Ab initio methods: Post-Hartree–Fock

Alston J. Misquitta

Centre for Condensed Matter and Materials Physics
Queen Mary, University of London

March 6, 2012
Hartree–Fock equations I

We have introduced basis sets in solving the Hartree–Fock equations, but what exactly are they?

Brief summary of Hartree–Fock theory:

- Born–Oppenheimer approximation: fixed nuclei.
- Ground-state $N$-electron wavefunction: single Slater determinant:
  \[
  \psi_0(x_1, x_2, \cdots, x_N) = |\chi_1 \chi_2 \cdots \chi_N\rangle
  \] (1)
- Variational Principle:
  \[
  E_0 \leq E^{\text{HF}} = \min \langle \psi_0 | \mathcal{H} | \psi_0 \rangle
  \] (2)

subject to the conditions that the spin-orbitals are orthonormal.
Hartree–Fock equations II

- Fock equations:
  \[ f(i)\chi(x_i) = \epsilon\chi(x_i) \]  
  where \( f(i) \) is an effective operator called the **Fock operator**

  \[ f(i) = -\frac{1}{2}\nabla_i^2 - \sum_{\alpha} \frac{Z_{\alpha}}{r_{i\alpha}} + \nu_{\text{HF}}(i) \]  
  where \( \nu_{\text{HF}}(i) \) is the Hartree–Fock *effective* potential.

- Self-consistent solution (iterations needed).

- Infinity of solutions to the Fock equations.

- Introduce a basis for the spatial part of spin-orbitals:
  \[ \chi_i(r) = \sum_m C_{im}\phi_m(r) \]
Hartree–Fock equations III

Leads to linear equations:

$$\mathbf{FC} = \epsilon \mathbf{SC}$$  \hspace{1cm} (6)

Q: How do we choose the basis?
A reasonable choice for basis sets for finite systems would be what are called Slater-type orbitals: these are very like solutions of the 1-electron Hamiltonian. They differ in two ways: (1) the radial part is simpler and (2) the exponent is not integral but can be varied to account for screening effects.

\[ \phi = R_{nl}(r)Y_{lm}(\theta, \phi) \quad (7) \]

where \( Y_{lm} \) is a (real) spherical harmonic and the radial part is given by

\[ R_{nl}(r) = \frac{(2\zeta)^{n+1/2}}{[(2n)!]^{1/2}} r^{n-1} e^{-\zeta r} \quad (8) \]
Slater-type orbitals II

- \( \zeta = 1.0, n = 1 \)
- \( \zeta = 1.0, n = 2 \)
Comments on Slater-type orbitals:

- **GOOD** Nuclear cusp condition satisfied.

\[ \frac{\partial}{\partial r} \langle \rho(r) \rangle_{\text{sph}} \bigg|_{r=0} = -2Z \langle \rho(0) \rangle_{\text{sph}} \]

- **GOOD** Exact wavefunction has the long-range form of a Slater orbital.

  If we pull one electron out of an \( N \)-electron molecule the wavefunction behaves like

\[ \psi(N) \rightarrow \psi(N - 1) \times e^{-\sqrt{2I}r} \]

  where \( I \) is the first (vertical) ionization energy.

- **BAD** Integrals very difficult for multi-atom systems.
In 1950 S. F. Boys pointed out that the problem of computing integrals could be resolved by using not Slater-type orbitals, but rather Gaussian-type orbitals (GTOs):

\[ R_{nl} \sim r^l e^{-\alpha(r-A)^2} \]  \hspace{1cm} (9)

where \( A \) is the centre of the GTO. The main reason for the efficacy of GTOs is that the product of two GTOs is a third GTO, centred at a point in between:

\[ \exp(-\alpha(r-A)^2) \exp(-\beta(r-B)^2) = \exp(-\gamma(A-B)^2) \exp(-\mu(r-P)^2) \]

where \( \mu = \alpha + \beta \), \( \gamma = \alpha \beta / \mu \) and \( P = (\alpha A + \beta B) / \mu \).
Gaussian-type orbitals (GTOs) II
Gaussian-type orbitals (GTOs) III

- **GOOD** GTOs makes the integrals that appear in the HF energy expression much simpler.
- **BAD** Nuclear cusp condition violated: zero derivative at origin.
- **BAD** Wrong long-range form: dies off too fast with distance.

The two negative points can, to some extent, be remedied by using not single GTOs, but *linear combinations* of GTOs. These groups of GTOs are termed *contractions*.

Basis sets consist of groups of contractions together with some un-contracted GTOs. The better the basis, the more of these there will be and the more GTOs in a contraction.
<table>
<thead>
<tr>
<th>Basis sets</th>
<th>CI</th>
<th>CC</th>
<th>MBPT</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian-type orbitals (GTOs) IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$$\zeta = 1.0, n = 1$$
Gaussian-type orbitals (GTOs) V

\[ \phi \]

- \( \zeta = 1.0, n = 1 \)
- STO-1G
Gaussian-type orbitals (GTOs) VI

\[ 4\pi|\phi|^2 \]

- Blue line: \( \zeta = 1.0, n = 1 \)
- Red line: STO-1G
Gaussian-type orbitals (GTOs) VII
Gaussian-type orbitals (GTOs) VIII

\[ \zeta = 1.0, n = 1 \]

STO-3G
Gaussian-type orbitals (GTOs) IX

Basis set recommendations:

- **GOOD** Complete basis set (CBS) limit
- Geometry optimization: moderate size basis sets. Double-\( \zeta \).
- Energies: At least triple-\( \zeta \) quality.
- Properties: Triple-\( \zeta \) or more.

We will have another look at basis sets after discussion correlated methods.
Q: How to we improve on the complete basis set HF results? I.e., how do we get beyond the HF limit?

- Perturbation Theory
- Coupled-cluster methods
- Density functional theory
**Configuration Interaction I**

**CI**: Increase the flexibility in the wavefunction by including in addition the HF ground state, excited states.

Q: What are excited states and how do we form them?
Configuration Interaction II

\[ n + m \quad \text{------} \quad \text{------} \]
\[ \quad \text{........} \quad \text{........} \]
\[ \quad \text{........} \quad \text{........} \]
\[ n + 1 \quad \text{------} \quad \text{------} \]
\[ \quad \uparrow\downarrow \quad \uparrow\downarrow \]
\[ \quad \text{........} \quad \text{........} \]
\[ 2 \quad \uparrow\downarrow \quad \uparrow\downarrow \]
\[ 1 \quad \uparrow\downarrow \quad \uparrow\downarrow \]

**Figure:** Left: HF ground state configuration. Right: An example of an excited state configuration. If there are \( n \) occupied levels (2 electrons each, so \( N = 2n \)) and \( m \) virtual (un-occupied) levels, in how many ways can we form excited states? Each of these states will correspond to a Slater determinant.
Configuration Interaction III

The Full CI (FCI) wavefunction:

\[ |\psi\rangle = |\psi_0\rangle + \sum_{ar} c_{ar}^r |\psi_{a}^r\rangle + \sum_{abrs} c_{ab}^{rs} |\psi_{ab}^{rs}\rangle + \cdots \]

\[ = |\psi_0\rangle + c_{S} |S\rangle + c_{D} |D\rangle + \cdots \]

where electrons are excited from the *occupied* orbitals \( a, b, c, \cdots \) to the *virtual* orbitals \( r, s, t, \cdots \).

- **GOOD** This expansion will lead to the **exact** energy within the basis set used.
- **BAD** There are too many determinants!

\[ \frac{(2(n + m))!}{(2n)!(2m)!} \]
One solution to the problem is to use only some of the many determinants. For example we could use only double excitations. This leads to the CID method.

\[
|\psi^{\text{CID}}\rangle = |\psi_0\rangle + \sum_{abrs} c_{rs}^{ab} |\psi_{ab}^{rs}\rangle
\]

\[
= |\psi_0\rangle + c_D |D\rangle
\]

BAD This theory, like all truncated CI methods, is not size extensive.

Size-extensivity: If \( E(N) \) is the energy of \( N \) non-interacting identical systems then a method is size-extensive if \( E(N) = N \times E(1) \).
Q: Is CID size-extensive?

If $\hat{T}_2$ is an operator that creates all double excitations, then we can write the CID wavefunction as

$$|\psi^{\text{CID}}\rangle = |\psi_0\rangle + \sum_{abrs} c_{ab}^{rs} |\psi_{ab}^{rs}\rangle$$

$$= |\psi_0\rangle + c_D |D\rangle$$

$$= (1 + \hat{T}_2) |\psi_0\rangle$$
The CID wavefunction for each of two identical non-interacting systems will be of that form, so the combined wavefunction will be

$$\begin{align*}
|\psi_A^{\text{CID}}\rangle |\psi_B^{\text{CID}}\rangle &= (1 + \hat{T}_2(A)) |\psi_0^A\rangle (1 + \hat{T}_2(B)) |\psi_0^B\rangle \\
&= (1 + \hat{T}_2(A) + \hat{T}_2(B) + \hat{T}_2(A)\hat{T}_2(B)) |\psi_0^A\rangle |\psi_0^B\rangle
\end{align*}$$

The last excitation term is a quadruple excitation so it will not be present in the CID wavefunction for the combined A and B systems. Therefore

$$E_{\text{CID}}(AB) \neq E_{\text{CID}}(A) + E_{\text{CID}}(B).$$
Cl is not size-extensive: $H_2$

$\sigma_u \quad \downarrow \quad \uparrow \uparrow$

$\sigma_g \quad \uparrow \downarrow \quad \uparrow \quad \text{———}$

**Figure:** An energy level diagram for $H_2$ with a minimal basis. The two MOs are as shown. We can form three determinants from them. Left is the ground state. Middle is a *singly excited* determinant. Right is a *doubly excited* determinant. For reasons of symmetry the middle configuration does not contribute to the Cl expansion.
The problem of truncated CI methods is severe enough that using them is very problematic. A resolution to the problem is the class of coupled-cluster theories. In these the wavefunction is defined as:

\[ |\psi^{CC}\rangle = \exp(\hat{T}) |\psi_0\rangle \]

where \( \hat{T} \) is an appropriate excitation operator.
Coupled-cluster Theory II

For example, in CCD theory we use $\hat{T}_2$ as the excitation operator. This gives:

$$ |\psi^{\text{CCD}}\rangle = \exp(\hat{T}_2) |\psi_0\rangle $$

$$ = (1 + \hat{T}_2 + \frac{1}{2!} \hat{T}_2 \hat{T}_2 + \cdots) |\psi_0\rangle $$

The first two terms give us CID theory. The rest are needed to make CCD size-extensive:

$$ \exp(\hat{T}_2(A)) |\psi_0^A\rangle \times \exp(\hat{T}_2(B)) |\psi_0^B\rangle = \exp(\hat{T}_2(A) + \hat{T}_2(B)) |\psi_0^A\rangle |\psi_0^B\rangle $$

$$ \equiv \exp(\hat{T}_2(AB)) |\psi_0^A\rangle |\psi_0^B\rangle $$
GOOD CC theories can be systematically improved.

GOOD CCSD(T) is a very accurate theory. Here single and double excitations are included as described above and triple excitations are included through a perturbative treatment.

GOOD Size-extensive.

BAD Computationally very expensive: CCSD(T) scales as $O(N^7)$. So double the system size and the calculation costs 128 times more.

BAD (kind of!) These are single-determinant theories as described. If the system is multi-configurational (more than one state contributing dominantly) the standard CC methods are not appropriate.
Møller–Plesset Perturbation Theory I

Brief recap of Raleigh–Schrödinger perturbation theory:

- Split the Hamiltonian into two parts:

\[ \mathcal{H} = \mathcal{H}_0 + \lambda \mathcal{V} \]

where \( \mathcal{H}_0 \) is a Hamiltonian which we know how to solve and \( \mathcal{V} \) contains that troublesome parts. We expect \( \mathcal{V} \) to be a perturbation so it must be small in some sense. \( \lambda \) is a complex number that will be 1 for the physical solution.
Møller–Plesset Perturbation Theory II

- Let the solutions of $\mathcal{H}_0$ be:

$$\mathcal{H}_0 \psi_i^{(0)} = E_i^{(0)} \psi_i^{(0)}$$

Here the ‘0’ indicates that these eigenvalues and eigenfunctions are of zeroth-order in the perturbation $\mathcal{V}$. We will use the short-form:

$$|\psi_i^{(0)}\rangle \equiv |i\rangle$$

- Express the solutions of $\mathcal{H}$ in a power-series:

$$\psi_i = \psi_i^{(0)} + \lambda \psi_i^{(1)} + \lambda^2 \psi_i^{(2)} + \cdots = \sum_n \lambda^n \psi_i^{(n)}$$

$$E_i = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \cdots = \sum_n \lambda^n E_i^{(n)}$$
Møller–Plesset Perturbation Theory III

- Energies can be calculated by collecting terms at various orders:

\[ E_i^{(0)} = \langle \psi_i^{(0)} | \mathcal{H}_0 | \psi_i^{(0)} \rangle = \langle i | \mathcal{H}_0 | i \rangle \]
\[ E_i^{(1)} = \langle \psi_i^{(0)} | \mathcal{V} | \psi_i^{(0)} \rangle = \langle i | \mathcal{V} | i \rangle \]
\[ E_i^{(2)} = \langle \psi_i^{(0)} | \mathcal{V} | \psi_i^{(1)} \rangle \]

etc. for higher order terms.

- The first-order correction to the wavefunction is given by

\[ | \psi_i^{(1)} \rangle = \sum_{n \neq 0} \frac{|n\rangle \langle n | \mathcal{V} | i \rangle}{E_i^{(0)} - E_n^{(0)}} \]
Møller–Plesset Perturbation Theory IV

- So we get

\[ E_i^{(2)} = \sum_{n \neq 0} \frac{|\langle n | V | i \rangle|^2}{E_i^{(0)} - E_n^{(0)}} \]
Møller–Plesset Perturbation Theory $V$

Many-body perturbation theory (MBPT) starts from Hartree–Fock theory:

$$\mathcal{H}_0 = \sum_{i=1}^{N} f(i) = \sum_{i=1}^{n} \left( h(i) + V_{\text{HF}}(i) \right)$$  \hspace{1cm} (10)

where $h(i) = -\frac{1}{2} \nabla_i^2 - \sum_\alpha \frac{Z_\alpha}{r_{i\alpha}}$

We can now define the perturbation as

$$V = \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{r_{ij}} - \sum_{i=1}^{N} V_{\text{HF}}(i)$$ \hspace{1cm} (11)

Unlike $V_{\text{HF}}$, the perturbation $V$ is a 2-electron operator.
Møller–Plesset Perturbation Theory VI

MBPT energy at $\lambda^0$:

$$E^{(0)}_0 = \langle 0|\mathcal{F}|0 \rangle = \sum_{a \in \text{occ}} \epsilon_a$$

At first-order we get (no proof):

$$E^{(1)}_0 = -\frac{1}{2} \sum_{a=1, b=1}^N \left[ \langle ab|r^{-1}_{12}|ab \rangle - \langle aa|r^{-1}_{12}|bb \rangle \right]$$

We have not seen this before, but the sum of $E^{(0)}_0$ and $E^{(1)}_0$ is just the Hartree–Fock ground state energy:

$$E^{HF} = E^{(0)}_0 + E^{(1)}_0 \quad (12)$$

This means that we need to get to at least second-order in perturbation theory to go beyond the Hartree–Fock description.
Møller–Plesset Perturbation Theory VII

Here is what the second-order MBPT energy expression looks like:

\[
E^{(2)} = \sum_{a,b>a}^{\text{occ}} \sum_{r,s>r}^{\text{vir}} \left[ \langle ab | r_{12}^{-1} | rs \rangle - \langle ab | r_{12}^{-1} | sr \rangle \right]^2 \frac{1}{\epsilon_r + \epsilon_s - \epsilon_a - \epsilon_b}
\]  (13)

This expression is termed as MBPT2 or MP2. The latter name comes from the other name for this kind of perturbation theory: Møller–Plesset perturbation theory.

GOOD MP2 contains correlation.

BAD But not enough correlation. Problems with systems with small HOMO-LUMO gaps (band gap — HOMO is highest occupied MO and LUMO is lowest unoccupied MO).

GOOD (kind of!) It has a computational cost of $O(N^5)$. I.e., double the system in size and it will cost 32 times more computational power.

GOOD MBPT is *size-consistent*
Once we’ve got the HF solution, how do we calculate expectation values of operators and forces?