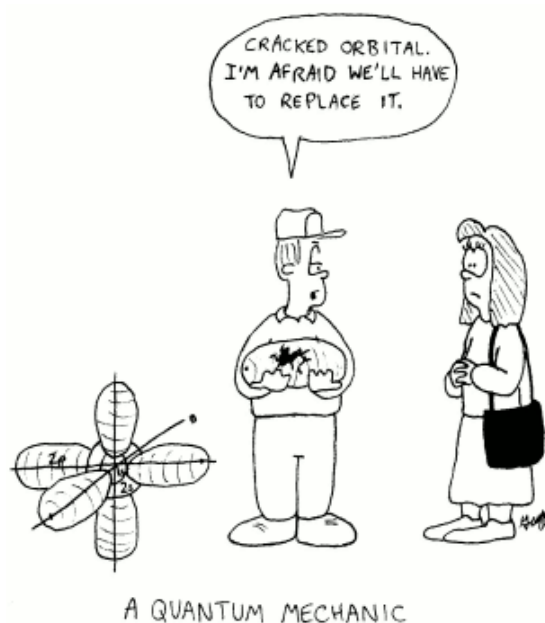




So theory guys have got it made  
in rooms free of pollution.  
Instead of problems with the reflux,  
they have only solutions...  
In other words, experimentalists  
will likely die of cancer  
From working hard, yet fruitlessly...  
till theory gives the answer.

*Thomas A. Holme, 1983*

## Practical Quantum Chemistry, L2: Methods and terminology, part B

*integral classification, semi-empirical methods, restricted and  
unrestricted SCF, correlation, M-P theory, configuration interaction*

Prof Ilya Kuprov, Weizmann Institute of Science, 2025

For all lecture notes and video records see <https://spindynamics.org>

# Integrals involved in QC calculations

The following integrals over one-electron orbitals are common:

$$\int \psi_i^*(\mathbf{r}_1) \left( -\frac{1}{2} \nabla_1^2 - \sum_n \frac{Z_n}{r_{n,1}} \right) \psi_j(\mathbf{r}_1) dV_1$$

**One-electron integrals** (over the coordinates of one electron) are cheap. There are  $O(n^2)$  such integrals for a basis set with  $n$  functions.

$$\iint \frac{\psi_i^*(\mathbf{r}_1) \psi_j(\mathbf{r}_1) \psi_k^*(\mathbf{r}_2) \psi_l(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} dV_1 dV_2$$

**Two-electron integrals** (over the coordinates of two electrons) are expensive. There are  $O(n^4)$  such integrals for a basis set with  $n$  functions.

Depending on the values of  $i, j, k, m$  indices, a two-electron integral may be called one-, two-, three- and four-centre integral. For combinatorial reasons, the number of four-centre integrals may be very large, particularly with diffuse functions.

Two specific index combinations have their own names:

$$\iint \frac{\psi_i^*(\mathbf{r}_1) \psi_i(\mathbf{r}_1) \psi_j^*(\mathbf{r}_2) \psi_j(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} dV_1 dV_2$$

**Coulomb integrals**

$$\iint \frac{\psi_i^*(\mathbf{r}_1) \psi_i(\mathbf{r}_2) \psi_j^*(\mathbf{r}_2) \psi_j(\mathbf{r}_1)}{|\mathbf{r}_1 - \mathbf{r}_2|} dV_1 dV_2$$

**Exchange integrals**

# Direct and semi-direct methods

For anything bigger than a benzene, the two-electron integral array would not fit into reasonable amounts (1-100 GB) of RAM. Two alternative strategies exist:

**Conventional methods:**  
store pre-computed integrals on disk

Integrals only need to be computed once.

Sequential access bandwidth – 100 MB/s, random access bandwidth – 1 MB/s, latency – 1-10 ms.

For very large molecules (100+ atoms) the integral arrays do not fit on the disk either.

On-the-fly data compression is possible.

**Direct methods:**  
(re-)compute integrals on-the-fly

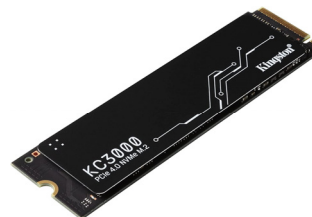
Integrals need to be computed as many times as they need to be used (typically 4-16 times).

Sequential throughput – 50 GB/s, random throughput – 10 GB/s, latency – 1-10 ns.

Only the basis description needs to be stored, which usually fits into main RAM.

Data compression is not necessary.

Semi-direct methods are the middle ground – they are useful for post-HF methods, which can store some very expensive quantities on disk.



► Latest PCIe SSDs are recommended, avoid spinning platter disks (mechanical = high latency).

# Resolution of identity approximation

**Resolution of identity** is an equation connecting a complete basis set of a vector space to the identity operator on that space:

$$\hat{E} = \sum_n |\psi_n\rangle\langle\psi_n|$$

This allows to approximate four-centre integrals with combinations of two- and three-centre integrals, reducing the integral calculation cost from  $O(n^4)$  to  $O(n^2)$ . Products of functions on either side of the four-centre integral:

$$\iint \frac{\psi_i^*(\mathbf{r}_1)\psi_j(\mathbf{r}_1)\psi_k^*(\mathbf{r}_2)\psi_l(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} dV_1 dV_2$$

are replaced by linear combinations of the functions from the *fitting basis set*:

$$\text{primary basis set} \rightarrow \psi_i^*(\mathbf{r}_1)\psi_j(\mathbf{r}_1) = \sum_k a_k^{ij} \alpha_k(\mathbf{r}_1) \leftarrow \text{fitting basis set}$$

This is a good approximation for plane-wave and Gaussian basis sets: both algebras have analytical structure constants.

# Semi-empirical approximations

One drastic approximation is *partial neglect of two-electron integrals*. Two-electron integrals are the most expensive part of the Hartree-Fock theory calculations, and the semi-empirical methods attempt to replace them with empirical parameters.

| Approximations                                                                                                                                                                                                                         | Method                                                                                                                | Average heat of formation error                                           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|
| Core electrons are not considered or approximated by effective core potentials.                                                                                                                                                        | PM6                                                                                                                   | 19 kJ/mol (HCNO), 26 kJ/mol (all)                                         |
| Minimal basis set (one function per orbital) is used.                                                                                                                                                                                  | PM5                                                                                                                   | 24 kJ/mol (HCNO), 42 kJ/mol (all)                                         |
| It is assumed that basis functions do not overlap.                                                                                                                                                                                     | PM3                                                                                                                   | 33 kJ/mol (HCNO), 80 kJ/mol (all)                                         |
| All three- and four-centre integrals are ignored. The remaining integrals are parameterised.<br><i>(the details of this approximation define the method)</i>                                                                           | AM1                                                                                                                   | 40 kJ/mol (HCNO), 101kJ/mol (all)                                         |
|                                                                                                                                                                                                                                        | MNDO                                                                                                                  | 77 kJ/mol (HCNO), 193 kJ/mol (all)                                        |
|                                                                                                                                                                                                                                        | ZINDO                                                                                                                 |                                                                           |
|                                                                                                                                                                                                                                        | MINDO                                                                                                                 | Very inaccurate (flat ammonia, incorrect rotation barriers, <i>etc.</i> ) |
|                                                                                                                                                                                                                                        | INDO                                                                                                                  |                                                                           |
|                                                                                                                                                                                                                                        | CNDO                                                                                                                  |                                                                           |
| The practical accuracy of semi-empirical methods varies greatly from system to system and from property to property.<br>As of 2025, it is only reasonable to use these methods for preliminary geometry optimisations in CHNO systems. | Wall clock time for geometry optimisation of strychnine (Gaussian16):<br><i>20 sec (PM3), 1.5 hours (HF/6-31G**).</i> |                                                                           |

# Restricted and unrestricted orbitals

The Hamiltonian used in Hartree-Fock theory does not explicitly mention spin:

Diagram illustrating the components of the Hartree-Fock Hamiltonian equation:

- electron momenta →  $-\frac{1}{2} \sum_k \nabla_k^2$
- nuclear momenta →  $-\frac{1}{2} \sum_N \frac{\nabla_N^2}{m_N}$
- electron-nuclear attraction →  $-\sum_{Nk} \frac{Z_N}{r_{Nk}}$
- inter-nuclear repulsion →  $+\sum_{N'>N} \frac{Z_N Z_{N'}}{r_{NN'}}$
- inter-electron repulsion →  $+\sum_{j>k} \frac{1}{r_{jk}}$
- energy →  $E\psi(\vec{r})$

$$\left[ -\frac{1}{2} \sum_k \nabla_k^2 - \frac{1}{2} \sum_N \frac{\nabla_N^2}{m_N} - \sum_{Nk} \frac{Z_N}{r_{Nk}} + \sum_{N'>N} \frac{Z_N Z_{N'}}{r_{NN'}} + \sum_{j>k} \frac{1}{r_{jk}} \right] \psi(\vec{r}) = E\psi(\vec{r})$$

**Restricted SCF:** spatial part the same for spin-up and spin-down.

- > Cheaper (fewer integrals to compute).
- > Wavefunction is an eigenstate of the total spin (it must be).
- > May fail in open-shell systems and bond breaking processes.
- > Smaller variational flexibility.

**Unrestricted SCF:** spin-up and spin-down orbitals are allowed to be different.

- > Can describe spin polarisation effects in open-shell systems.
- > Wavefunction is not an eigenstate of the total spin (spin contamination).
- > May display stability issues and produce errors in magnetic properties.

► Always check! Specific magnetic properties are accurate in one and dead wrong in the other.

# Spin contamination in unrestricted SCF

Unrestricted SCF methods produce wavefunctions that are not eigenfunctions of the total spin operator. This can lead to distortions in magnetic properties.

$$\langle \hat{S}^2 \rangle_{\text{UHF}} \geq \langle \hat{S}^2 \rangle_{\text{RHF}} = S(S+1) = \left( \frac{n_\alpha - n_\beta}{2} \right) \left( \frac{n_\alpha - n_\beta}{2} + 1 \right)$$

The three common approaches to dealing with spin contamination are:

1. **Annihilation** method projects out the terms with incorrect total spin at the end of the SCF procedure.
2. **Projection** method projects out the terms with incorrect total spin at each iteration within the SCF loop.
3. **Spin constraint** method introduces the deviation from the desired total spin as a Lagrange term in a constrained energy optimisation problem.

Strong spin contamination means that single determinant is not a good approximation – consider using MCSCF.

From a log produced by Gaussian16:

```
SCF Done: E(UB+HF-LYP) = -306.904337176      A.U. after 31 cycles
               Conv = 0.1424D-08              -V/T = 2.0044
               S**2 = 0.7779
```

```
KE= 3.055508118982D+02 PE=-1.230645883764D+03 EE= 3.606118207110D+02
```

Annihilation of the first spin contaminant:

```
S**2 before annihilation 0.7779, after 0.7505
```

# Static and dynamic correlation

**Electron correlation:** (in effect) “all inter-electron interactions that Hartree-Fock theory does not properly account for”.

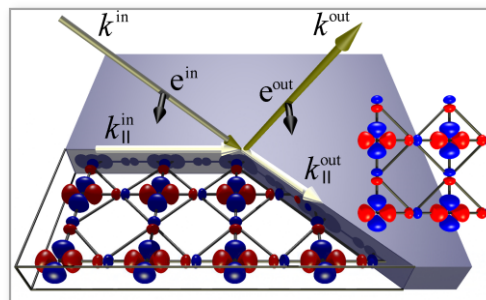
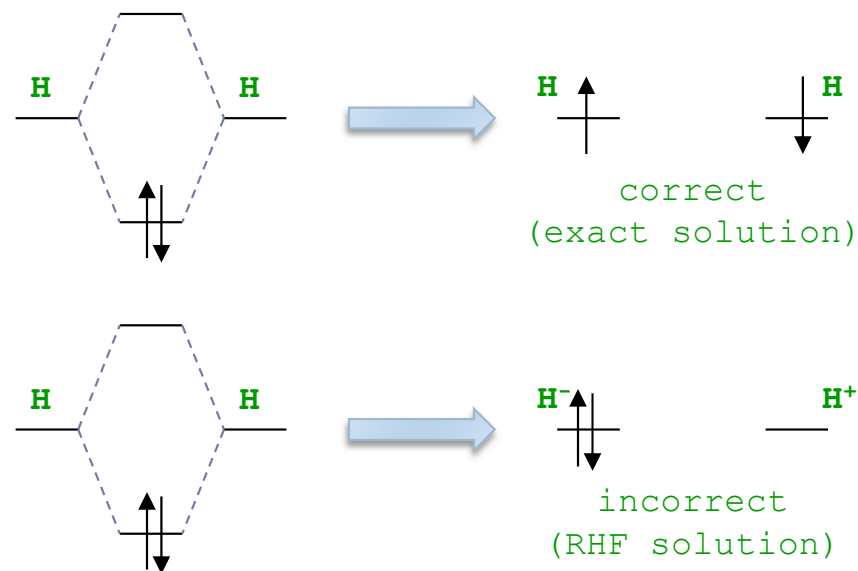
**Static correlation:** when a single Slater determinant (*i.e.* a single population pattern of molecular orbitals) is not a good approximation to the ground state. A famous example is the dissociation of the hydrogen molecule.

Important for: bond breaking, transition states, *d*- and *f*-element complexes, degenerate ground states, biradicals, large aromatic and conjugated systems.

**Dynamic correlation:** correlated *motion* of electrons due to Coulomb repulsion; not captured by effective potentials.

Important for: accurate energies, excited states, Van der Waals interactions, magnetic and optical properties.

**Example: dissociation of  $H_2$**



Accurate description of inelastic X-ray scattering cannot be obtained without the inclusion of dynamic correlation effects.

# Møller-Plesset perturbation theory

**Perturbation theory** is used when a (complicated) problem at hand differs only slightly from a (simple) problem already solved, and the Hamiltonian may be partitioned into a “reference” and a “perturbation” part:

$$\hat{H} = \hat{H}_0 + \hat{H}_1, \quad \|\hat{H}_1\| \ll \|\hat{H}_0\|$$

**Møller-Plesset perturbation theory** uses Hartree-Fock theory as a reference problem and includes everything that HF theory had ignored into the perturbation part. The calculations involve sums over all excited states:

$$E_{\text{MP2}} = \sum_{i < j}^{\text{occ}} \sum_{a < b}^{\text{vir}} \frac{\langle \Phi_0 | \hat{H}_1 | \Phi_{ij}^{ab} \rangle \langle \Phi_{ij}^{ab} | \hat{H}_1 | \Phi_0 \rangle}{E_0 - E_{ij}^{ab}}$$

two electrons excited  
from orbitals  $i$  and  $j$  to  
orbitals  $a$  and  $b$

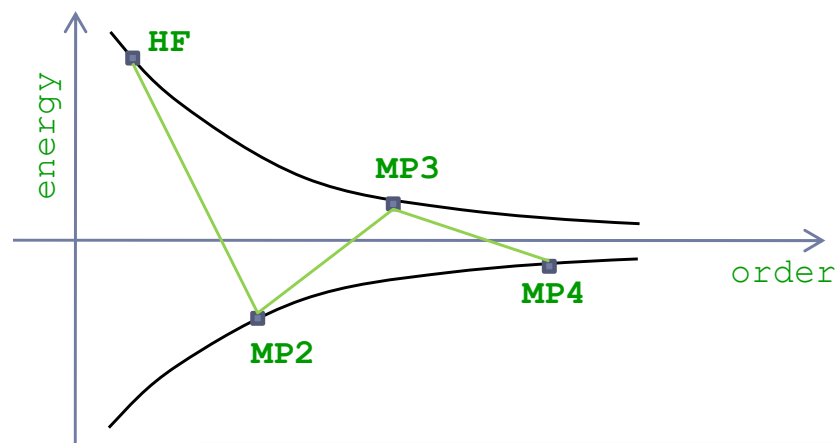
| Level | Complexity scaling with basis size   |
|-------|--------------------------------------|
| MP2   | $O(n^5)$                             |
| MP3   | $O(n^6)$                             |
| MP4   | $O(n^7)$ , but $O(n^6)$ for MP4(SDQ) |

MP2 typically accounts for 80-90% and MP3 for 95-98% of the electron correlation energy. MP2 is the cheapest way of including it to a reasonable accuracy.

# Møller-Plesset perturbation theory

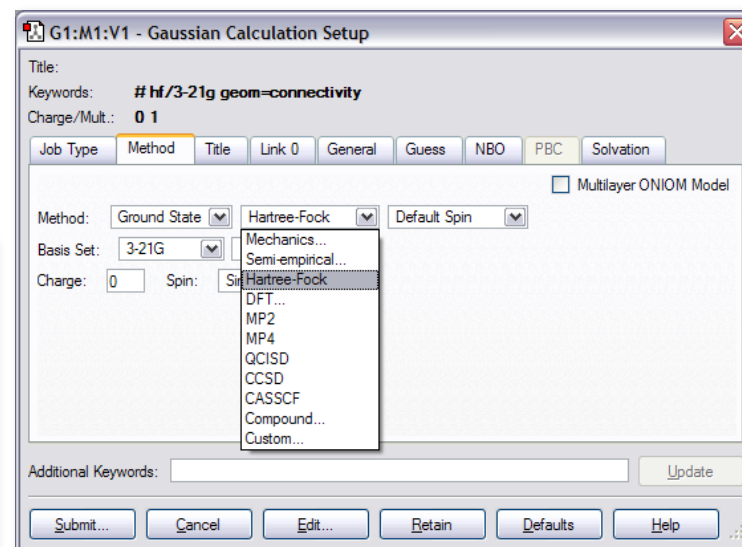
MP $n$  is not variational – that is, the energy obtained is not guaranteed to always be greater than the true energy. In fact, the convergence may be oscillatory:

| Level | Correlation energy recovery for a beryllium atom |
|-------|--------------------------------------------------|
| MP2   | 67.85%                                           |
| MP3   | 86.90%                                           |
| MP4   | 94.58%                                           |
| MP5   | 98.15%                                           |
| MP6   | 99.64%                                           |
| MP7   | 100.15% (oops...)                                |



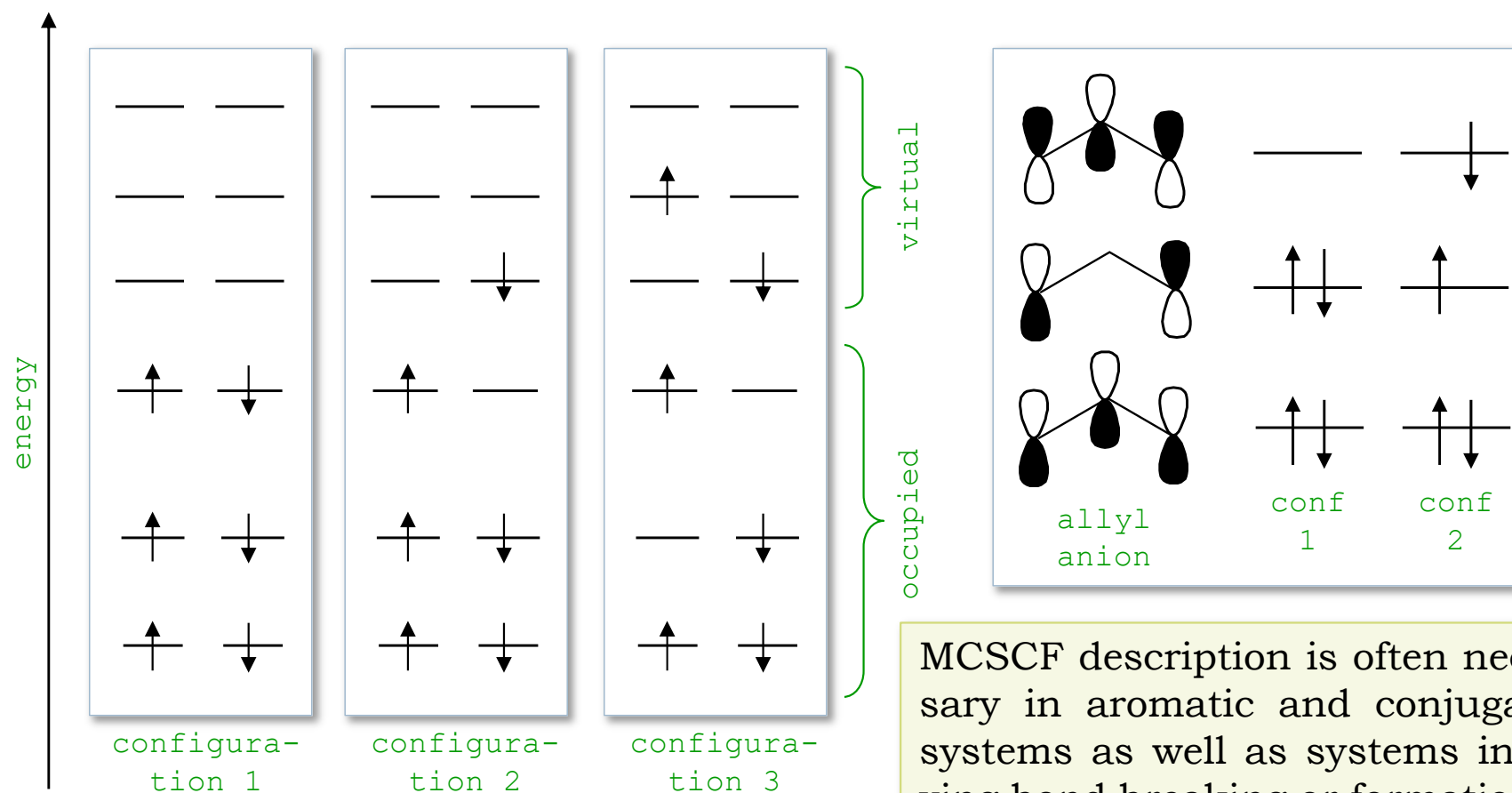
In some systems the electron correlation is not a small perturbation, and the convergence of a perturbation series is not guaranteed.

MP $n$  has convergence problems if HF is a poor starting point or if spin contamination is large.



# Multi-configuration SCF method

If a single configuration is a poor description of the system, multiple configurations (differing in the electron arrangement in the orbitals) may be used, with weights adjusted using the variational principle.



MCSCF description is often necessary in aromatic and conjugated systems as well as systems involving bond breaking or formation.

# Configuration interaction

We can take MCSCF to the limit and include all excited configurations into the calculation. This is expensive (for combinatorial reasons), and truncation strategies are therefore required.

$$\Psi^{CISD\dots} = a_0 \Psi^{HF} + \sum_i^{occ} \sum_r^{vir} a_i^r \Psi_i^r + \sum_{i<j}^{occ} \sum_{r<s}^{vir} a_{ij}^{rs} \Psi_{ij}^{rs} + \dots$$

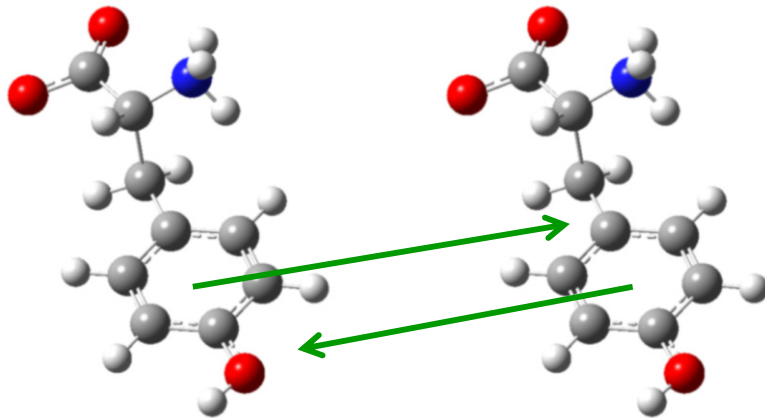
| Method  | Configurations                                     | Scaling with basis size |
|---------|----------------------------------------------------|-------------------------|
| CID     | all doubly excited determinants                    | $O(n^5)$                |
| CISD    | all singly and doubly excited determinants         | $O(n^6)$                |
| CISDT   | all singly, doubly and triply excited determinants | $O(n^8)$                |
| full CI | all configurations                                 | $O(n!)$                 |

When full CI is used in some physically motivated subset of orbitals, the method is called CASSCF (complete active space SCF). Choosing CASSCF orbitals is a dark art.

Due to the astronomical number of configurations, full CI is only possible for very small (diatomics) systems. Fortunately, full CI is rarely necessary, as CISD tends to be sufficiently accurate for most practical purposes.

# Size consistency and basis set superposition error

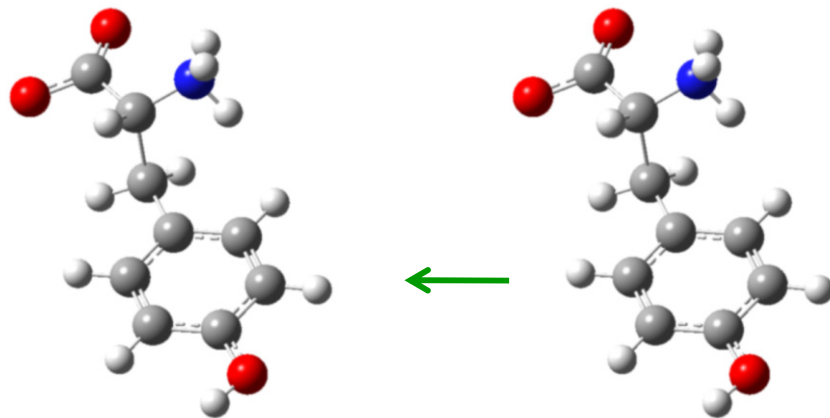
Size consistency problem in configuration interaction calculations:



When molecules are simulated together, new excitations become available to each of them, resulting in a greater variational flexibility in the CI subspace.

$$E^{CI}(1+2) < E^{CI}(1) + E^{CI}(2)$$

Basis set superposition error in atom-centred basis sets:



When molecules are brought close together, they augment each other's basis, thereby increasing the variational flexibility. The energy will go down even in the absence of any interaction. This results in non-physical attraction.

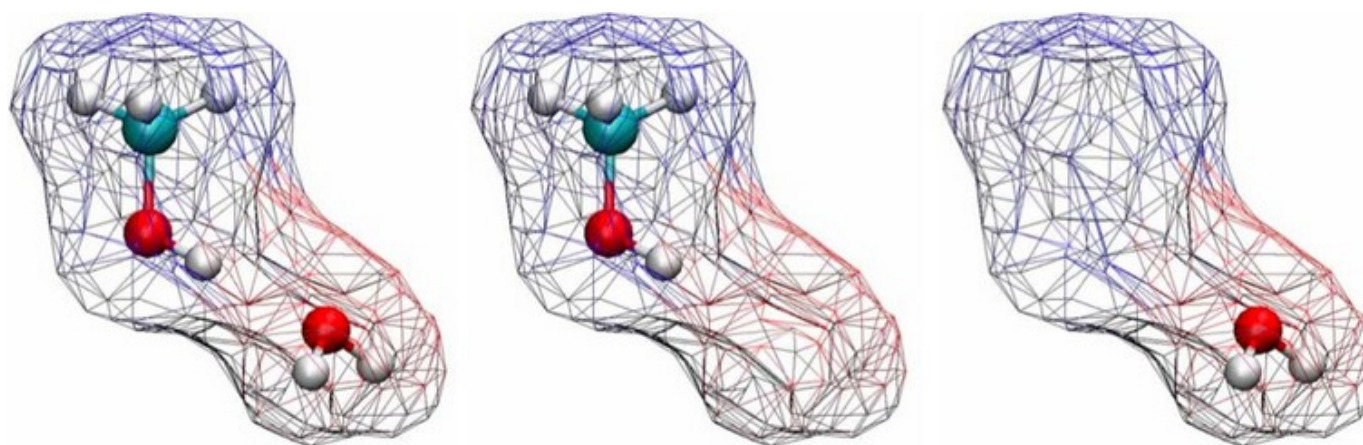
$$E(\textit{proximate}) < E(\textit{distant})$$

► Plane-wave basis sets do not suffer from BSSE because they are uniform.

# Counterpoise correction to BSSE

Atom-centred basis sets are rarely big enough to avoid the basis set superposition error. Counterpoise correction aims to compensate this error.

1. *The energy of the complex is calculated in the usual way.*
2. *Two additional calculations are carried out for each fragment in the presence of the orbitals (but not electrons or nuclei) of the other fragment.*



$$\Delta E_{\text{CP}} = E(A)_{\text{AB}} + E(B)_{\text{AB}} - E(A)_{\text{A}} - E(B)_{\text{B}}$$

| Basis Set | $H_0$ (no CP correction) (in hartree) | Counterpoise corrected energy (in hartree) | Counterpoise BSSE Energy (in kcal/mol) |
|-----------|---------------------------------------|--------------------------------------------|----------------------------------------|
| 3-21G     | -151.1894                             | -151.13462                                 | 4.066                                  |
| 6-31G     | -151.9342                             | -151.93274                                 | 0.916                                  |
| 6-31+G(d) | -151.99457                            | -151.99364                                 | 0.584                                  |