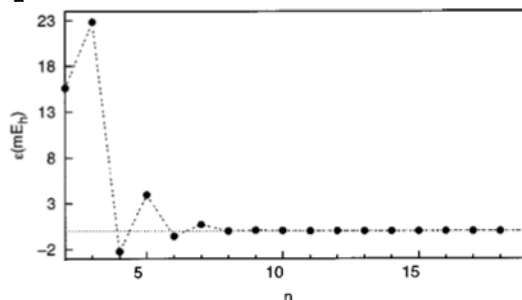
which has somewhat slower convergence as well as some oscillation. For neon we see very fast near-convergence, followed by divergence(!):
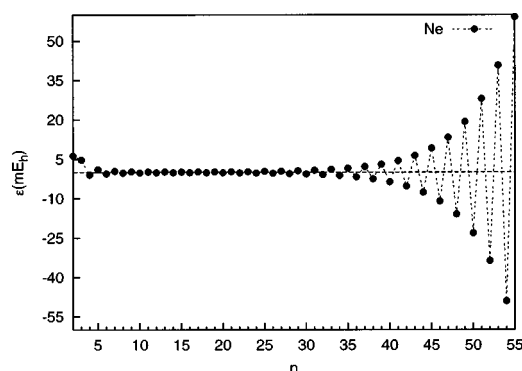
Coupling these findings with the fact that the cost of an MPn calculation increases dramatically as we increase n, it is very rarely advisable to do MP3 or MP4 calculations in practice. MP2 is the least expensive of the MP methods and statistically it is about as good as any of the other orders.

Configuration Interaction (CI)
In order to arrive at something that has reliable convergence, we have to attack the wave function directly This is the idea behind configuration interaction (CI), which is based on the familiar steps of matrix mechanics. First, we find a complete basis. Then operators become matrices, wave functions become vectors and all our quantum operations just become matrix algebra. For example, we can find the ground and excited state energies by looking for the eigenvalues of $\mathbf{H}$. It may seem like we have already done this with MO theory, but in MO theory we restricted our attention to one electron functions (orbitals) and in CI we take the brute force approach of looking at N particle functions, which results in the exact answer, albeit at a much higher cost.

In order to define a complete basis, we use what is called Fock's theorem. Given a complete set of orthonormal one electron functions (e.g. all the occupied and virtual molecular orbitals, ψ_i) then any many electron wave function can be written as a sum of all possible Slater determinants:

$$\Psi(1,2,3,...) = \sum_{p<q<r<...} C_{pqr...} \begin{vmatrix} \psi_p(1) & \psi_q(1) & \psi_r(1) & \cdots \\ \psi_p(2) & \psi_q(2) & \psi_r(2) & \\ \psi_p(3) & \psi_q(3) & \psi_r(3) & \\ \vdots & & & \ddots \end{vmatrix} \equiv \sum_{p<q<r<...} C_{pqr...} \left| \psi_p \psi_q \psi_r ... \right|$$

Here, the dots just indicate that the number of electrons is arbitrary. For any particular number of electrons, the list terminates. Thus, for four electrons:

$$\Psi(1,2,3,4) = \sum_{p<q<r<s} C_{pqrs} \begin{vmatrix} \psi_p(1) & \psi_q(1) & \psi_r(1) & \psi_s(1) \\ \psi_p(2) & \psi_q(2) & \psi_r(2) & \psi_s(2) \\ \psi_p(3) & \psi_q(3) & \psi_r(3) & \psi_s(3) \\ \psi_p(4) & \psi_q(4) & \psi_r(4) & \psi_s(4) \end{vmatrix}$$

Using this expansion, we can represent all wave functions as vectors in this space and all operators as matrices. Thus, we can find all the energy eigenstates by solving an orthogonal eigenvalue problem:

$$\mathbf{H} \cdot \mathbf{C}_i = E_i \mathbf{C}_i.$$

This expansion is known as Full CI and it is guaranteed to give us the right answer – unlike perturbation theory (where there are an infinite number of terms) here the expansion is finite and there is no question of convergence. The problem is that the dimension of this eigenvalue problem is <u>very</u> large. Say we have a benzene molecule in a minimal basis, which has 42 electrons and around 100 AOs. To enumerate the number of terms in the sum, we recognize that there are (100 possible values of p) times (99 possible values of q) times (98 possible values of r) ... for a total of $\begin{pmatrix} 100 \\ 42 \end{pmatrix} \approx 1x10^{30}$ unique

basis states! Thus something the size of benzene is out of the question, and FCI will be limited to very small molecules (like diatomics).

In order to get something more tractable, we first rearrange the expansion. We note that one of the determinants involved will be the HF determinant, Ψ_{HF}. There will also be a number of determinants that differ from the HF state by only one orbital (e.g. orbital i gets replaced by orbital a). This singly substituted states will be denoted Ψ_i^a. Next, there will be determinants that differ by two orbitals (e.g. i and j get and replaced by a and b, denoted Ψ_{ij}^{ab}). Similarly, we will have triple, quadruple, quintuple ... substitutions which allows us to reorganize the expansion as

$$\Psi = \Psi_{HF} + \sum_{ia} C_i^a \Psi_i^a + \sum_{ia} C_{ij}^{ab} \Psi_{ij}^{ab} + ...$$

This expansion will terminate only after we have performed N substitutions (where N is the number of electrons). Because this is an exact re-writing of the CI expansion, we have saved no effort in this step. However, this way of writing things suggests a useful approximation. We recognize that the HF

state is a pretty good approximation to the ground state. As a result, determinants that differ by many, many orbitals from Ψ_{HF} shouldn't be that important in the expansion, and so it is natural to truncate the expansion at some level. Thus, if we only include Single substitutions, we have CIS. If we include Double substitutions, too, we have CISD. Adding Triples gives CISDT and so on. In any of these cases, we still obtain the eigenstates by solving the linear eigenvalue problem $\mathbf{H} \cdot \mathbf{C}_i = E_i \mathbf{C}_i$.

However, we have reduced the dimension of our Hilbert space (e.g. so that it only includes HF plus single and double substitutions) so the dimension of our matrices is much, much smaller. These forms of truncation allow CI expansions to be routinely used in quantum chemistry.

There is one other way to truncate CI that is based more on chemical intuition. First, we recognize that each of the terms above can be associated with a group of stick diagrams:

$$\left| \Phi_{HF} \right\rangle =$$

$$\sum_{i\alpha} C_i^a \left| \Phi_i^a \right\rangle =$$
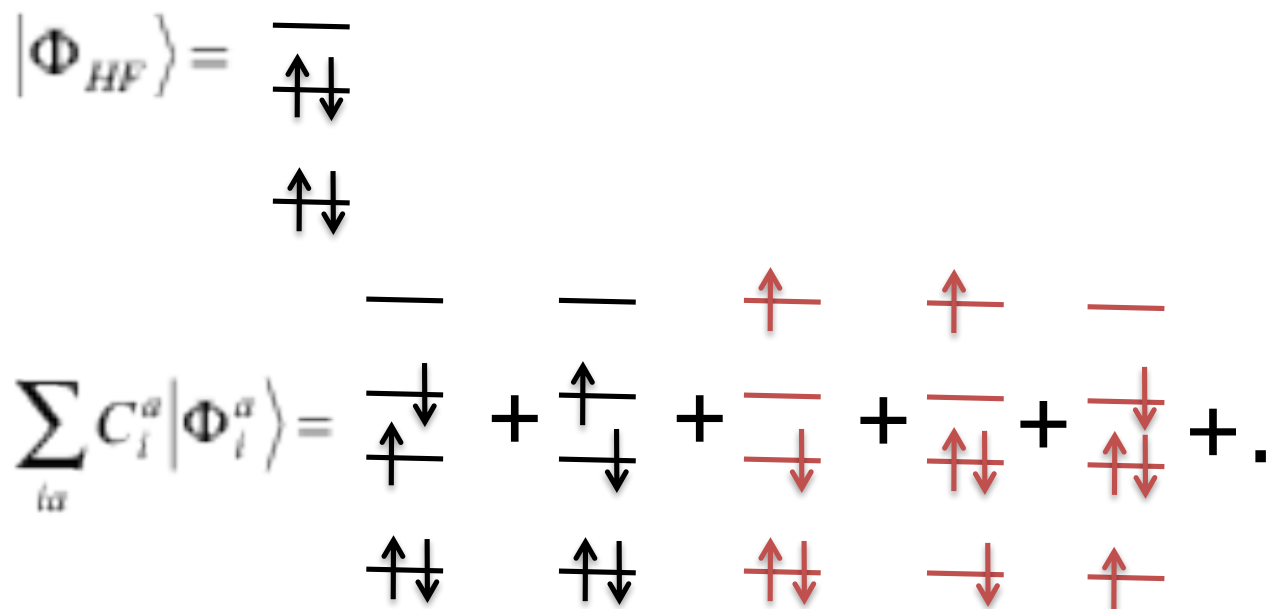
$$\sum_{ijab} C_{ij}^{ab} \left| \Phi_{ij}^{ab} \right\rangle =$$

Thus, each of the basis states corresponds to a sum of many electron configurations. The off-diagonal elements of the Hamiltonian matrix encode interactions between the configurations, which is where CI gets its name. Note that this picture is somewhat different than the physical picture we have in our head of *electrons* interacting, but mathematically the result is the same. In any case, writing the expansion in this way, one sees that there are many strange terms that look as if they can hardly have any chemical relevance – ones that excite out of core orbitals and into high-energy virtual orbitals, excitations that unpair numerous spins…. It thus seems reasonable to manually pick the configurations one thinks will be most relevant to the problem at hand. Thus, one might include the black configurations and ignore the red ones. This is the strategy behind multiconfiguration SCF (MCSCF). One particularly successful variant of MCSCF is CASSCF, where one selects a subset of important orbitals (e.g. HOMO, LUMO and LUMO+1) and then includes all the excitations that involve <u>only</u> these orbitals. In either case, it is clear that the answer obtained here will depend on how good our intuition is: if we pick the right configurations, we can get very good results. But if our intuition is bad and we pick bad configurations we will get bad results. In practice, matters are not quite so dire as this, because one always has the convergence of the CI expansion to fall back on: if all else fails, including more and more configurations will eventually lead to the correct result if we have enough patience.

There is one significant problem with truncated CI methods. Suppose you approximate the ground state of a given molecule (M) with CISD:

$$\Psi_M = \Psi_{HF} + \sum_{ia} C_i^a \Psi_i^a + \sum_{ia} C_{ij}^{ab} \Psi_{ij}^{ab} \equiv \Psi_{HF} + \Psi_S + \Psi_D$$

Where Ψ_S and Ψ_D are sums of singly and doubly excited sates, respectively. Next, consider a pair of M molecules very far apart. At the same level of approximation, the correct dimer wave function will be a product

$$\Psi_{M...M} = \Psi_M = \left(\Psi_{HF} + \Psi_S + \Psi_D\right)\left(\Psi_{HF} + \Psi_S + \Psi_D\right)$$

$$= \Psi_{HF}\Psi_{HF} + \left(\Psi_{HF}\Psi_S + \Psi_S\Psi_{HF}\right) + \left(\Psi_{HF}\Psi_D + \Psi_D\Psi_{HF} + \Psi_S\Psi_S\right)$$

$$+ \left(\Psi_S\Psi_D + \Psi_D\Psi_S\right) + \Psi_D\Psi_D$$

In the last equality, we have grouped the terms according to the number of substitutions in the dimer: the first term has no substitutions (since it is just the product of the Hartree-Fock determinants for the two monomers), the second has singles, the third doubles, the fourth triples and the last quadruples [Note that the exictation levels add here. Hence, if we have two orbitals substituted on the left *and* two substituted on the right, the total wave function has four substitutions.] Thus, we see that if we include only double substitutions for the monomer, that implies we need <u>quadruple</u> substitutions for the dimer (and six-fold substitutions for the trimer, eight-fold for the tetramer…) Hence, *any fixed level of truncation (e.g. to doubles or triples or quadruples) will fail to describe non-interacting dimer wave functions correctly.* This failure is called the size consistency problem. It manifests itself in a number of ways. One prominent mistake here is that, since CISD wave functions never decompose as products, the energies do not add properly:

$$E_{CISD}(A...X) \neq E_{CISD}(A) + E_{CISD}(X).$$

In fact, one can even prove the more stringent fact that for very large systems, the correlation energy per particle predicted by CISD goes to zero! Thus, one finds strange results like the following: if you compute the CISD energy of an infinite number copies of molecule A that are all far from one another, the total energy per molecule will actually come out the same as the HF energy for A:

$$\lim_{n\to\infty} \frac{E_{CISD}(A_n)}{n} = E_{HF}(A)$$

This error is called the size extensivity error. It reflects the fact that, while we know the energy is an extensive quantity, in CISD the correlation energy is not extensive (i.e. it does not grow linearly with *N*). Thus, when the system is large enough, the CISD correlation per particle is negligible.

Coupled Cluster (CC)

Because of the problems with size extensivity and size consistency, it is not meaningful to discuss the quality of results one expects with CI. The rules would depend strongly on the size of the system and even on the presence or absence of spectator molecules. Instead, we need to modify the CI expansion. Toward this end, we define an operator, $\hat{C}$, that generates all the substituted configurations in the CI expansion:

$$\hat{C}\Psi_{HF} \equiv \sum_{ia} C_i^a \Psi_i^a + \sum_{ia} C_{ij}^{ab} \Psi_{ij}^{ab} + ...$$

We can use this operator to generate the entire CISD wave function:

$$\Psi_{CISD} = \left(1 + \hat{C}\right)\Psi_{HF} \equiv \Psi_{HF} + \sum_{ia} C_i^a \Psi_i^a + \sum_{ia} C_{ij}^{ab} \Psi_{ij}^{ab}$$

As we saw, this CI expansion led to difficulties with size extensivity. Rather than this linear prescription, Coupled Cluster postulates an *exponential* form for the wave function. Thus, the CCSD wave function is

$$\Psi_{CCSD} = e^{\hat{C}}\Psi_{HF} = \left(1 + \hat{C} + \tfrac{1}{2}\hat{C}^2\right)\Psi_{HF}$$

$$= \Psi_{HF} + \hat{C}\Psi_{HF} + \tfrac{1}{2}\hat{C}^2\Psi_{HF} + ...$$

In the last expression, the first term has zero substitutions, the second has up to double substitutions, the third has **up to quadruple** substitutions,.... While this exponential form may seem artificial at first, it is easily confirmed that this format alleviates the size consistency problem. Suppose molecules A and B are well represented by CCSD

$$\Psi_{CCSD}^A = e^{\hat{T}_A}\Psi_{HF}^A \qquad \Psi_{CCSD}^B = e^{\hat{T}_B}\Psi_{HF}^B$$

Then the noninteracting A...B heterodimer will be well-represented by a product:

$$\Psi_{CCSD}^{A...B} = \Psi_{CCSD}^A \Psi_{CCSD}^B = e^{\hat{T}_A}\Psi_{HF}^A e^{\hat{T}_B}\Psi_{HF}^B = e^{\hat{T}_A}e^{\hat{T}_B}\Psi_{HF}^A \Psi_{HF}^B$$

$$= e^{\hat{T}_A + \hat{T}_B}\Psi_{HF}^A \Psi_{HF}^B = e^{\hat{T}_{AB}}\Psi_{HF}^{A...B}$$

Where on the last line we have made the definitions $\hat{T}_{AB} \equiv \hat{T}_A + \hat{T}_B$ and $\Psi_{HF}^{A...B} \equiv \Psi_{HF}^A \Psi_{HF}^B$. Now, if $\hat{T}_A$ and $\hat{T}_B$ involve at most double substitutions, then $\hat{T}_{AB} \equiv \hat{T}_A + \hat{T}_B$ **will also involve at most doubles**. Thus, the product of two noninteracting CCSD wave functions is another CCSD wave function. We conclude that a truncated CC prescription is capable of describing noninteracting fragments correctly. As a result, the CCSD energy is properly size consistent and size extensive and consequently preferable to CISD for looking at large systems.

There are truncated CC cousins for every CI approximation (e.g. CCS, CCSD, CCSDT, CCSDTQ...). There is one approximation that has gained widespread

appeal – CCSD(T). As the name suggests, this includes all single and double substitutions and an approximate (perturbative) treatment of triple substitutions. Because of the exponentiation, the algebra involved in the CC equations is quite complex – note that we haven't even touched on how one would solve for the C_i^a and C_{ij}^{ab} coefficients. However, from a conceptual point of view, these details are less important. Grossly, we understand CCSD as involving the same chemistry as CISD, but with a more elegant treatment of the excitations that deals properly with independent fragments. The additional terms in CCSD do not increase the computational cost significantly relative to CISD; as a result CCSD calculations now far outnumber CISD calculations.

CC theory can give very accurate descriptions of chemical properties. Filling in our table:

Property	HF	DFT	MP2	CCSD(T)
IPs and EAs	±0.5 eV	±0.2 eV	±0.2 eV	±0.05 eV
Bond Lengths	-1%	±1 pm	±1 pm	±0.5 pm
Vibrational Frequencies	+10%	+3%	+3%	±5 cm^{-1}
Barrier Heights	+30-50%	-25%	+10%	±2 kcal/mol
Bond Energies	-50%	±3 kcal/mol	±10 kcal/mol	±1 kcal/mol

Thus we see that CCSD(T) is a phenomenal improvement over DFT. The accuracy of CCSD(T) is such that the statistical errors are getting down to where small terms we have neglected (like relativity and the finite mass of the nucleus) actually become about as large as the inherent errors in CCSD(T). For this reason, CCSD(T) has been termed the gold standard of quantum chemistry. Indeed, CCSD(T) is as good or better than many experiments at determining precise molecular properties.

Unfortunately, for correlated methods there is no free lunch – the increased accuracy comes at a cost: CCSD is exponentially more expensive than even MP2:

Resource	HF/DFT	MP2	CCSD
Storage	$O(N^2)$	$O(N^3)$	$O(N^4)$
CPU Time	$O(N^3)$	$O(N^{4-5})$	$O(N^6)$
Maximum Feasible N	$\approx 10{,}000$	$\approx 1{,}000$	≈ 500
Maximum Feasible Atoms	$\approx 1{,}000$	≈ 50	≈ 10

As computers get faster and algorithms get smarter, the number of feasible atoms with each method will increase, but the correlated methods like CCSD and MP2 will always be relatively more expensive.
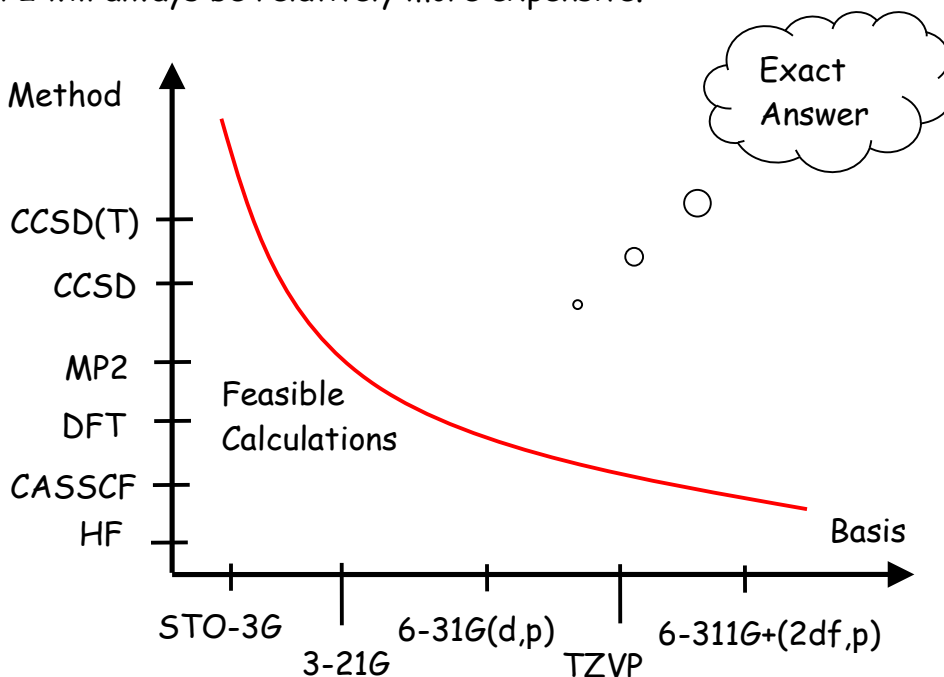
Summary

Combining what we have learned, then, the approximations we can make fit nicely into a two-dimensional parameter space (see above). On the one axis, we have the basis sets we can choose from. On the other, we have the different methods for approximating the energy. The get close to the exact answer, we need to employ a large basis set and an accurate energy method. Unfortunately, both increasing the size of the basis and improving the method tend to slow our calculations down. Given that we don't want to wait years and years to find out the result of a calculation, modern

computers therefore limit how accurate our answers can be (as illustrated with the red line above). As we become experts at what is and is not feasible with current computing power, we become better able to get good accuracy for a variety of problems.

5.61 Physical Chemistry
Fall 2017