

The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble.

P. A. M. Dirac, 1929.

Practical Aspects of Quantum Chemistry

A brief overview of tools and methods used in practical applications of quantum chemistry

Tools of the trade

Expert-level quantum chemical codes:

Name	Capabilities	Ease of use	Environment	Price
GAMESS*	Mainstream + cutting edge	Reasonable	Unix/Linux	GPL
Gaussian*	Mainstream	Easy	Unix/Linux	£3000
DALTON	Mainstream + cutting edge	Unforgiving	Unix/Linux	GPL

Mid-level quantum chemistry programs (many more exist):

Name	Capabilities	Ease of use	Environment	Price
HyperChem*	Limited mainstream	Easy	Windows	£500
Chem3D	Limited mainstream	Easy	Windows	£1500+

Visualization and post-processing:

Name	Capabilities	Ease of use	Environment	Price
GaussView*	Limited mainstream	Easy	All	£300
MOLDEN	Limited mainstream	Easy	Unix/Linux	GPL
MATLAB*	Sky's the limit	Reasonable	All	£10000+

Tools of the trade: where to look

GAMESS-US

<http://www.msg.ameslab.gov/GAMESS/>

MOLDEN

<http://www.cmbi.ru.nl/molden/molden.html>

DALTON

<http://www.kjemi.uio.no/software/dalton/dalton.html>



Gaussian, ChemOffice, Matlab, Mathematica etc

Site-licensed by the University: check with your supervisor, IT support officer or directly with OUCS and Oxford Supercomputing Centre.

Two new supercomputers are coming online in April

168-core SSI Itanium 2 system

512-core cluster with HPC interconnect

Roothaan's formulation of SCF

Given a complete *basis set*, the integrodifferential Fock equations may be recast in a more tractable matrix form:

$$\hat{f}(\mathbf{r})|\psi_i(\mathbf{r})\rangle = \varepsilon_i|\psi_i(\mathbf{r})\rangle$$

Basis set
↙

$$|\psi_i\rangle = \sum_k C_{ki}|\phi_k\rangle$$

↓

$$\hat{f} \sum_k C_{ki}|\phi_k\rangle = \varepsilon_i \sum_k C_{ki}|\phi_k\rangle$$

↙

$$\langle\phi_j|\times\downarrow$$
$$\sum_k \langle\phi_j|\hat{f}|\phi_k\rangle C_{ki} = \varepsilon_i \sum_k \langle\phi_j|\phi_k\rangle C_{ki} \longrightarrow$$

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon}$$
$$F_{jk} = \langle\phi_j|\hat{f}|\phi_k\rangle$$
$$S_{jk} = \langle\phi_j|\phi_k\rangle$$
$$\boldsymbol{\varepsilon} = \text{diag}(\{\varepsilon_i\})$$

For full derivation and discussion see

Szabo and Ostlund, *Modern Quantum Chemistry*, 1996 (Dover), Chapter 3.

Roothaan's formulation of SCF

The need to iterate is still there, but the complexity is now confined to the evaluation of Fock and basis overlap matrices:

$$\langle \phi_j | \hat{f} | \phi_k \rangle = \langle \phi_j | \hat{h} | \phi_k \rangle + \sum_m \langle jm || km \rangle$$

Expensive 6+2
dimensional
integrals



Along the MO LCAO lines, one would think of using Laguerre polynomials and spherical harmonics as basis:

$$G_{n,l,m}(r, \theta, \varphi) = L_n(r) Y_{l,m}(\theta, \varphi) e^{-r/a_n}$$

Or perhaps Slater functions and spherical harmonics:

$$G_{n,l,m}(r, \theta, \varphi) = S_n(r) Y_{l,m}(\theta, \varphi)$$

The numerical integration problem

With numerical quadratures, all but the smallest HF solutions are impractical on even the very latest computer hardware.

Gaussian basis sets

The anatomy of a Gaussian basis set:

$$G_{n,l,m}(r, \theta, \varphi) = \underbrace{Y_{l,m}(\theta, \varphi)}_{\text{spherical harmonic}} \underbrace{e^{-r^2/b_n^2}}_{\substack{\text{Gaussian} \\ \text{exponent} \\ \text{(each n may have several)}}$$

Spherical harmonics are a complete basis for the angular part of the wavefunction in a spherically symmetric system.

Contracted Gaussian basis sets (a time-saver):

$$G_{n,l,m}(r, \theta, \varphi) = Y_{l,m}(\theta, \varphi) \sum_i a_{ni} e^{-r^2/b_{ni}^2}$$

The impact of contraction on the variational flexibility is usually minimal, so the most commonly used basis sets are contracted.

Augmented and multi- ζ Gaussian basis sets (greater variational flexibility):

$$G_{n,l,m}^{(*)} = Y_{l+k,m}(\theta, \varphi) e^{-r^2/b_n^2}; \quad G_{n,l,m}^{(+)} = Y_{l,m}(\theta, \varphi) e^{-r^2/b_{n+k}^2}$$

Gaussian basis sets

Advantages of Gaussian basis sets:

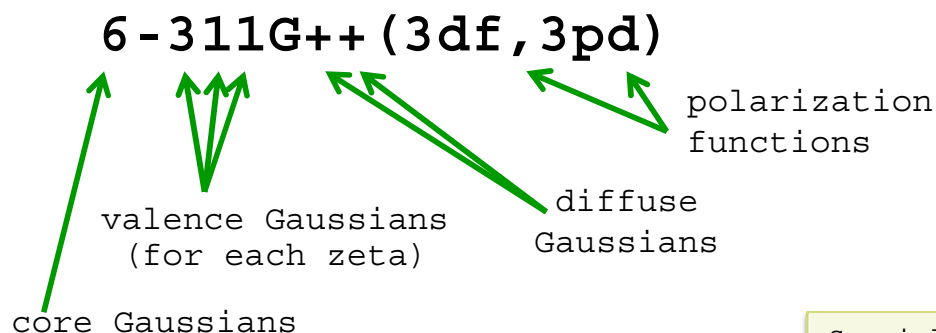
Property	Benefits
A product of two Gaussians is another Gaussian at their centre of gravity. Spherical harmonics are a semigroup under multiplication.	All two-electron integrals are analytical. Massive time saving compared to numerical quadratures.
Gaussians fall off very steeply with distance.	Wide use of distance-based cut-offs possible in two-electron integral evaluation routines, further improving performance.

Important drawbacks:

Property	Consequences
Gaussians fall off very steeply with distance.	Incorrect asymptotic behaviour at large distances. Special measures required to describe anions, Rydberg states and long-range orbital overlap.
Gaussians are bounded and have zero derivative at the origin.	Incorrect asymptotic behaviour at small distances. Special measures required to adequately describe wavefunction cusp at the nucleus point.
Gaussians forming the basis are centred at the nuclei.	Non-uniform variational flexibility in the basis. Basis set superposition error.

Gaussian basis sets

People's basis sets:



Correlation-consistent basis sets:

aug-cc-pVTZ

"correlation consistent polarized valence triple-zeta basis set augmented with diffuse functions"

Specialized basis sets exist for almost any particular purpose. Make sure you search the literature for any case-specific bases before using a general-purpose one.

What to look for:

Zeta number	Polarization functions	Diffuse functions	Tight functions	Effective core potentials	Rydberg functions
The more, the better/slower, usually 2-3.	The more, the better/slower, Usually one or two $l+1$ sets.	The more, the better/slower, usually a single set for anions.	Only used for computing HFCs and J -couplings.	Used to approximate inner shells of heavy elements	Only used to describe Rydberg states

Where to find Gaussian basis sets

<https://bse.pnl.gov/bse/portal>

(Gaussian basis set exchange portal)

Cookbook: ground state energies

Fast approximate methods (but capturing the essentials):

HF 6-31G(d,p) or HF cc-pVDZ

$$O(n^{2-3})$$

Slow accurate methods:

MP2 6-311G(2d,2p) or MP2 cc-pVTZ

$$O(n^{4-5})$$

Very slow, very accurate methods:

CISD(T) aug-cc-pVQZ and higher

$$O(n^{7+})$$

Specialized high-accuracy compound methods:

MP2-R12, G1-G3, G2MP2, G3MP2 etc

$$O(n^{5-6})$$

Dunning-Feller extrapolation to CBS limit:

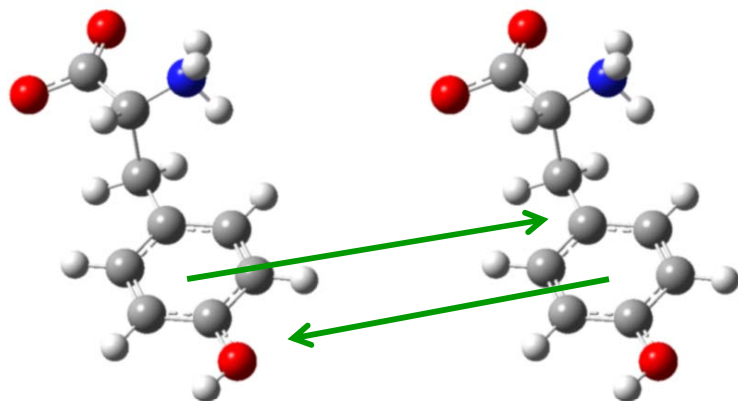
$$E_n = E_{CBS} + \alpha e^{-\beta n} \quad (\text{for } E_n \text{ computed with (aug-)cc-pVnZ basis sets})$$

For other extrapolation schemes see:

de Lara-Castells *et al*, Journal of Chemical Physics 2001, 115, 22, 10438.

Size-consistency and BSSE

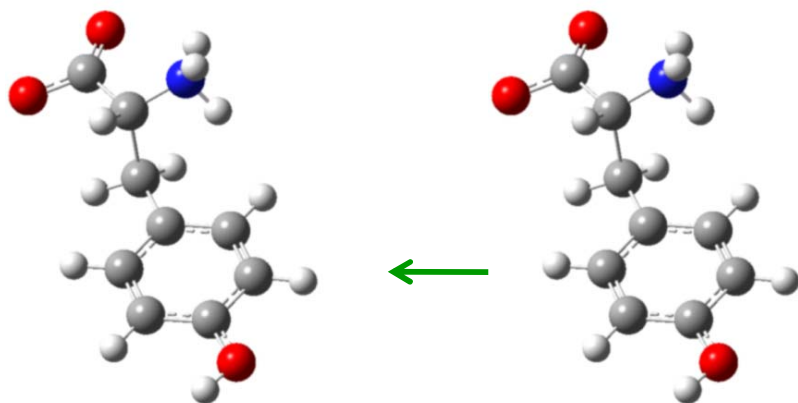
Size-consistency problem in configuration interaction:



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:



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.

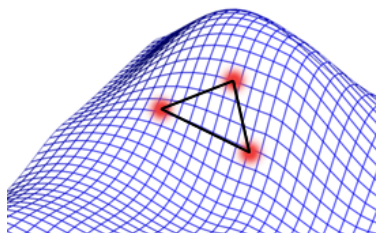
$$E(\textit{proximate}) < E(\textit{distant})$$

Molecular geometry

Energy-only algorithms

Simplex

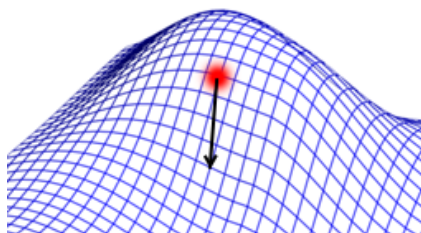
Downward move by polygon node reflection.



The problem consists in bringing the system to a stationary state, which is a state with zero energy gradient with respect to the coordinates of the nuclei. It might not correspond to a minimum or a maximum.

Gradient descent/ascent algorithms

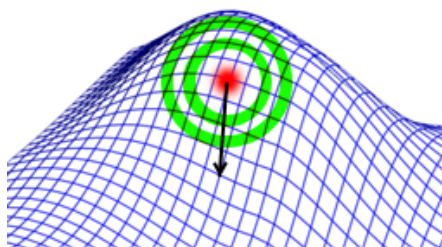
Line search, backtracking, gradient descent.



Gradient evaluation is required at every step (which is about as expensive as energy itself). Much faster than simplex, but less robust.

Second and higher order algorithms

Newton-Raphson, Broyden, RFO, GDIIS, Levenberg-Marquardt.



Hessian and/or higher derivatives are required at every step (very expensive). Very fast convergence assuming the initial guess is good.

Molecular geometry

Hessians and stationary point classification

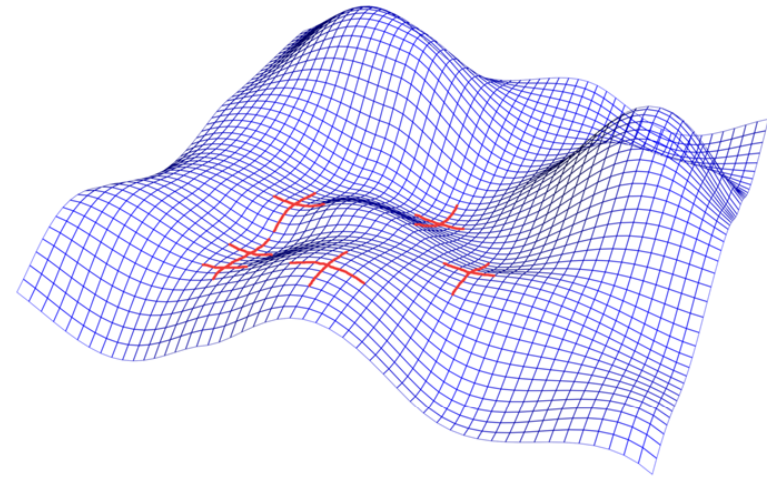
$$g_i = \frac{\partial E}{\partial x_i}; \quad H_{ij} = \frac{\partial^2 E}{\partial x_i \partial x_j}$$

If the gradient is zero and

All \mathbf{H} eigenvalues positive \Rightarrow **Minimum**

All \mathbf{H} eigenvalues negative \Rightarrow **Maximum**

\mathbf{H} eigenvalues of both signs \Rightarrow **Saddle**



BFGS (Broyden-Fletcher-Goldfarb-Shanno) formula

$$\mathbf{H}_{k+1} = \left[\mathbf{H} + \frac{\mathbf{y}\mathbf{y}^T}{\mathbf{s}^T\mathbf{y}} - \frac{\mathbf{H}\mathbf{s}\mathbf{s}^T\mathbf{H}}{\mathbf{s}^T\mathbf{H}\mathbf{s}} \right]_k$$

$$\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$$

$$\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$$

The exact Hessian is very expensive to compute. In many cases however, it can be estimated with reasonable accuracy from gradients computed at the previous steps using BFGS formula or one of its many descendants.

Properties

Densities and orbital plots

Commonly encountered ground state densities of interest are:

Orbital density

Not of much use, apart from pretty pictures and common sense checks. Might be used for Hückel-style intuitive predictions.

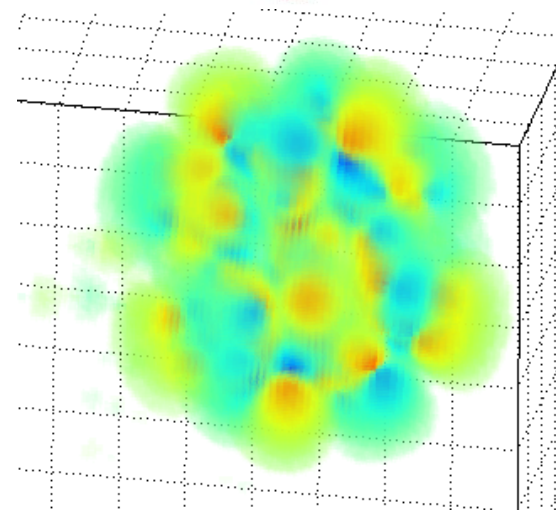
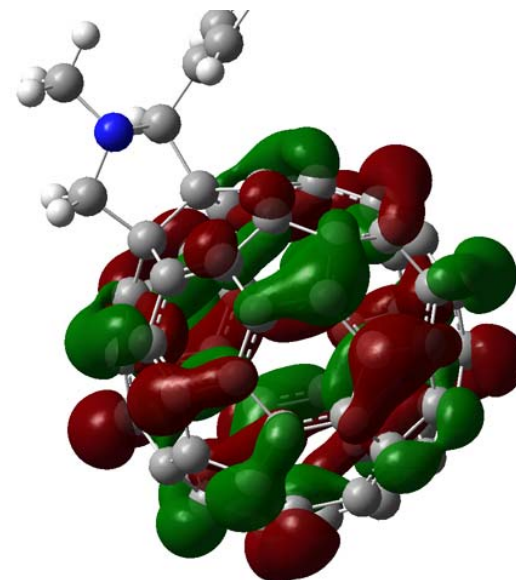
Charge density

Used for modelling solvent effects, for parameterizing MD force fields and computing multipole moments.

Spin density

Used for computing magnetic dipolar couplings and hyperfine coupling tensors.

Densities are commonly visualized using either isosurface plots or transparency density plot.



Properties

Vibrational spectra

In a harmonic approximation: $E = \frac{kx^2}{2} \Rightarrow k = \frac{\partial^2 E}{\partial x^2}$

Therefore Hessian eigenvalues are required. Exact Hessian is necessary, so the calculation will be rather expensive.

In practice, due to non-Born-Oppenheimer effects and anharmonicities, the vibrational frequencies computed from Hessian eigenvalues differ from experimental values by a scaling factor (of about 0.9). The scaling factors have been tabulated in the literature:

Scott & Radom, *Journal of Physical Chemistry* **1996** 100 (41),16502

Anharmonicities may be obtained by numerical differentiation along the normal modes.

Vibrational corrections to properties may be obtained by thermal averaging over the available vibrations. This is currently very expensive and only done in special cases.

Other properties

One hour is not enough for the whole Practical QC course!

Excited state energies and geometries:

CI approach
TD-DFT approach
ZINDO methods

Solvents and solvation:

Explicit solvent
Onsager solvent model
PCM, IPCM and SCI-PCM methods

Magnetic resonance:

Chemical shielding tensors
Hyperfine coupling tensors
g-tensors
Scalar coupling constants

Reaction dynamics:

CPMD and BOMD
Potential energy surfaces
Intrinsic reaction coordinates

Semi-empirical Quantum Chemistry:

INDO, MINDO etc
AMI, PM3, PM3MM
ZINDO and ZINDO/S

MCSCF and VBT:

Restricted state spaces
MCQDPT and CASPT2

Orbital and population analysis:

Orbital localization
Mulliken and Lowdin populations

Boundary conditions:

Periodic boundary conditions
Plane-wave basis sets
ONIOM and related methods

The best starting point is Gaussian
and GAMESS documentation.

Miscellaneous

Exploit symmetry whenever possible

Split-up into irreducible representations may speed up calculations by orders of magnitude. However, in some cases non-symmetric energy minima and excitations may be missed if the symmetry is followed strictly.

Check your Hessians

Some algorithms, particularly the vanilla Newton-Raphson, will converge to the *nearest stationary point* rather than a minimum. For production calculations always double-check that hessian eigenvalues are indeed all positive.

Use Density Functional Theory methods

Density Functional Theory methods are frequently faster and more accurate for ground state properties. Unless your calculation critically depends on the excited states, consider using DFT.

A little bit of brain time saves a lot of CPU time

Subscribe to technical support mailing lists, take a look through the source code, try different compiler options, read up on what exactly goes on inside a program. Clever choice of options frequently results in speed-ups by orders of magnitude.