Advanced Physical Chemistry

Lecture notes for the computational chemistry lab (CHM 5111)

Roberto Peverati · Florida Institute of Technology

Contents

Preface 2
1 The molecular Schrödinger equation 3
2 Approximating the wave function: basis sets 16
3 From theory to a calculation: IQmol, Q-Chem and Experiment 1 28
4 Approximating the Hamiltonian I: Hartree–Fock and the SCF 38
5 Approximating the Hamiltonian II: density functional theory 53
6 Approximating the Hamiltonian III: Jacob's ladder 65
7 How we test a functional 79
8 Dispersion and noncovalent interactions 91
9 Why some reactions are hard for DFT 102
10 Basis-set error vs functional error 112
11 Numerical integration grids 121
12 SCF stability, open shells and transition metals 134
13 From electronic energies to experiment 143
14 Choosing, running and reporting a method 154
15 What we learned 162
Bibliography 166

Preface

These notes collect the fifteen lecture handouts of the computational lab of Advanced Physical Chemistry at the Florida Institute of Technology, one chapter per lecture. Chapters 1–3 approximate the **wave function** (basis sets) and come before Experiment 1. Chapters 4–6 approximate the **Hamiltonian** (Hartree–Fock and density functional theory) and come before Experiment 2. From Chapter 7 on, each chapter first looks back at the experiment you just finished and then prepares the next one; read those debrief sections only **after** you submit the experiment.

The experiments, the guides and the interactive tools are on the course website, https://peverati.github.io/advpchem/. For the quantum mechanics behind Chapters 1–5, see also the Live Textbook of Physical Chemistry 2.

1 The molecular Schrödinger equation

Learning goals

After studying this chapter you should be able to:

- say what a computational chemist actually predicts, and how the ten experiments of this lab fit together;
- write the molecular Hamiltonian term by term, in SI and in atomic units, and convert between hartree, kcal/mol, kJ/mol and eV;
- explain the Born–Oppenheimer approximation and what the "total energy" printed by Q-Chem contains (and what it does not);
- explain why the Schrödinger equation cannot be solved exactly for any system with two or more electrons;
- build a Slater determinant and explain why it is antisymmetric;
- state the variational principle and use it to judge which of two energies is "better";
- explain what "approximating Ψ" and "approximating $\hat{H}$" mean, and decode a model chemistry such as B3LYP/def2-TZVP.

Review from P-Chem 2

This chapter builds on the Live Textbook of Physical Chemistry 2. If any of the following feels rusty, reread it first:

- The time-independent Schrödinger equation (chapter 3)
- Operators, eigenfunctions and eigenvalues and the Hamiltonian operator (chapter 6)
- The hydrogen atom: radial wave functions (chapter 5)
- Spin and Postulate 6: the Pauli exclusion principle (chapters 7–8)
- Many-electron wave functions and the variational method (chapter 10)
- The molecular Hamiltonian and the Born–Oppenheimer approximation (chapter 11)

1.1 Why compute?

Quantum chemistry programs such as Q-Chem [1] solve, approximately, one equation: the Schrödinger equation for the electrons of a molecule. From its solution we get an **energy** for a given arrangement of atoms. Almost everything a computational chemist predicts is built from such energies:

- **Reaction energies.** Compute the energy of every reactant and product, then combine them with the stoichiometric coefficients. Is the reaction uphill or downhill, and by how much?
- **Barrier heights.** Compute the energy of a transition state and compare it with the reactants. The barrier controls the rate.
- **Structures.** Move the atoms until the energy is as low as possible. The result is a predicted equilibrium geometry, with bond lengths and angles.
- **Spectroscopic and thermodynamic properties.** The curvature of the energy around a minimum gives vibrational frequencies; from those, statistical mechanics gives zero-point energies, enthalpies and free energies.
- **Noncovalent interactions.** How strongly do two molecules stick together? This needs very accurate energy *differences*, often of only one or two kcal/mol.

The energies involved are tiny compared with the total energies the program prints. A water molecule has a total energy of about −76 hartree, which is about −48,000 kcal/mol. A typical reaction energy is a few to a few tens of kcal/mol, and a hydrogen bond is a few kcal/mol. In this lab you will constantly subtract large numbers to get small ones. That is why the accuracy of each calculation matters so much, and why you must keep all the digits.

1.1.1 The course roadmap

This lab has ten experiments. Each one isolates one source of error or one practical decision:

Exp.	Topic	What it teaches
1	Basis-set incompleteness	how big a basis set must be
2	Jacob's ladder	how density functionals are classified
3, 4	Organic chemistry, parts A and B	how functionals perform on "easy" and dispersion-bound reactions
5	Difficult cases	where approximate functionals break down
6	B3LYP/6-31G*	error cancellation between basis and method
7	Integration grids	numerical noise in DFT
8	Stability analysis	whether the SCF found the right solution
9	Theory vs experiment	from electronic energies to measured quantities
10	Final review	choosing and defending a method yourself

Behind all ten sits a single idea, which this chapter develops: **every practical calculation makes two approximations.** We approximate the *wave function* Ψ by expanding it in a finite set of functions (the basis set, Lecture 2 and Experiment 1). And we approximate the *Hamiltonian* $\hat{\mathrm{H}}$, or more precisely the way electrons interact, by choosing a method such as Hartree–Fock or a density functional (Lectures 4–6 and Experiments 2–5). Most of the semester is about understanding the errors these two approximations introduce, and how they can add up or cancel.

1.2 Review: the Schrödinger equation and the hydrogen atom

1.2.1 Operators, eigenfunctions, eigenvalues

The time-independent Schrödinger equation (TISE) [2] is an eigenvalue equation:

$$\hat{H}\Psi = E\Psi. \tag{1}$$

Here $\hat{H}$ is the **Hamiltonian operator**, the operator associated with the total energy. An operator acts on a function and returns another function. When the result is the *same* function multiplied by a number, the function is an **eigenfunction** of the operator and the number is the **eigenvalue**. Equation 1 says that

the allowed stationary states of a system are the eigenfunctions Ψ of $\hat{H}$, and the allowed energies are its eigenvalues E.

Three facts from P-Chem 2 matter for everything that follows:

1. The Hamiltonian is a **Hermitian** operator, so its eigenvalues are real and its eigenfunctions can be chosen orthonormal: $\int \Psi_i^* \Psi_j \, d\tau = \delta_{ij}$.
2. The Hamiltonian is a sum of a **kinetic** part and a **potential** part, $\hat{H} = \hat{T} + \hat{V}$. For one particle of mass m, $\hat{T} = -\frac{\hbar^2}{2m}\nabla^2$, and $\hat{V}$ simply multiplies the wave function by the potential energy $V(\mathbf{r})$.
3. For any normalized function Ψ, the **expectation value** of the energy is $\langle E \rangle = \int \Psi^* \hat{H} \Psi \, d\tau$. If Ψ is an eigenfunction, $\langle E \rangle$ equals its eigenvalue exactly.

> **i** Notation in this course
>
> The P-Chem 2 textbook writes kinetic-energy operators as $\hat{K}$ and uses χ for spin orbitals. In these handouts we follow the convention of most quantum-chemistry programs and papers: $\hat{T}$ for kinetic energy, Ψ for the many-electron wave function, ϕ_i (or ψ_i) for orbitals, and χ_μ for **basis functions** (from Lecture 2 on). Bold symbols such as $\mathbf{r}$ are vectors.

1.2.2 The hydrogen atom: the one exactly solvable atom

The hydrogen atom has one electron moving in the Coulomb field of one proton. Its Schrödinger equation can be solved exactly. In atomic units (Section 1.3.2) the energies are

$$E_n = -\frac{1}{2n^2}\,E_\mathrm{h}, \qquad n = 1, 2, 3, \ldots \tag{2}$$

so the ground state lies at exactly $-0.5\,E_\mathrm{h}$. The eigenfunctions factor into a radial part and an angular part,

$$\psi_{n\ell m_\ell}(r, \theta, \varphi) = R_{n\ell}(r)\, Y_\ell^{m_\ell}(\theta, \varphi),$$

labeled by three quantum numbers: the principal quantum number $n = 1, 2, \ldots$; the angular-momentum quantum number $\ell = 0, \ldots, n-1$ (s, p, d, f, …); and the magnetic quantum number $m_\ell = -\ell, \ldots, +\ell$. The radial function $R_{n\ell}$ has $n - \ell - 1$ **radial nodes**, and the angular part has ℓ angular nodes. Figure 1 shows the first six radial functions and the corresponding radial probability densities.

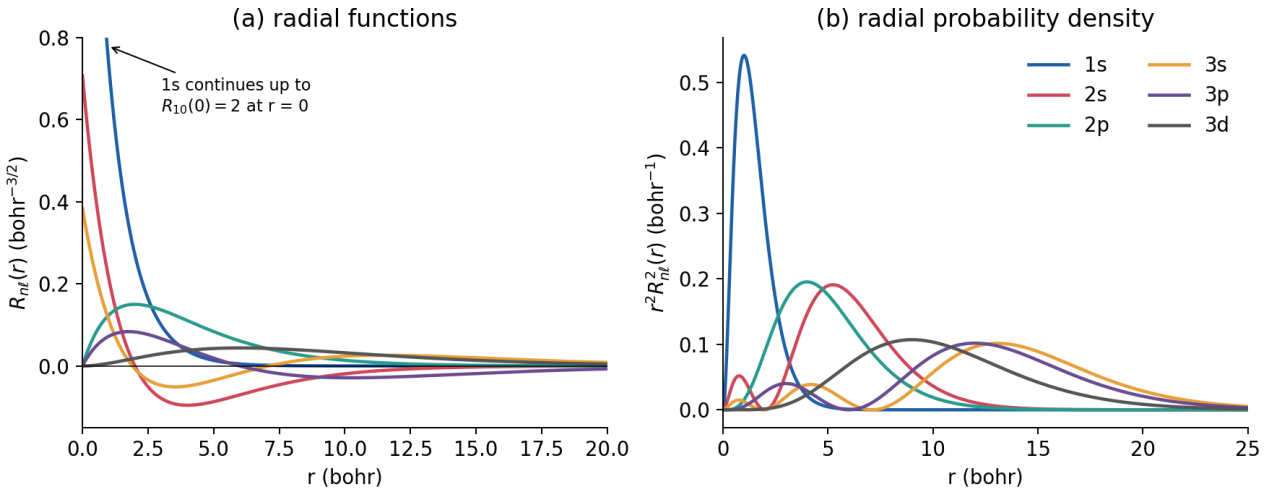


Figure 1: Hydrogen-atom radial functions $R_{n\ell}(r)$ (a) and radial probability densities $r^2 R_{n\ell}^2(r)$ (b), in atomic units. The 2s function has one radial node, 3s has two, 3p has one; 2p and 3d have none ($n - \ell - 1 = 0$). Note two features that will matter in Lecture 2: s functions have a sharp, finite **cusp** at the nucleus, and every function decays **exponentially**, as $e^{-r/n}$, at large r.

The hydrogen orbitals are **one-electron wave functions**. That idea is the seed of everything we do: for many-electron systems we will build approximate wave functions out of one-electron functions (orbitals), and we will build the orbitals themselves out of simpler functions that look like the hydrogen solutions near each nucleus. Two properties of $\psi_{1s} \propto e^{-r}$ are worth noticing now, because Gaussian basis functions will get both of them wrong (Lecture 2):

- **The cusp.** At $r = 0$ the 1s function has a finite value and a sharp corner: its slope is $d\psi/dr = -Z\psi(0)$, not zero. This is not an accident of hydrogen; it is required of any exact wave function wherever an electron meets a nucleus [3].
- **The exponential tail.** Far from the nucleus, bound wave functions decay as $e^{-\sqrt{2I}\,r}$, where I is the ionization energy in hartree. For hydrogen, $I = 0.5\ E_\mathrm{h}$ and the decay is e^{-r}.

1.3 The molecular Hamiltonian

1.3.1 Five terms

A molecule is a collection of M nuclei (charges Z_A, masses M_A, positions $\mathbf{R}_A$) and N electrons (mass m_e, charge $-e$, positions $\mathbf{r}_i$). Its nonrelativistic Hamiltonian has five terms:

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn}. \tag{3}$$

In SI units, each term reads

$$
\begin{aligned}
\hat{T}_n &= -\sum_{A=1}^{M} \frac{\hbar^2}{2M_A}\nabla_A^2 && \text{kinetic energy of the nuclei} \\
\hat{T}_e &= -\sum_{i=1}^{N} \frac{\hbar^2}{2m_e}\nabla_i^2 && \text{kinetic energy of the electrons} \\
\hat{V}_{ne} &= -\sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A e^2}{4\pi\varepsilon_0\,|\mathbf{r}_i - \mathbf{R}_A|} && \text{nucleus–electron attraction} \\
\hat{V}_{ee} &= \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{e^2}{4\pi\varepsilon_0\,|\mathbf{r}_i - \mathbf{r}_j|} && \text{electron–electron repulsion} \\
\hat{V}_{nn} &= \sum_{A=1}^{M}\sum_{B>A}^{M} \frac{Z_A Z_B e^2}{4\pi\varepsilon_0\,|\mathbf{R}_A - \mathbf{R}_B|} && \text{nucleus–nucleus repulsion}
\end{aligned}
\tag{4}
$$

The two attraction/repulsion sums over pairs use $j > i$ and $B > A$ so that each pair is counted once. The P-Chem 2 textbook writes the same five terms in §11.1, with $\hat{K}$ for kinetic energy.

This Hamiltonian is called "exact", but it already leaves out some physics: relativistic effects (important for core electrons of heavy atoms) and spin–orbit coupling. For the light molecules in most of this lab they are small. For heavy elements (Experiment 8 has ruthenium; Experiment 10 has palladium) the def2 basis sets include relativistic effects through *effective core potentials*, which you will meet in Lecture 2.

1.3.2 Atomic units

All the constants in Equation 4 are a nuisance. Quantum chemistry therefore works in **atomic units**, in which

$$\hbar = m_e = e = 4\pi\varepsilon_0 = 1.$$

With this choice the Hamiltonian becomes much cleaner:

$$\hat{H} = -\sum_{A}\frac{1}{2M_A}\nabla_A^2 - \sum_{i}\frac{1}{2}\nabla_i^2 - \sum_{i}\sum_{A}\frac{Z_A}{r_{iA}} + \sum_{i}\sum_{j>i}\frac{1}{r_{ij}} + \sum_{A}\sum_{B>A}\frac{Z_A Z_B}{R_{AB}}, \tag{5}$$

where $r_{iA} = |\mathbf{r}_i - \mathbf{R}_A|$, $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$, $R_{AB} = |\mathbf{R}_A - \mathbf{R}_B|$, and the nuclear masses M_A are now in units of the electron mass (about 1836 for a proton).

The unit of length is the **bohr** (a_0), the most probable distance of the 1s electron from the proton in hydrogen. The unit of energy is the **hartree** (E_h), twice the ionization energy of hydrogen. These are the units of every number Q-Chem prints internally, so you need the conversions at your fingertips:

Table 1: Atomic units used in this course. In this lab, total energies are always reported in hartree and reaction energies in kcal/mol, with 1 E_h = 627.5095 kcal/mol.

Quantity	Atomic unit	Conversion
length	1 bohr (a_0)	0.529177 Å
energy	1 hartree (E_h)	627.5095 kcal/mol
		2625.50 kJ/mol
		27.2114 eV
		219,474.6 cm^{-1}
mass	m_e	1/1822.89 of an atomic mass unit

Worked example: units

(a) The ground-state energy of hydrogen is $-0.5\ E_\mathrm{h}$. In other units: $-0.5 \times 627.5095 = -313.75$ kcal/mol, $-0.5 \times 27.2114 = -13.606$ eV, which is the familiar 13.6 eV ionization energy.

(b) Experiment 1 studies the dissociation of the HCl dimer, with a reference energy of 2.01 kcal/mol. In hartree, that is $2.01/627.5095 = 0.003203\ E_\mathrm{h}$. The total energy of one HCl molecule is about $-460\ E_\mathrm{h}$. The quantity you want is therefore about one part in 10^5 of the numbers you subtract. Rounding each total energy to 4 decimals introduces up to $0.00005\ E_\mathrm{h}$ per energy. In $\Delta E = 2E(\mathrm{HCl}) - E(\mathrm{dimer})$ the worst case is $3 \times 0.00005 = 0.00015\ E_\mathrm{h}$, or 0.09 kcal/mol: almost 5% of the answer.

(c) The bond length of HCl in the Experiment 1 geometry file is 1.27907 Å. In bohr: $1.27907/0.529177 = 2.41709\ a_0$.

1.4 The Born–Oppenheimer approximation

1.4.1 Clamped nuclei

Equation 5 still describes all particles at once. Even for H_2 (two nuclei, two electrons) it is a problem in 12 coordinates. The first and most important simplification uses the fact that nuclei are heavy: a proton is 1836 times heavier than an electron, and a ^{35}Cl nucleus about 64,000 times. For the same kinetic energy, nuclei move far more slowly. To the electrons, the nuclei look almost frozen; to the nuclei, the electrons adjust instantly to wherever the nuclei are.

The **Born–Oppenheimer approximation** [4] makes this precise. We fix ("clamp") the nuclei at positions $\mathbf{R}$, drop the nuclear kinetic energy $\hat{T}_n$, and solve for the electrons alone. The **electronic Hamiltonian** is

$$\hat{H}_\mathrm{el} = -\sum_i \frac{1}{2}\nabla_i^2 - \sum_i \sum_A \frac{Z_A}{r_{iA}} + \sum_i \sum_{j>i} \frac{1}{r_{ij}}, \tag{6}$$

and the **electronic Schrödinger equation** is

$$\hat{H}_\mathrm{el}\,\Psi_\mathrm{el}(\mathbf{r};\mathbf{R}) = E_\mathrm{el}(\mathbf{R})\,\Psi_\mathrm{el}(\mathbf{r};\mathbf{R}). \tag{7}$$

The semicolon means that the electronic wave function depends on the electron coordinates $\mathbf{r}$ as variables and on the nuclear positions $\mathbf{R}$ only as **parameters**: change the geometry, and you get a different equation to solve. With the nuclei fixed, V_{nn} is just a number, and we add it at the end.

1.4.2 The potential energy surface

The sum

$$U(\mathbf{R}) = E_{\mathrm{el}}(\mathbf{R}) + V_{nn}(\mathbf{R}) \tag{8}$$

is the **potential energy surface** (PES): the energy the nuclei feel as they move. Its minima are stable structures, its first-order saddle points are transition states, and its curvature at a minimum gives the vibrational frequencies. Every concept of a "molecular structure" lives on this surface, and it exists *only because of* the Born–Oppenheimer approximation.

Figure 2 shows the PES of H_2 (a single coordinate, the bond length), split into its two pieces. The electronic energy alone is attractive at all distances; it is the nuclear repulsion that turns it into a curve with a minimum.

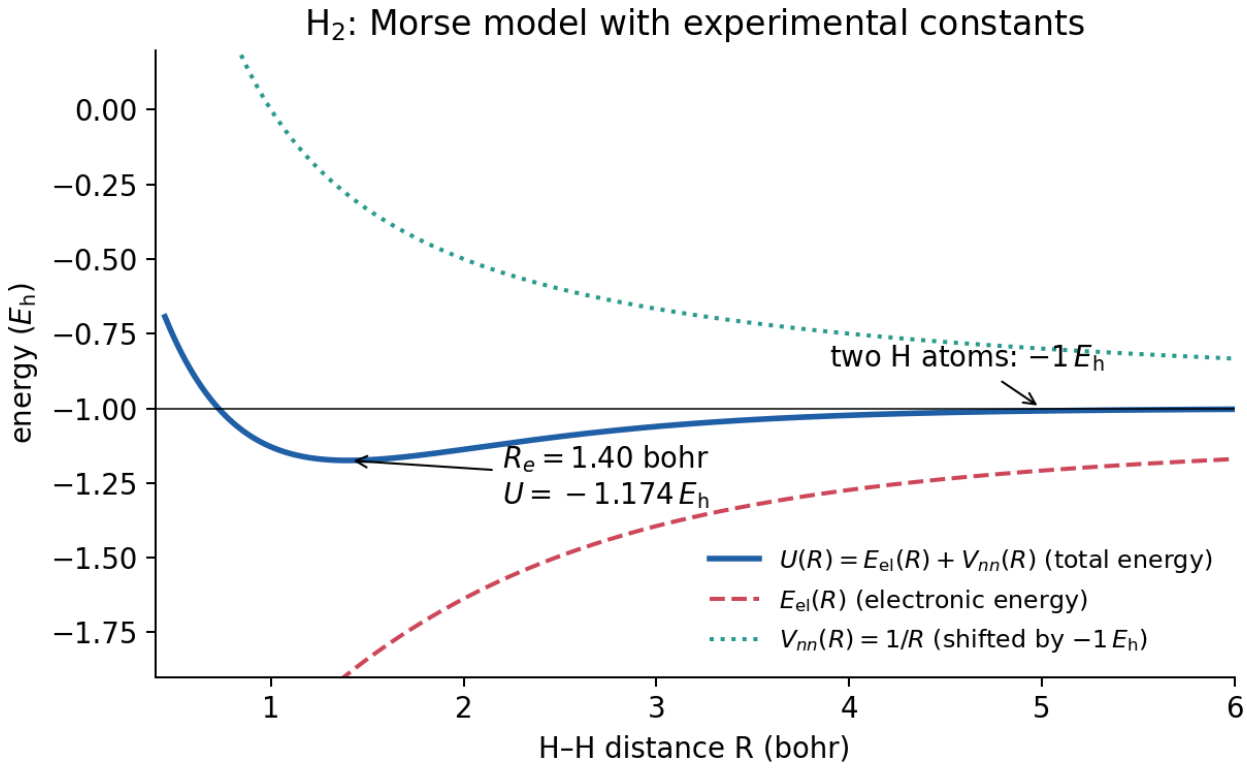


Figure 2: Potential energy curve of H_2, modeled with a Morse function using experimental constants (D_e = 4.75 eV, R_e = 0.741 Å, $\tilde{\nu}_e$ = 4401 cm^{-1}), in atomic units. The total energy $U(R)$ (solid) is the sum of the electronic energy $E_{\mathrm{el}}(R)$ (dashed) and the nuclear repulsion $V_{nn} = 1/R$ (dotted, shifted down by 1 E_{h} to fit on the same axes). At large R the curve approaches the energy of two separate H atoms, $2 \times (-0.5) = -1\ E_{\mathrm{h}}$.

1.4.3 What "total energy" means in a Q-Chem output

When Q-Chem prints

```
Total energy in the final basis set =     -460.xxxxxxxxxx
```

that number is $U(\mathbf{R})$ of Equation 8, computed *at the one geometry you gave in the input*: the electronic energy plus the nuclear repulsion. It is the energy of a motionless molecule. In particular, it does **not** contain:

- the **zero-point vibrational energy** (ZPE): even at 0 K, a molecule vibrates, and that costs energy;

- **thermal energy** at room temperature (translation, rotation, populated vibrational levels);
- **entropy**, so it is not a free energy.

The reaction energies you compute in Experiments 1–8 are therefore *electronic* reaction energies, ΔE_{el}, and they are compared with reference values of the same kind. Comparing with experiment requires the corrections listed above; that is the subject of Lecture 13 and Experiment 9.

Worked example: nuclear repulsion of HCl

For HCl at $R = 2.41709\ a_0$ (from the previous worked example), $V_{nn} = Z_{\mathrm{Cl}} Z_{\mathrm{H}} / R = 17 \times 1/2.41709 = 7.0332\ E_{\mathrm{h}}$. That is a large positive number, about 4,400 kcal/mol. The total energy is about $-460\ E_{\mathrm{h}}$, so the electronic energy is about $-467\ E_{\mathrm{h}}$. You never need to add V_{nn} yourself: the "total energy" already includes it. Q-Chem also prints the nuclear repulsion energy separately near the top of the output, which is a quick check that the geometry was read correctly.

! Common mistake

"Total energy" in a quantum-chemistry output does *not* mean "energy including everything". It means the electronic energy plus nuclear repulsion at a fixed geometry. Never compare it directly with an experimental enthalpy.

1.5 Many electrons

1.5.1 Why we cannot solve it

For a one-electron system, such as H or He^+ or H_2^+, the electronic Schrödinger equation can be solved exactly (analytically for atoms, numerically to any precision for H_2^+). With two or more electrons the term

$$\hat{V}_{ee} = \sum_i \sum_{j>i} \frac{1}{r_{ij}}$$

couples the coordinates of every electron to every other one. The motion of electron 1 depends on where electron 2 is, and vice versa. The equation no longer separates into one-electron problems, and no analytic solution exists. Numerically, the difficulty is the dimension: Ψ is a function of $3N$ spatial coordinates (plus N spin coordinates). A benzene molecule has 42 electrons, so its wave function lives in a 126-dimensional space. Storing it on even a coarse grid of 10 points per coordinate would take 10^{126} numbers.

The helium atom shows how big the effect of $\hat{V}_{ee}$ is:

Table 2: The helium atom. Ignoring electron repulsion is a disaster; a single determinant of the best orbitals gets most of the energy; the remaining 0.042 E_h (26 kcal/mol) is the *correlation energy* of Lecture 4.

Model for He	Energy (E_h)
Ignore $\hat{V}_{ee}$ entirely (two independent hydrogen-like electrons with $Z = 2$)	$2 \times (-Z^2/2) = -4.000$
Best possible single determinant (Hartree–Fock limit)	-2.8617
Exact nonrelativistic energy	-2.9037

Notice the size of the last difference. The error of the best single-determinant wave function is "only" 1.4% of the total energy, yet 26 kcal/mol is far larger than the reaction energies we care about. This is why approximations to the electron–electron interaction are the heart of quantum chemistry.

1.5.2 Spin and the Pauli principle

Electrons have spin, an intrinsic angular momentum with two possible projections, $m_s = +\tfrac{1}{2}$ (α, "up") and $m_s = -\tfrac{1}{2}$ (β, "down"). It has no classical analogue and does not appear in the Hamiltonian of Equation 6, yet it has dramatic consequences through the **Pauli principle**: the wave function of a system of electrons must be **antisymmetric** with respect to the exchange of any two electrons,

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = -\Psi(\mathbf{x}_2, \mathbf{x}_1, \ldots, \mathbf{x}_N), \tag{9}$$

where $\mathbf{x}_i = (\mathbf{r}_i, s_i)$ collects the three spatial coordinates and the spin coordinate of electron i. The familiar "no two electrons in the same state" rule is a consequence of Equation 9, not the other way around.

Spin matters for every calculation you run. The second line of each geometry file in this lab gives the charge and the **spin multiplicity** $2S + 1$, where S is the total spin quantum number. Closed-shell molecules such as HCl have all electrons paired: $S = 0$, a **singlet** (multiplicity 1). A hydrogen atom has one unpaired electron, $S = \tfrac{1}{2}$, a **doublet** (2). The ground state of O_2 and of the O atom has two unpaired electrons with parallel spins, $S = 1$, a **triplet** (3). Get the multiplicity wrong and you compute a different electronic state, with a different energy.

1.5.3 Orbitals and the Slater determinant

How do we write down a wave function that is automatically antisymmetric? Start from one-electron functions. A **spatial orbital** $\phi(\mathbf{r})$ times a spin function, α or β, is a **spin orbital** $\varphi(\mathbf{x})$. Each spatial orbital can therefore hold two electrons, one with each spin.

A simple product of spin orbitals, $\varphi_1(\mathbf{x}_1)\varphi_2(\mathbf{x}_2)$, is not antisymmetric. The fix, for two electrons, is to subtract the swapped product:

$$\Psi(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{\sqrt{2}}\Big[\,\varphi_1(\mathbf{x}_1)\varphi_2(\mathbf{x}_2) - \varphi_2(\mathbf{x}_1)\varphi_1(\mathbf{x}_2)\Big] = \frac{1}{\sqrt{2}}\begin{vmatrix} \varphi_1(\mathbf{x}_1) & \varphi_2(\mathbf{x}_1) \\ \varphi_1(\mathbf{x}_2) & \varphi_2(\mathbf{x}_2) \end{vmatrix}. \tag{10}$$

The second form is a determinant, and it generalizes to N electrons. The **Slater determinant** [5] is

$$\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \varphi_1(\mathbf{x}_1) & \varphi_2(\mathbf{x}_1) & \cdots & \varphi_N(\mathbf{x}_1) \\ \varphi_1(\mathbf{x}_2) & \varphi_2(\mathbf{x}_2) & \cdots & \varphi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(\mathbf{x}_N) & \varphi_2(\mathbf{x}_N) & \cdots & \varphi_N(\mathbf{x}_N) \end{vmatrix}. \tag{11}$$

Rows are electrons, columns are spin orbitals. Two properties of determinants do all the work:

- **Swapping two rows changes the sign** of a determinant. Swapping two rows means exchanging two electrons, so Equation 11 obeys Equation 9 automatically.
- **A determinant with two identical columns is zero.** Two electrons in the same spin orbital give $\Psi = 0$: this is the Pauli exclusion principle.

The Slater determinant is the *simplest* antisymmetric wave function we can build from orbitals. It is the wave function of Hartree–Fock theory (Lecture 4), and the Kohn–Sham determinant of density functional theory (Lecture 5) has the same form. It is not the exact wave function: a single determinant describes each electron as moving in the *average* field of the others, and misses the way electrons instantaneously avoid each other. That missing piece is **electron correlation**.

> **i** Orbitals are building blocks, not observables
>
> For one-electron systems, orbitals are exact wave functions. For many-electron systems they are a *mathematical device* for building an approximate Ψ. This is why the orbital pictures in IQmol are so useful for intuition, and also why you should not over-interpret them: different methods give somewhat different orbitals for the same molecule.

1.5.4 The variational principle

We cannot solve Equation 7 exactly, so how can we tell whether one approximate wave function is better than another? The **variational principle** gives the answer. For any normalized, well-behaved trial function Ψ_{trial} that satisfies the right boundary conditions and symmetry,

$$E[\Psi_{\text{trial}}] = \int \Psi^*_{\text{trial}} \hat{H} \Psi_{\text{trial}} \, d\tau \geq E_0, \tag{12}$$

where E_0 is the exact ground-state energy. Equality holds only if Ψ_{trial} is the exact ground state.

The proof takes three lines. Expand the trial function in the (unknown) exact eigenfunctions, $\Psi_{\text{trial}} = \sum_k c_k \Psi_k$, with $\sum_k |c_k|^2 = 1$. Then $E[\Psi_{\text{trial}}] = \sum_k |c_k|^2 E_k$. Since every $E_k \geq E_0$, the weighted average can never be below E_0.

The consequences are practical:

1. **Lower is better.** Among trial wave functions of the same Hamiltonian, the one with the lower energy is closer to the exact ground state (in energy).
2. **Optimize parameters.** Give the trial function adjustable parameters and minimize the energy with respect to them. The result is the best approximation of that form.
3. **More flexibility can only help.** If one family of trial functions contains another (for example, a bigger set of basis functions that includes all the functions of a smaller one), the minimum energy of the bigger family is lower or equal.

Point 3 is the reason basis sets work, as you will see in Lecture 2.

Worked example: the variational principle for helium

Take as a trial function for He a product of two hydrogen-like 1s orbitals with an adjustable exponent ζ (an "effective nuclear charge"), $\Psi = \frac{\zeta^3}{\pi} e^{-\zeta r_1} e^{-\zeta r_2}$, times the singlet spin function. Evaluating $\langle \hat{H} \rangle$ for helium ($Z = 2$) gives

$$E(\zeta) = \zeta^2 - 2Z\zeta + \frac{5}{8}\zeta.$$

Setting $dE/d\zeta = 0$ gives $\zeta = Z - \frac{5}{16} = 1.6875$ and $E = -\zeta^2 = -2.8477\ E_\mathrm{h}$. Compare with Table 2. The optimized energy is above the exact $-2.9037\ E_\mathrm{h}$, as the variational principle guarantees, and well below the unscreened $\zeta = 2$ result ($E = 4 - 8 + 1.25 = -2.75\ E_\mathrm{h}$). The optimized $\zeta < 2$ has a physical meaning: each electron partly *screens* the nucleus from the other. Adding more flexibility (more parameters) would bring the energy closer to -2.9037.

! When "lower is better" does not apply

The variational principle compares energies computed **with the same Hamiltonian**. Hartree–Fock and the basis-set expansion are variational, so a bigger basis always gives a lower or equal HF energy. But approximate density functionals (Lectures 5–6) use *approximate* Hamiltonians: a functional that gives a lower total energy is not, for that reason, a better functional. Comparing total energies across methods is meaningless; comparing **reaction energies** with a reliable reference is how methods are judged (Lecture 7).

1.6 The two approximations of every calculation

We can now state precisely what any practical electronic-structure calculation does. It makes two approximations, and they are independent of each other.

Approximation 1: the wave function (the basis set). The orbitals are unknown functions in three dimensions. A computer cannot store an arbitrary function, so we expand each orbital in a finite set of known functions, the **basis set**. The finite expansion restricts the shapes the orbitals can take, and by the variational principle the energy is too high. As the basis grows, this **basis-set incompleteness error** (BSIE) shrinks toward the *complete-basis-set* (CBS) limit. This is Lecture 2 and Experiment 1.

Approximation 2: the Hamiltonian (the method). Even with a complete basis, we cannot handle $\hat{V}_{ee}$ exactly. A **method** decides how to treat it. Hartree–Fock replaces it by an average field (Lecture 4). Density functional theory replaces the hard part by an approximate *exchange–correlation functional* (Lectures 5–6). Wave-function methods such as MP2 or CCSD(T) add correlation on top of Hartree–Fock, at a much higher cost.

The pair of choices is called a **model chemistry** or **level of theory**, and it is written *method/basis*:

Table 3: Decoding a model chemistry. You will meet each of these names in Experiment 1.

Model chemistry	Method (approximates $\hat{H}$)	Basis set (approximates Ψ)
HF/cc-pVTZ	Hartree–Fock	Dunning triple-ζ
B3LYP/def2-SVP	B3LYP hybrid functional	Karlsruhe double-ζ
ωB97M-V/def2-QZVP	ωB97M-V range-separated hybrid meta-GGA	Karlsruhe quadruple-ζ

The total error of a computed reaction energy is, roughly, method error + basis-set error (+ smaller numerical errors such as integration grids and SCF convergence). One of the goals of this lab is to learn to keep these errors separate. Experiment 1 holds the method fixed and varies the basis; Experiments 3–5 hold the basis fixed and vary the method; Experiment 6 shows what happens when the two errors cancel each other.

i What comes next

Lecture 2 opens the first approximation: what basis functions look like, why everyone uses Gaussians, and how to read names such as 6-31G*, def2-TZVP and cc-pVQZ. Lecture 3 turns all of this into a real calculation in IQmol and Q-Chem.

1.7 Key points

- A quantum-chemistry program computes an energy for a fixed arrangement of nuclei. Reaction energies, barriers, structures and frequencies are all built from such energies.
- The molecular Hamiltonian has five terms: $\hat{T}_n$, $\hat{T}_e$, $\hat{V}_{ne}$, $\hat{V}_{ee}$ and $\hat{V}_{nn}$. In atomic units ($\hbar = m_e = e = 4\pi\varepsilon_0 = 1$) energies are in hartree and lengths in bohr.
- 1 E_h = 627.5095 kcal/mol = 2625.50 kJ/mol = 27.2114 eV; 1 bohr = 0.529177 Å.
- The Born–Oppenheimer approximation clamps the nuclei and gives the electronic Schrödinger equation; $E_\mathrm{el}(\mathbf{R}) + V_{nn}(\mathbf{R})$ is the potential energy surface.
- Q-Chem's "total energy" is $E_\mathrm{el} + V_{nn}$ at one fixed geometry: no zero-point energy, no thermal corrections, no entropy.
- The electron–electron repulsion couples all electrons, so no system with two or more electrons can be solved exactly.
- Electrons are fermions: the wave function must be antisymmetric. A Slater determinant of spin orbitals is the simplest antisymmetric wave function and enforces the Pauli principle automatically.
- The variational principle: any trial wave function gives an energy at or above the exact ground-state energy, for the same Hamiltonian. Lower is better, and more flexibility can only lower the energy.
- Every calculation approximates Ψ (the basis set) and $\hat{H}$ (the method). A model chemistry is written method/basis.

1.8 Study questions

1. Write the five terms of the molecular Hamiltonian for the H_2^+ ion in atomic units. Which term is absent compared with H_2, and why does its absence make H_2^+ exactly solvable within the Born–Oppenheimer approximation?
2. The total energy of a molecule is printed as −76.4 E_h. Express it in kcal/mol, kJ/mol and eV. Which unit would you use to report a reaction energy in this lab?

3. A reaction energy is computed as $0.0158\ E_\mathrm{h}$. Convert it to kcal/mol. If each of the two total energies used to compute it had been rounded to 3 decimal places, what is the largest error (in kcal/mol) the rounding could introduce?
4. Why does the Born–Oppenheimer approximation work so well? Give the mass ratio between a ^{35}Cl nucleus and an electron, and explain why the approximation is expected to be slightly worse for molecules containing hydrogen.
5. Compute the nuclear repulsion energy of H_2 at its equilibrium bond length of 0.741 Å, in hartree and in kcal/mol.
6. Explain in your own words what the "total energy" in a Q-Chem output contains, and list three contributions that would be needed to compare it with an experimental reaction enthalpy at 298 K.
7. Write the Slater determinant for the ground state of the lithium atom (configuration $1s^2\ 2s^1$) using spin orbitals $1s\alpha$, $1s\beta$, $2s\alpha$. Show that putting the third electron in $1s\alpha$ instead gives $\Psi = 0$.
8. Give the spin multiplicity of: HCl, the H atom, the Cl atom, the O atom (ground state), O_2 (ground state), and the CH_3 radical.
9. Calculation A gives $E = -2.8477\ E_\mathrm{h}$ for He and calculation B gives $-2.8617\ E_\mathrm{h}$, both with the exact Hamiltonian. Which wave function is better, and how do you know? Could a calculation with the exact Hamiltonian give $-2.95\ E_\mathrm{h}$?
10. Two students compute the total energy of water with two different density functionals and get −76.42 and −76.38 E_h. One argues the first functional is better "because of the variational principle". Is the argument correct? Explain.

1.9 Further reading

- The Live Textbook of Physical Chemistry 2, chapters 3, 5–8, 10 and 11, for the review material of this chapter.
- A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* [6], chapters 1–2: the classic, careful introduction to the electronic Hamiltonian, Slater determinants and their algebra.
- T. Helgaker, P. Jørgensen and J. Olsen, *Molecular Electronic-Structure Theory* [7]: the advanced reference for everything in Lectures 1–4.
- W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory* [8], chapter 1: a compact review of the many-electron problem from the point of view of DFT users.

2 Approximating the wave function: basis sets

Learning goals

After studying this chapter you should be able to:

- explain the LCAO expansion and why, by the variational principle, a bigger basis gives a lower (or equal) energy;
- compare Slater-type and Gaussian-type functions and explain why almost every program uses contracted Gaussians;
- explain what split valence (ζ), polarization and diffuse functions add, and when each matters;
- decode basis-set names such as STO-3G, 6-31G*, 6-311+G**, def2-TZVP, cc-pVQZ and aug-cc-pVTZ;
- count the basis functions of a small molecule from the shell composition of each atom;
- describe the complete-basis-set (CBS) limit, how the cost grows with the number of basis functions, and what the basis-set incompleteness error (BSIE) and basis-set superposition error (BSSE) are.

Review from P-Chem 2

- The chemical bond in H_2^+ and H_2 (chapter 12): the LCAO idea in its simplest form.
- The water molecule with a minimal basis and Hartree–Fock calculation for water (§13.1–13.2): seven basis functions, the overlap matrix and the integrals.
- The Hartree–Fock method (§11.3.1): orbitals expanded in basis functions.
- The variational method (§10.2.1).

2.1 Molecular orbitals from basis functions

2.1.1 The LCAO expansion

Lecture 1 ended with the first of the two approximations of every calculation: the orbitals are unknown functions of three coordinates, and a computer cannot store an arbitrary function. The solution is to write each molecular orbital ϕ_i as a **linear combination** of a finite number K of known functions χ_μ, the **basis functions**:

$$\phi_i(\mathbf{r}) = \sum_{\mu=1}^{K} c_{\mu i}\, \chi_\mu(\mathbf{r}). \tag{13}$$

The basis functions are fixed in advance; only the **coefficients** $c_{\mu i}$ are optimized. In P-Chem 2 you met this as LCAO, the *linear combination of atomic orbitals*: for H_2^+, one 1s function on each proton; for water, a 1s, 2s and three 2p functions on oxygen plus one 1s on each hydrogen, seven functions in total. Modern basis functions are not atomic orbitals in the strict sense (they are not eigenfunctions of any atom), but they are still **centered on the atoms** and look like atomic orbitals, so the name has stuck.

With Equation 13, the hard problem of finding unknown *functions* becomes the much easier problem of finding unknown *numbers*: a $K \times K$ matrix of coefficients. In Hartree–Fock and Kohn–Sham DFT this leads to a matrix eigenvalue equation, the Roothaan–Hall equation $\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon}$, which you will meet in Lecture 4. Here $\mathbf{S}$ is the **overlap matrix**, $S_{\mu\nu} = \int \chi_\mu^* \chi_\nu \, d\mathbf{r}$. Basis functions on different atoms overlap, so $\mathbf{S}$ is not the identity matrix; §13.2 of the P-Chem 2 textbook shows it for water.

K basis functions give exactly K molecular orbitals. Of these, $N/2$ are occupied in a closed-shell molecule with N electrons; the rest are **virtual** (unoccupied) orbitals. The number of basis functions K is printed in every Q-Chem output and is one of the four numbers you record in Experiment 1.

2.1.2 The variational principle and the CBS limit

If the set $\{\chi_\mu\}$ were **complete**, any function could be expanded exactly in it, and Equation 13 would be no approximation at all. A complete set is infinite, so any real basis is a **truncation**. The orbitals can only take the shapes that the chosen functions allow.

The variational principle of Lecture 1 tells us what happens as we enlarge the basis. For Hartree–Fock and for Kohn–Sham DFT with a given functional, the energy is minimized with respect to the coefficients. Adding basis functions enlarges the space of allowed orbitals, so the minimum can only go down or stay the same:

$$E(\text{basis}_1) \geq E(\text{basis}_2) \quad \text{if basis}_2 \supseteq \text{basis}_1.$$

As the basis grows toward completeness, the energy converges from above to a limit, the **complete-basis-set (CBS) limit**. The difference between the energy in a finite basis and the CBS energy (with the same method) is the **basis-set incompleteness error** (BSIE). It is the error that Experiment 1 is designed to measure.

> **i** Strictly nested or not?
>
> The inequality is guaranteed only if the bigger basis *contains* the smaller one. Real basis-set families are re-optimized at each size, so def2-TZVP does not literally contain def2-SVP. In practice the total energy still decreases monotonically along a well-designed family, which is what matters for us. Reaction energies, being differences, do not have to converge monotonically, and Experiment 1 asks you to look at exactly this.

2.2 Slater-type and Gaussian-type functions

What should the basis functions look like? Two choices dominate the history of quantum chemistry.

2.2.1 Slater-type orbitals

Slater-type orbitals (STOs) mimic the hydrogen-atom solutions:

$$\chi^{\text{STO}}(r,\theta,\varphi) = N\, r^{n-1} e^{-\zeta r}\, Y_\ell^m(\theta,\varphi). \tag{14}$$

The **exponent** ζ (zeta) sets the size: a large ζ gives a tight function close to the nucleus; a small ζ gives a diffuse one. STOs have the right physics. They have a **cusp** at the nucleus (a finite slope at $r = 0$, as every exact wave function must [3]) and an **exponential tail** at large r, like real bound states (Lecture 1).

Their weakness is computational. A molecular calculation needs huge numbers of integrals over basis functions, the worst being the **two-electron integrals**

$$(\mu\nu|\lambda\sigma) = \iint \frac{\chi_\mu(\mathbf{r}_1)\chi_\nu(\mathbf{r}_1)\,\chi_\lambda(\mathbf{r}_2)\chi_\sigma(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}\, d\mathbf{r}_1 d\mathbf{r}_2, \tag{15}$$

in which the four functions may sit on four different atoms. For STOs these multicenter integrals have no closed form and must be computed numerically, which is slow.

2.2.2 Gaussian-type orbitals

In 1950, Boys proposed using **Gaussian-type orbitals** (GTOs) instead [9]:

$$\chi^{\mathrm{GTO}}(x,y,z) = N\, x^a y^b z^c\, e^{-\alpha r^2}, \qquad \ell = a+b+c. \tag{16}$$

The sum $\ell = a+b+c$ sets the angular momentum: $\ell = 0$ is an s function, $\ell = 1$ a p function (x, y or z in front), $\ell = 2$ a d function, and so on. The exponent α plays the role of ζ.

GTOs have the *wrong* physics in two places, visible in Figure 4:

- **No cusp.** A Gaussian is flat at the nucleus ($d\chi/dr = 0$ at $r = 0$), and its value there is too small.
- **Wrong tail.** $e^{-\alpha r^2}$ decays much faster than $e^{-\zeta r}$. Far from the nucleus, a single Gaussian underestimates the wave function by orders of magnitude.

2.2.3 The Gaussian product theorem

Why use the wrong functions? Because of one beautiful property. The product of two Gaussians centered on two different points $\mathbf{A}$ and $\mathbf{B}$ is a single Gaussian centered at a point $\mathbf{P}$ on the line between them:

$$e^{-\alpha|\mathbf{r}-\mathbf{A}|^2}\, e^{-\beta|\mathbf{r}-\mathbf{B}|^2} = K_{AB}\, e^{-p|\mathbf{r}-\mathbf{P}|^2}, \qquad p = \alpha+\beta, \quad \mathbf{P} = \frac{\alpha\mathbf{A}+\beta\mathbf{B}}{\alpha+\beta}, \quad K_{AB} = e^{-\frac{\alpha\beta}{\alpha+\beta}|\mathbf{A}-\mathbf{B}|^2}. \tag{17}$$

This is the **Gaussian product theorem** (Figure 3). It collapses every pair of functions in Equation 15 to a single center, so a four-center integral becomes a two-center one, which has an analytic formula. Gaussian integrals are so cheap that it pays to use several Gaussians to imitate one STO. This trade, more functions but much faster integrals, is the reason virtually every molecular quantum-chemistry program, Q-Chem included, uses Gaussian basis functions.

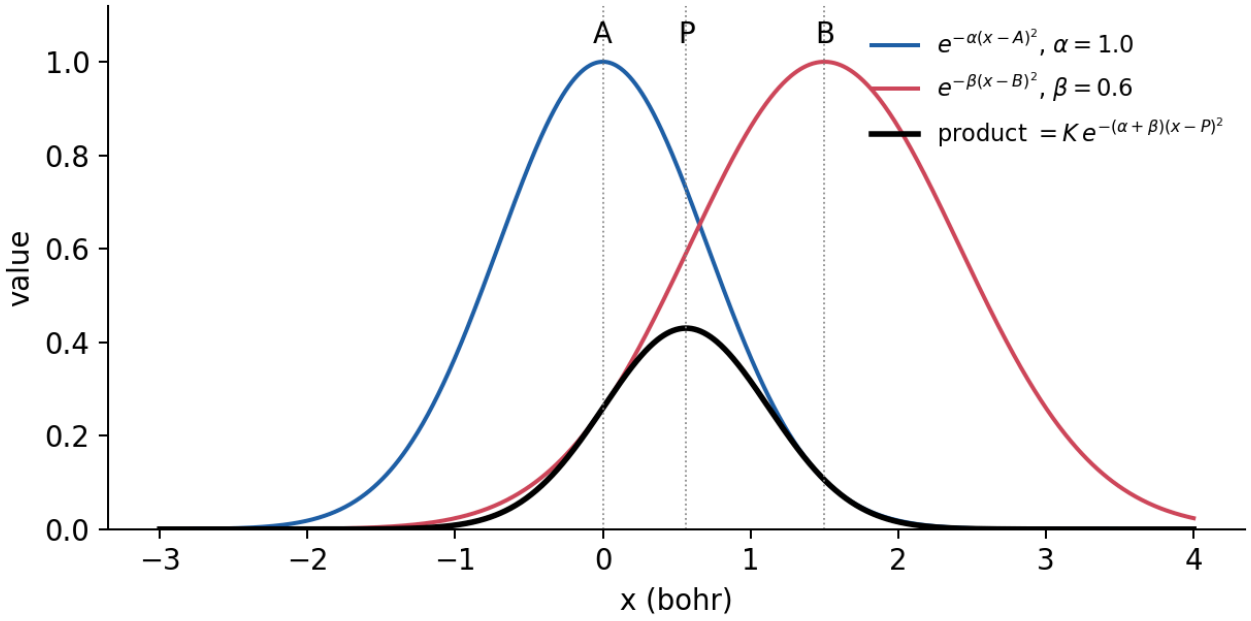


Figure 3: The Gaussian product theorem in one dimension. The product (black) of a Gaussian centered at A (exponent 1.0) and one centered at B (exponent 0.6) is exactly a single Gaussian centered at P, with exponent 1.6 and height reduced by the constant K_{AB} of Equation 17.

2.2.4 Contracted Gaussians: STO-3G

A **contracted Gaussian** is a fixed linear combination of several **primitive** Gaussians:

$$\chi_\mu(\mathbf{r}) = \sum_{k=1}^{L} d_{k\mu}\, g_k(\alpha_k, \mathbf{r}). \tag{18}$$

The exponents α_k and the **contraction coefficients** $d_{k\mu}$ are chosen once, when the basis set is designed, and never change during a calculation. The program optimizes only the coefficients $c_{\mu i}$ of Equation 13, one per *contracted* function. Contraction gives functions with a better shape without increasing K.

The simplest example is **STO-3G** [10]: each Slater-type orbital is replaced by a contraction of three primitive Gaussians, fitted to reproduce it as closely as possible. For the hydrogen 1s function, which in STO-3G mimics a Slater function with $\zeta = 1.24$ (slightly tighter than the free-atom value of 1.0, because a hydrogen in a molecule holds its electron more tightly), the published primitives are:

Table 4: STO-3G contraction for the hydrogen 1s function (coefficients for normalized primitives).

k	exponent α_k (bohr^{-2})	coefficient d_k
1	3.42525091	0.15432897
2	0.62391373	0.53532814
3	0.16885540	0.44463454

The tight primitive ($\alpha \approx 3.4$) builds up the peak near the nucleus; the diffuse one ($\alpha \approx 0.17$) supplies the tail. Figure 4 compares the result with the Slater function it imitates and with the best *single* Gaussian (the one with maximum overlap with the STO, $\alpha = 0.417$). The single Gaussian has an overlap of 0.978 with the STO; the three-Gaussian contraction reaches 0.9998. Even so, STO-3G still misses the cusp, and its tail still falls too fast beyond about 5 bohr.

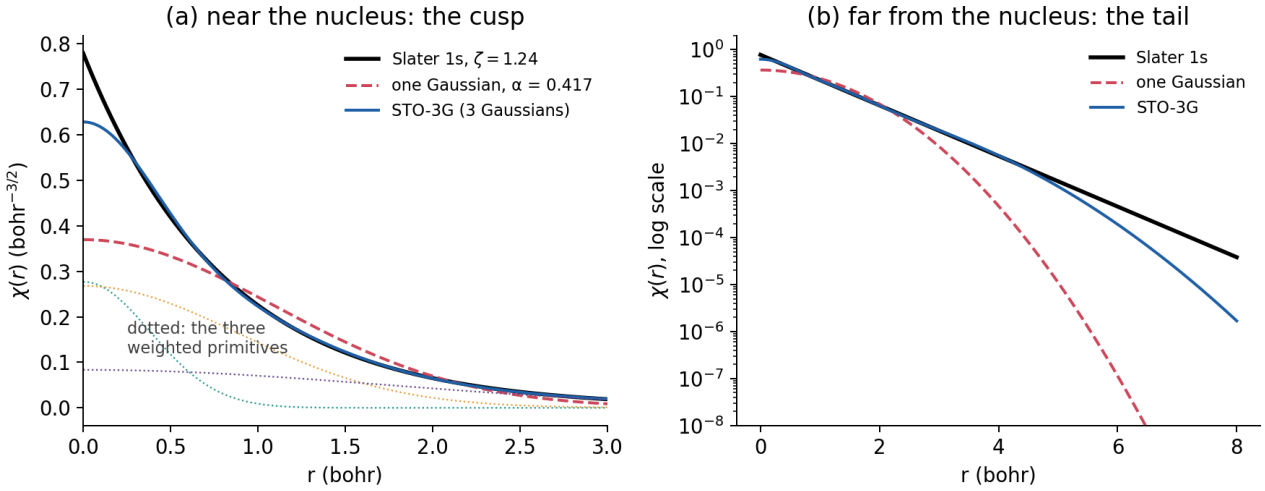


Figure 4: The hydrogen 1s Slater function with $\zeta = 1.24$ (black), the best single Gaussian (dashed red) and the STO-3G contraction of three Gaussians (blue), using the exponents and coefficients of Table 4. (a) Near the nucleus, the Slater function has a cusp; Gaussians are flat at $r = 0$. The dotted curves are the three weighted primitives that add up to the STO-3G function. (b) On a logarithmic scale, the Slater function decays as a straight line ($e^{-\zeta r}$); Gaussians bend down ($e^{-\alpha r^2}$).

The shell composition of a basis set is often written as (primitives)/[contracted functions]. For STO-3G on oxygen, (6s3p)/[2s1p]: six s primitives contracted to two s functions (1s and 2s), and three p primitives

contracted to one set of p functions. A "set" of p functions means p_x, p_y and p_z, so [2s1p] is $2 + 3 = 5$ basis functions.

2.2.5 Cartesian and spherical d functions

Equation 16 with $\ell = 2$ gives six **Cartesian** d functions: x^2, y^2, z^2, xy, xz, yz (times the Gaussian). But a true d shell has only five components. The extra one is the combination $x^2 + y^2 + z^2 = r^2$, which is spherical: it behaves like an s function. **Spherical** (or *pure*) d functions keep only the five real combinations with $\ell = 2$ (d_{z^2}, $d_{x^2-y^2}$, d_{xy}, d_{xz}, d_{yz}). The same happens for higher ℓ:

Table 5: Number of functions per shell. Cartesian: $(\ell + 1)(\ell + 2)/2$; spherical: $2\ell + 1$.

Shell	Cartesian components	Spherical (pure) components
s	1	1
p	3	3
d	6 ("6d")	5 ("5d")
f	10	7
g	15	9

The choice changes K, the energy (slightly) and the cost. Each basis set was designed with one convention in mind: the Pople sets such as 6-31G* were defined with six Cartesian d functions, while the def2 and correlation-consistent sets were designed for spherical functions. Programs have defaults that you can override, so **always check the number of basis functions in your output** against what you expect. This is one of the checks you will do in Experiment 1.

2.3 Building better basis sets

A minimal basis such as STO-3G has a fundamental limitation: each occupied atomic orbital gets exactly one function, with a fixed size and shape. In a molecule, orbitals need to change in three ways, and each one is addressed by a different kind of extra function.

2.3.1 Split valence: letting orbitals change size

The core orbitals (1s on C, N, O; 1s2s2p on Cl) hardly change when atoms form bonds. The **valence** orbitals change a lot: they expand or contract depending on the bonding partner and on the charge. A **split-valence** basis gives each valence orbital two or more functions with different exponents, a tight one and a loose one. The variational optimization then chooses the right mixture, effectively tuning the size of the orbital in each direction and in each molecule.

The number of functions per valence orbital is called **ζ** ("zeta", from the STO exponent):

- **minimal** (single-ζ): one function per occupied orbital (STO-3G);
- **double-ζ** (DZ): two functions per valence orbital;
- **triple-ζ** (TZ): three;
- **quadruple-ζ** (QZ): four; and so on.

In the Dunning families, ζ is labeled by the **cardinal number** n = 2 (D), 3 (T), 4 (Q), 5, 6.

Worked example: what does split valence buy?

For oxygen, a minimal basis has 1s, 2s, 2p: $1 + 1 + 3 = 5$ functions. A split-valence double-ζ basis keeps one core 1s function but doubles the valence: $1 + 2 + 2 \times 3 = 9$ functions (written [3s2p]). Water goes from $5 + 2 \times 1 = 7$ functions (minimal) to $9 + 2 \times 2 = 13$ (double-ζ valence on both O and H, no polarization). Splitting the valence nearly doubles the size of the basis, before adding anything else.

2.3.2 Polarization functions: letting orbitals change shape

An s function is spherical; a p function points along one axis. When an atom sits in a molecule, the electric field of its neighbors distorts its electron cloud: the orbitals need to bend and shift off-center. The way to allow this is to add functions with **higher angular momentum** than any occupied orbital of the free atom: p functions on hydrogen, d functions on C, N, O and Cl, f functions on top of those, and so on. Mixing a little p into an s function shifts its center of charge; mixing d into p bends it. These are **polarization functions**.

Polarization functions are essential for almost every property: bond angles (NH_3 is too flat without them), reaction energies, dipole moments and, very importantly for this lab, the weak interactions between molecules. In larger basis sets, they come in several layers: def2-TZVP has 2d1f on oxygen; cc-pVQZ has 3d2f1g.

2.3.3 Diffuse functions: the far tail

Diffuse functions are extra functions with very small exponents, so they extend far from the nucleus. They are needed whenever electrons spend time far out:

- **anions**, whose extra electron is weakly bound and spread out;
- **weak intermolecular interactions** (hydrogen bonds, dispersion), which depend on the overlap of the outer tails of two molecules;
- **Rydberg excited states** and properties such as polarizabilities.

Diffuse functions are expensive and can cause numerical problems (near-linear dependence of the basis) in larger molecules. Use them when the chemistry needs them.

Three questions, three kinds of functions

Size → more ζ (split valence). Shape → polarization (higher ℓ). Reach → diffuse (small α). A good basis set is **balanced**: it improves all three together rather than piling up one kind.

2.4 Pople basis sets

The basis sets developed by John Pople and co-workers in the 1970s and 1980s are the most widely known, and you will see their names in thousands of papers. Their names are a code for how the functions are contracted.

Table 6: Reading a Pople basis-set name. The numbers before "G" describe contractions: the digit before the hyphen is the core, the digits after it are the valence functions.

Name	Meaning
STO-3G	minimal; each STO fitted by 3 Gaussians [10]
3-21G	core: 3 primitives in 1 function; valence split in two: 2 primitives + 1 primitive [11]
6-31G	core: 6 primitives; valence: 3 + 1 (double-ζ) [12], [13]
6-311G	core: 6; valence: 3 + 1 + 1 (triple-ζ) [14]
6-31G* = 6-31G(d)	6-31G plus one set of d polarization functions on heavy (non-hydrogen) atoms [15], [16]
6-31G** = 6-31G(d,p)	as above, plus p functions on hydrogen
6-31+G*	plus diffuse s and p functions on heavy atoms
6-31++G**	plus diffuse s functions on hydrogen too
6-311G**	triple-ζ valence, d on heavy atoms, p on H

Worked example: counting functions in 6-31G*

Water. Oxygen in 6-31G: one core 1s function, two 2s functions and two sets of 2p functions, [3s2p] = $3 + 6 = 9$. The star adds one set of d functions: 6 if Cartesian (as 6-31G* was defined), 5 if spherical. Hydrogen in 6-31G has two s functions and no polarization in 6-31G* (only the second star adds p on H). Total: $9 + 6 + 2 \times 2 = 19$ (Cartesian d) or 18 (spherical d).

HCl. Chlorine in 6-31G: one 1s, then the 2s2p shell (one s + one p set), then the valence 3s3p split in two (two s + two p sets): [4s3p] = $4 + 9 = 13$; plus d: 6 or 5. Hydrogen: 2. Total: $13 + 6 + 2 = 21$ (Cartesian) or 20 (spherical). For comparison, STO-3G gives HCl only 10 functions: [3s2p] = 9 on Cl, 1 on H.

Pople sets were designed for Hartree–Fock in the era of slow computers, and they were extraordinarily successful. Their limitations are well known today: the s and p functions of each valence shell share exponents (to save time, which no longer matters), polarization is added in an ad hoc way, and the sequence 6-31G* → 6-311G** is **not designed to converge systematically** toward the CBS limit. You should recognize them, and you will meet 6-31G* again in Experiment 6, where its famous partnership with B3LYP gets a closer look. For new work, the families in the next section are better choices.

2.5 The modern families used in this course

2.5.1 Karlsruhe def2 basis sets

The **def2** basis sets of Weigend and Ahlrichs [17] (the "Ahlrichs" or "Karlsruhe" family) were designed to give consistent quality across almost the whole periodic table, H to Rn. The names describe the valence quality:

- **def2-SVP**: split valence (double-ζ) plus polarization;
- **def2-TZVP**: triple-ζ valence plus polarization;
- **def2-QZVP**: quadruple-ζ valence plus polarization.

Variants with an extra P (def2-TZVPP, def2-QZVPP) add more polarization, and variants ending in D add diffuse functions. "def2" stands for the second generation of the Karlsruhe *default* sets. For elements beyond krypton, def2 replaces the inner-core electrons with an **effective core potential** (ECP): a fixed potential that represents the core and includes the scalar-relativistic effects that matter for heavy atoms. Only the outer electrons are treated explicitly. You will use this in Experiments 8 and 10 for ruthenium and palladium.

The def2 family has been found to be among the most cost-effective choices for DFT [18], which is one reason it is the default basis of most experiments in this lab.

2.5.2 Dunning correlation-consistent basis sets

The **correlation-consistent** basis sets of Dunning and co-workers are written **cc-pVnZ**: *correlation-consistent polarized valence n-zeta*, with n = D, T, Q, 5, 6 [19], [20]. They were built for correlated wave-function methods (MP2, CCSD(T)). The design rule is that functions are added in groups that each recover a similar amount of **correlation energy**. Going from cc-pVDZ to cc-pVTZ adds one more s, p and d function *and* the first f function on oxygen, because these contribute comparably. The result is a family that converges **systematically** toward the CBS limit, so results at several n can be extrapolated (Section 2.6.2).

The prefix **aug-** adds one diffuse function of *each* angular momentum present [21]: aug-cc-pVTZ is the standard choice for anions and weak interactions with correlated methods. (Other prefixes, such as cc-pCVnZ for core correlation, exist but are not used here.)

2.5.3 Jensen polarization-consistent basis sets

Jensen's **pc-n** basis sets (pc-0, pc-1, pc-2, pc-3, pc-4) apply the same systematic idea, but are optimized for Hartree–Fock and DFT instead of correlated methods [22]. pc-1 is double-ζ quality. You will use pc-1 in Experiment 6 to compare double-ζ basis sets from different families.

2.5.4 Counting basis functions

The size of every basis set follows from the shell composition of each atom. With spherical (pure) functions, each shell contributes $2\ell + 1$ functions: s = 1, p = 3, d = 5, f = 7, g = 9.

Table 7: Number of basis functions for water, from the published shell compositions (Basis Set Exchange). Spherical (5d, 7f, 9g) functions unless noted.

Basis set	ζ	H	O	K for H_2O
STO-3G	1	[1s] = 1	[2s1p] = 5	7
6-31G* (5d)	2	[2s] = 2	[3s2p1d] = 14	18 (19 with 6d)
6-311G** (5d)	3	[3s1p] = 6	[4s3p1d] = 18	30 (31 with 6d)
def2-SVP	2	[2s1p] = 5	[3s2p1d] = 14	24
def2-TZVP	3	[3s1p] = 6	[5s3p2d1f] = 31	43
def2-QZVP	4	[4s3p2d1f] = 30	[7s4p3d2f1g] = 57	117
cc-pVDZ	2	[2s1p] = 5	[3s2p1d] = 14	24
cc-pVTZ	3	[3s2p1d] = 14	[4s3p2d1f] = 30	58
cc-pVQZ	4	[4s3p2d1f] = 30	[5s4p3d2f1g] = 55	115
aug-cc-pVTZ	3	[4s3p2d] = 23	[5s4p3d2f] = 46	92
pc-1	2	[2s1p] = 5	[3s2p1d] = 14	24

Worked example: def2-TZVP versus cc-pVTZ for water

def2-TZVP. Oxygen [5s3p2d1f]: $5 \times 1 + 3 \times 3 + 2 \times 5 + 1 \times 7 = 5 + 9 + 10 + 7 = 31$. Hydrogen [3s1p]: $3 + 3 = 6$. Water: $31 + 2 \times 6 = 43$.

cc-pVTZ. Oxygen [4s3p2d1f]: $4 + 9 + 10 + 7 = 30$. Hydrogen [3s2p1d]: $3 + 6 + 5 = 14$. Water: $30 + 2 \times 14 = 58$.

Both are "triple-ζ", but they distribute the functions differently. cc-pVTZ gives hydrogen much more polarization (2p1d instead of 1p), because correlated methods need it. def2-TZVP puts more s functions on oxygen. Same label, different size: the name tells you the *quality class*, not the cost.

Notice three things in Table 7. First, double-ζ sets from different families can have the same size (24 for def2-SVP, cc-pVDZ and pc-1 here) yet different exponents and different results. Second, the size grows quickly with ζ: about a factor of 2 from double- to triple-ζ, and another factor of 2–3 to quadruple-ζ. Third, hydrogen atoms are cheap at double-ζ but not at quadruple-ζ (30 functions each).

2.6 Convergence, cost and basis-set errors

2.6.1 More functions, better results

The central message of this chapter fits in one sentence: **in a systematic family, more functions give results that converge toward the CBS limit.** Figure 5 (a) sketches the typical behavior of the total energy. The first step, double- to triple-ζ, is large; each further step is smaller.

For Hartree–Fock and DFT, the total energy converges *fast*, roughly exponentially with the cardinal number [23]:

$$E(n) = E_{\mathrm{CBS}} + A\, e^{-\alpha n}. \tag{19}$$

The **correlation energy** of wave-function methods converges much more slowly, as an inverse power [24], [25]:

$$E_{\mathrm{corr}}(n) = E_{\mathrm{corr,CBS}} + A\, n^{-3}. \tag{20}$$

The reason is physical. Near the point where two electrons meet, the exact wave function has a cusp in the inter-electron distance r_{12}, and smooth orbital products converge to it only slowly. HF and DFT have no such cusp to describe, so they converge faster.

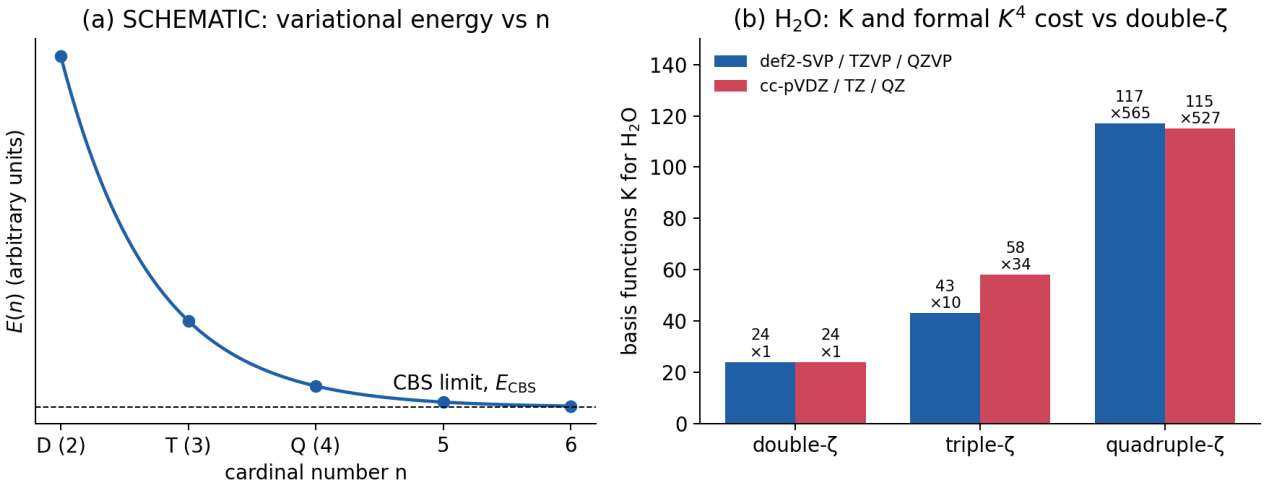


Figure 5: (a) **Schematic** convergence of a variational energy with the cardinal number n of a systematic basis-set family, toward the CBS limit (Equation 19; the curve is illustrative, not data). (b) Real numbers of basis functions K for water in the def2 and cc-pVnZ families (Table 7), with the formal K^4 cost relative to the double-ζ basis written above each bar.

2.6.2 Extrapolation to the CBS limit

Because the convergence is systematic, two or three results at consecutive n can be fit to Equation 19 or Equation 20 to estimate the CBS limit without ever computing it. For the n^{-3} form, two points $m < n$ give

$$E_{\mathrm{CBS}} = \frac{n^3 E_n - m^3 E_m}{n^3 - m^3}. \tag{21}$$

Worked example: two-point extrapolation

Suppose (hypothetical numbers) the MP2 correlation energy of a molecule is $-0.2950\ E_\mathrm{h}$ with cc-pVTZ ($m = 3$) and $-0.3080\ E_\mathrm{h}$ with cc-pVQZ ($n = 4$). Then

$$E_{\mathrm{corr,CBS}} = \frac{64(-0.3080) - 27(-0.2950)}{64 - 27} = \frac{-19.712 + 7.965}{37} = -0.3175\ E_\mathrm{h}.$$

The extrapolated value lies $0.0095\ E_\mathrm{h}$ (6 kcal/mol) below the QZ result: the correlation energy is far from converged even at quadruple-ζ. This is why reference values such as those you compare with in this lab are usually CCSD(T)/CBS estimates.

The optional Colab notebook of Experiment 1 lets you try an extrapolation on the HCl dimer yourself.

2.6.3 The price: cost grows steeply with K

Better results cost more. The number of two-electron integrals (Equation 15) grows as K^4: with permutational symmetry there are about $K^4/8$ unique ones. For water, that is about 41,000 with def2-SVP ($K = 24$), 430,000 with def2-TZVP (43), and 23 million with def2-QZVP (117). Going from double- to quadruple-ζ multiplies the formal cost by about 565 (Figure 5 b).

This K^4 is the *formal* scaling of Hartree–Fock and hybrid DFT. In practice, programs do better: integrals between functions far apart are negligible and are skipped (screening), and approximations such as density

fitting reduce the work. For small molecules, fixed overheads (starting the program, reading files) also blur the picture. The *measured* exponent is therefore lower than 4, and it depends on the method and the size of the system. In Experiment 1 you will measure it yourself from your own wall times.

> ! The 10-minute limit is a basis-set limit
>
> On the free IQmol server every job must finish in 10 minutes. For the molecules in this lab, the basis set is usually what decides whether a job fits. A one-character typo, such as def2-QZVP instead of def2-SVP, can multiply the cost by hundreds and get your job killed.

2.6.4 Two basis-set errors: BSIE and BSSE

- **Basis-set incompleteness error (BSIE).** The difference between a result in a finite basis and the CBS limit, with the same method. It affects every calculation. It shrinks as the basis grows, and it is what Experiment 1 measures for a reaction energy.
- **Basis-set superposition error (BSSE).** A specific artifact of incomplete basis sets in systems made of fragments, such as a dimer. In the dimer, each monomer can "borrow" the basis functions centered on its partner to improve its own description, lowering its energy. The isolated monomers have no partner functions to borrow. The dimer is therefore artificially stabilized, and the interaction energy is too large in magnitude. BSSE also vanishes in the CBS limit. The standard correction, the counterpoise method of Boys and Bernardi [26], recomputes each monomer in the full dimer basis. We return to BSSE in Lectures 8 and 10.

> Practical rules of thumb
>
> - Double-ζ (def2-SVP, cc-pVDZ) is fine for quick tests and geometries of large systems, but rarely converged for energies.
> - Triple-ζ (def2-TZVP) is the usual compromise between accuracy and cost for DFT, and it is the default basis of Experiments 3–5 and 9.
> - Quadruple-ζ is close to the CBS limit for HF and DFT, and expensive.
> - Add diffuse functions for anions and when weak interactions matter most.
> - Report the basis set exactly, including stars, plus signs and prefixes: "6-31G" and "6-31G*" are different basis sets.

2.7 Key points

- Molecular orbitals are expanded in a finite set of K atom-centered basis functions, $\phi_i = \sum_\mu c_{\mu i}\chi_\mu$. Only the coefficients are optimized.
- For HF and DFT, a bigger basis gives a lower (or equal) energy; the limit is the complete-basis-set (CBS) limit, and the difference is the basis-set incompleteness error.
- Slater functions have the correct cusp and tail but expensive integrals; Gaussians have neither, but the Gaussian product theorem makes their integrals fast. Contracted Gaussians combine several primitives to imitate one Slater function (STO-3G).
- Cartesian d shells have 6 components and spherical ones 5 (f: 10 vs 7). The convention changes K; check it in the output.
- Split valence (more ζ) lets orbitals change size; polarization (higher ℓ) lets them change shape; diffuse functions (small α) extend their reach.

- Pople names encode contractions (6-31G*, 6-311+G**). Pople sets are historical and not designed to converge systematically.
- def2-SVP/TZVP/QZVP cover almost the whole periodic table, with ECPs beyond Kr. cc-pVnZ (n = D, T, Q, 5) converges systematically for correlated methods; aug- adds diffuse functions. pc-n is optimized for DFT.
- The number of basis functions follows from the shell composition: for water, 24 (def2-SVP, cc-pVDZ), 43 (def2-TZVP), 58 (cc-pVTZ), 117 (def2-QZVP), 115 (cc-pVQZ).
- The formal cost of HF and hybrid DFT grows as K^4; the measured exponent is lower. HF/DFT energies converge roughly exponentially with n; correlation energies as n^{-3}.
- BSSE makes complexes artificially too stable in small basis sets; it vanishes at the CBS limit.

2.8 Study questions

1. Explain in your own words why the energy cannot increase when functions are added to a basis set (for HF or DFT). Why is this guarantee weaker when comparing def2-SVP and def2-TZVP, and why does it not apply to reaction energies?
2. List two physical shortcomings of Gaussian functions compared with Slater functions, and the one mathematical property that makes Gaussians the standard anyway.
3. Using Equation 17, find the center P and the exponent p of the product of an s Gaussian with $\alpha = 2.0$ at $x = 0$ and one with $\beta = 0.5$ at $x = 1.0$ bohr. Which original Gaussian is P closer to, and why?
4. What is the difference between a primitive and a contracted Gaussian? In STO-3G, how many primitives and how many contracted functions does a hydrogen atom have? How many coefficients $c_{\mu i}$ does the SCF optimize for an H atom?
5. Decode each name: 3-21G, 6-31+G**, 6-311++G(d,p), def2-TZVPP, aug-cc-pVQZ. For each, say whether it has polarization functions on H, and whether it has diffuse functions.
6. Count the basis functions of HCl in def2-SVP from the shell compositions H [2s1p] and Cl [4s3p1d], with spherical functions. Then predict K for the HCl dimer. (Check your answer against your first Experiment 1 output.)
7. Count the basis functions of methane (CH_4) in cc-pVDZ, cc-pVTZ and def2-TZVP. Carbon has the same shell composition as oxygen in Table 7.
8. A water calculation reports 25 basis functions with 6-31G**. Another program reports 24 for the same basis. Explain the difference.
9. If the time of a calculation grew exactly as K^4, how much longer would a water job take with def2-QZVP than with def2-TZVP? Give two reasons why the real ratio is smaller.
10. Why does BSSE make the binding energy of a dimer too large in a small basis set, rather than too small? Why does it disappear at the CBS limit?

2.9 Further reading

- F. Jensen, "Atomic orbital basis sets" [27], and B. Nagy and F. Jensen, "Basis sets in quantum chemistry" [28]: modern, readable reviews of everything in this chapter.
- A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* [6], §3.6: the classic derivation of STO-nG and the minimal-basis H_2 calculation.
- T. Helgaker, P. Jørgensen and J. Olsen, *Molecular Electronic-Structure Theory* [7], chapters 6 and 8: Gaussian functions, basis-set design and convergence in depth.
- F. Weigend and R. Ahlrichs [17]: the original def2 paper, with tables of the shell compositions for every element.
- The Basis Set Exchange lists the exponents and coefficients of nearly every published basis set; you can look up any of the sets in Table 7 there.

3 From theory to a calculation: IQmol, Q-Chem and Experiment 1

Learning goals

After studying this chapter you should be able to:

- decode a model chemistry and translate it into the `$rem` section of a Q-Chem input;
- write a complete Q-Chem single-point input by hand, including charge and multiplicity;
- set up, submit and retrieve a job with IQmol on the free Q-Chem server, and use the experiment pages to generate inputs and track jobs;
- find the total energy, the number of SCF cycles, the wall time and the number of basis functions in an output, and check that the job finished correctly;
- compute a reaction energy as $\Delta E = \sum_i \nu_i E_i$ with signed coefficients, convert it to kcal/mol, and compute signed and absolute errors;
- explain what Experiment 1 measures, why, and what you have to do and submit.

Before this lecture

This chapter is a companion to the lab guides, not a replacement. Have them open:

- Guide 1 · Install IQmol and connect to Q-Chem: do this **before** class.
- Guide 2 · Your first Q-Chem job
- Guide 3 · Reading a Q-Chem output
- Guide 4 · Working on a shared machine
- Guide 5 · Units, reaction energies and errors

3.1 Recap: what a model chemistry is

Lecture 1 showed that every calculation makes two approximations: the **method** approximates the Hamiltonian (how electrons interact), and the **basis set** approximates the wave function (which shapes the orbitals may take). Lecture 2 opened the second one. A pair of choices, written *method/basis*, is a **model chemistry**:

Table 8: Three model chemistries you will run in Experiment 1. The method column will make full sense after Lectures 4–6; for now, treat the names as labels.

Model chemistry	Method	Rung (Lectures 4–6)	Basis set	ζ	Family
HF/cc-pVTZ	Hartree–Fock	no correlation at all	cc-pVTZ	triple	Dunning, correlation-consistent
B3LYP/def2-SVP	B3LYP	hybrid GGA	def2-SVP	double	Karlsruhe def2
ωB97M-V/def2-QZVP	ωB97M-V	range-separated hybrid meta-GGA with VV10 dispersion	def2-QZVP	quadruple	Karlsruhe def2

Two more choices complete the definition of a calculation, and you must also report them: the **geometry** (here always given to you, as an `.xyz` file) and the **electronic state**, set by the total **charge** and the **spin multiplicity**. Every number you compute in this lab is a total energy for one model chemistry, one geometry and one electronic state.

3.2 The tools

The workflow of every experiment is the same (Figure 6).

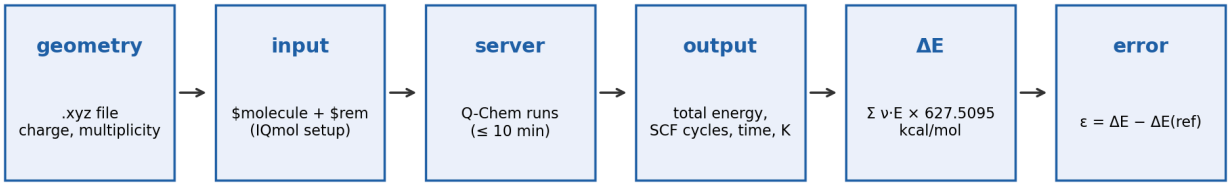


Figure 6: The workflow of every job in this lab. You prepare the geometry and input on your computer, Q-Chem runs on a remote server, and you extract four numbers from the output to compute reaction energies and their errors.

3.2.1 Q-Chem and IQmol

Q-Chem [1] is the quantum-chemistry program that actually solves the equations. It reads a plain-text **input file** and writes a plain-text **output file**. You will not install it: it runs on a remote server.

IQmol [29] is the free graphical program you install on your laptop (Guide 1). It does four jobs for you:

1. **Builder and viewer.** It opens `.xyz` files (**File → Open**) and shows the molecule in 3D. You can also build molecules, though in this lab all geometries are provided.
2. **Q-Chem setup window. Calculation → Q-Chem Setup** opens a window where you choose the job type, method, basis, charge and multiplicity, and which shows the text of the input file it will send. You can also paste a ready-made input there.
3. **Submission to a server.** In the **Server** menu of the setup window you select the free Q-Chem server and click **Submit**.
4. **Job Monitor.** It shows each job as *queued* → *running* → *finished* and lets you open or save the output when the job is done.

> ! The free server and its 10-minute limit
>
> The free Q-Chem server is shared by many users (Guide 4). Each job is stopped after **10 minutes** of run time, and a killed job gives no energy. All jobs in Experiment 1 are small and well within the limit. If a job is killed, suspect a typo in the input first, for example a much larger basis than intended. Start early: queues are longest right before deadlines.

3.2.2 Anatomy of a Q-Chem input

A Q-Chem input is a set of sections, each starting with `$name` and ending with `$end`. For a single-point energy you need two. Here is the test job of Guide 2, with the geometry of the HCl molecule from Experiment 1:

```
$molecule
0 1
Cl     0.00000000     0.00000000     0.00000000
H      0.00000000     0.00000000     1.27907000
$end

$rem
   JOBTYPE              SP
   METHOD               HF
   BASIS                def2-SVP
   SCF_CONVERGENCE      8
$end
```

`$molecule` describes the system:

- the first line holds the **total charge** and the **spin multiplicity** $2S + 1$, here `0 1` (neutral singlet);
- each following line holds an element symbol and its x, y, z coordinates **in ångström** (Q-Chem converts to bohr internally).

`$rem` holds the instructions, one keyword and one value per line:

Table 9: The four `$rem` keywords used in Experiment 1.

Keyword	Value here	Meaning
`JOBTYPE`	`SP`	single-point energy at the given geometry (no optimization, no frequencies)
`METHOD`	`HF`	the method: `HF`, or a functional name such as `B3LYP`, `PBE`, `wB97M-V`
`BASIS`	`def2-SVP`	the basis set
`SCF_CONVERGENCE`	`8`	a tight convergence threshold for the SCF iterations, 10^{-8} (Lecture 4 explains what is converged)

A model chemistry maps directly onto two lines: *method/basis* becomes `METHOD` and `BASIS`. To run the same molecule with B3LYP/def2-TZVP, you change only those two values. Everything else stays.

i Charge and multiplicity: line 2 of every .xyz file

Standard `.xyz` files use their second line as a free comment. In this lab, line 2 always holds `charge multiplicity`, so you never have to guess. Copy it into the first line of `$molecule`. Recall from Lecture 1: 1 = singlet (all electrons paired, like HCl), 2 = doublet (one unpaired electron, like an H atom or a radical), 3 = triplet (two parallel unpaired electrons, like ground-state O_2). All the species of Experiment 1 are closed-shell singlets. A wrong multiplicity gives a perfectly converged energy for the wrong electronic state.

3.2.3 The experiment pages do the typing for you

You will run hundreds of jobs this semester. Typing inputs by hand is slow and error-prone, so each experiment page has a **Run the calculations** section that generates them from the experiment's data:

- a **job checklist** with one entry per job, named systematically (for example `ex01_hf_def2-svp_hcl`), with a rough cost flag (*quick*, *a few minutes*, *may hit the 10-min limit*);
- a **Copy input** button for each job, to paste into IQmol's input editor, replacing what is there;
- an **All inputs (.zip)** button to download every input at once;
- the molecule gallery, with the 3D structures and a button to download all geometries.

The checklist ticks itself off as you enter results, so you can see at a glance what is missing. Use the generated inputs: they are already consistent with the reference calculations, and they avoid typos.

3.3 A first job, step by step

In class we run the test job of Guide 2 live: HCl with HF/def2-SVP. Repeat it on your own before starting Experiment 1. The steps are:

1. Download the HCl geometry from the molecule gallery of Experiment 1 and open it in IQmol (**File → Open**).
2. Open **Calculation → Q-Chem Setup**. Either fill in the fields (Energy, HF, def2-SVP, charge 0, multiplicity 1), or paste the input from the experiment page.
3. Choose the free server in the **Server** menu, give the job a systematic name, and click **Submit**.
4. Follow the job in the **Job Monitor**. When it has finished, open the output file and save it in a folder for Experiment 1.
5. Find the total energy (next section). It should be close to −460 E_h. If it is, everything works.

Then try the most important experiment of the semester in miniature: **change only `BASIS`** (for example to def2-TZVP), submit again, and compare. The total energy changes, and so do the number of basis functions and the time. Everything in Experiment 1 is a systematic version of this one change. Do not draw conclusions from a single molecule, though: what matters is how the *reaction* energy changes, and that needs both species of the reaction computed at the same level.

> Good habits from day one
>
> - Run **one** job first, check its output, then submit the rest in batches.
> - Keep every output file, one folder per experiment, with systematic names.
> - Never resubmit a failed job blindly: read the end of its output first.

3.4 Reading the output

An output file is long, but you need only a few lines. Search for the phrases below with your editor's find function (Guide 3 has a fully annotated example):

Table 10: Where to find the numbers in a Q-Chem output (example values from Guide 3, not from Experiment 1).

You need	Search for	Example	Record
Number of basis functions K	`basis functions`	`There are 26 shells and 74 basis functions`	74
SCF cycles	`Convergence criterion met`	the cycle number at the start of that line	12
Total energy (E_h)	`Total energy in the final basis set`	`Total energy in the final basis set = -920.6431907520`	all digits
Wall time (s)	`Total job time`	`Total job time: 14.52s(wall), 52.10s(cpu)`	14.52
Did it finish?	`Thank you very much for using Q-Chem`	last lines of the file	yes/no

What each number means:

- **Number of basis functions.** K of Lecture 2. Check it against what you expect from the shell compositions; it confirms that the right basis was used. It is also the N in the cost scaling you will measure.
- **SCF cycles.** Hartree–Fock and Kohn–Sham DFT are solved iteratively (Lecture 4). The table of cycles shows the energy settling down; the line ending in *Convergence criterion met* is the last one. If instead you see a message that the SCF failed to converge, the energy is meaningless.
- **Total energy.** Electronic energy plus nuclear repulsion at the fixed input geometry, in hartree (Lecture 1): no zero-point energy, no thermal corrections. If the output also prints an "SCF energy", always use the **total** energy: with extra terms such as a dispersion correction (-D4), the two can differ.
- **Wall time vs CPU time.** *Wall* time is what a stopwatch on the wall would measure. *CPU* time adds up the time of every processor core, so on a multi-core server it is larger. We always compare **wall** times, and only between jobs run on the same server. Timings are noisy: a busy server or a slow disk adds seconds that have nothing to do with the chemistry.
- **"Thank you very much for using Q-Chem. Have a nice day."** The last line of a finished job. If it is missing, the job crashed or was killed, and nothing in the file should be trusted.

! Five checks before you trust a number

1. The job **finished** (the "Thank you" line is there).
2. The SCF **converged** (*Convergence criterion met*).
3. The **charge and multiplicity** echoed near the top are right.
4. The **method and basis** echoed in `$rem` are the ones you intended.
5. The **number of basis functions** is what you expect for that basis.

3.5 From total energies to a reaction energy

3.5.1 Signed stoichiometric coefficients

For any reaction, the reaction energy is a single sum over **all** the species i that take part:

$$\Delta E = \sum_i \nu_i E_i, \tag{22}$$

where E_i is the total energy of species i and ν_i is its **signed** stoichiometric coefficient: $\nu_i > 0$ for products and $\nu_i < 0$ for reactants. The minus sign of "products minus reactants" is already inside the coefficients, so you never subtract separately. This is the convention of the whole course, and the results forms check your ΔE with exactly this sum (Guide 5).

For Experiment 1, the reaction is the dissociation of the HCl dimer:

$$\begin{gathered}(\mathrm{HCl})_2 \rightarrow 2\,\mathrm{HCl}, \qquad \nu_{\mathrm{dimer}} = -1, \quad \nu_{\mathrm{HCl}} = +2, \\ \Delta E = (+2)\,E(\mathrm{HCl}) + (-1)\,E\left((\mathrm{HCl})_2\right) .\end{gathered} \tag{23}$$

A positive ΔE means the dimer is lower in energy than the separated molecules: energy must be supplied to pull them apart.

Total energies come in hartree. Convert the reaction energy to kcal/mol at the end, with

$$1\ E_{\mathrm{h}} = 627.5095\ \mathrm{kcal/mol}\ .$$

3.5.2 Keep all the digits

A reaction energy is a small difference between large numbers. In Experiment 1 the total energies are hundreds of hartree and the reference reaction energy is 2.01 kcal/mol, which is only 0.0032 E_{h}. Rounding each total energy to 4 decimals can introduce errors of almost 0.1 kcal/mol (Lecture 1). Copy **every digit** Q-Chem prints, and round only the final reaction energy, to two decimals in kcal/mol.

3.5.3 Signed and absolute errors

To judge a method, compare the computed reaction energy with a reliable **reference** value (for the reactions in this lab, high-level wave-function results from benchmark databases):

$$\varepsilon = \Delta E_{\mathrm{comp}} - \Delta E_{\mathrm{ref}}. \tag{24}$$

The **signed error** ε tells you the direction: positive means the method *overestimates* the reaction energy. The **absolute error** $|\varepsilon|$ tells you how far off it is, regardless of direction. Over several reactions, the mean signed error (MSE) reveals systematic bias and the mean absolute error (MAE) measures overall accuracy; Lecture 7 returns to these statistics. An error of about 1 kcal/mol is traditionally called **chemical accuracy**.

Worked example: a reaction energy and its error

Consider the water dimer, $(H_2O)_2 \rightarrow 2\,H_2O$, with **hypothetical** total energies $E(H_2O) = -76.4203184\ E_h$ and $E\left((H_2O)_2\right) = -152.8487052\ E_h$, and a reference of 5.00 kcal/mol.

1. Coefficients: $\nu(H_2O) = +2$, $\nu(\text{dimer}) = -1$.
2. Sum: $\Delta E = 2(-76.4203184) + (-1)(-152.8487052) = -152.8406368 + 152.8487052 = 0.0080684\ E_h$.
3. Convert: $0.0080684 \times 627.5095 = 5.06$ kcal/mol.
4. Error: $\varepsilon = 5.06 - 5.00 = +0.06$ kcal/mol (slight overestimate); $|\varepsilon| = 0.06$ kcal/mol.

Now repeat step 2 with the total energies rounded to 3 decimals, -76.420 and -152.849: $\Delta E = -152.840 + 152.849 = 0.009\ E_h = 5.65$ kcal/mol. The rounding alone created an error ten times larger than the method's.

! Common mistakes with reaction energies

- **Forgetting a coefficient.** The dimer dissociates into *two* monomers: $E(HCl)$ must be multiplied by 2.
- **Sign flip.** Writing reactants minus products gives the right magnitude and the wrong sign. With signed ν this cannot happen.
- **Mixing levels of theory.** Every species in one reaction energy must be computed with the *same* method and basis. Combining a def2-SVP dimer with a def2-TZVP monomer gives a meaningless number.
- **Forgetting the conversion**, or dividing instead of multiplying by 627.5095.

3.6 Experiment 1 briefing

Everything above comes together in Experiment 1: Basis Set Incompleteness Error. Read the experiment page in full before you start; this section explains what you will do and why it matters.

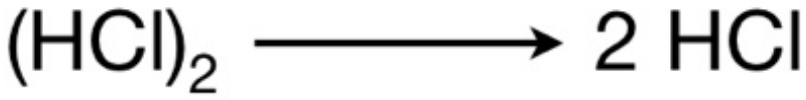


Figure 7: The reaction of Experiment 1: dissociation of the HCl dimer into two HCl molecules.

3.6.1 The question

Lecture 2 claimed that larger basis sets give results that converge toward the CBS limit, at a steeply growing cost. Experiment 1 tests that claim on a real reaction energy, for four different methods, with two families of basis sets. The reaction is the dissociation of the HCl dimer, a weakly bound complex held together by noncovalent interactions, from the NCCE30 subset [30] of the Minnesota 2015B database [31]. Its reference energy is **2.01 kcal/mol**. Weakly bound complexes are a demanding test for basis sets: the interaction energy is small, so even small basis-set errors are large in relative terms.

3.6.2 What you will run

Table 11: The Experiment 1 design. Each method/basis pair gives one reaction energy, so you obtain 24 reaction energies.

	Choices	Count
Species	$(HCl)_2$ and HCl, both neutral singlets (`0 1`)	2
Methods	HF, PBE (GGA), B3LYP (hybrid GGA), ωB97M-V (range-separated hybrid meta-GGA)	4
Basis sets	def2-SVP, def2-TZVP, def2-QZVP (Karlsruhe); cc-pVDZ, cc-pVTZ, cc-pVQZ (Dunning)	6
Jobs	2 species × 4 methods × 6 basis sets	**48**

All jobs are single points (`JOBTYPE SP`) with `SCF_CONVERGENCE 8`, and all are short. The two basis-set families are matched by ζ: double (SVP, DZ), triple (TZVP, TZ) and quadruple (QZVP, QZ). The methods span the range from no electron correlation at all (HF) to a modern, heavily parametrized functional (ωB97M-V); Lectures 4–6 explain what each one is.

For each job, record from the output the **total energy**, the **number of SCF cycles**, the **wall time** and the **number of basis functions**. Then compute each reaction energy yourself with Equation 23, convert it to kcal/mol, and enter it. The results form recomputes ΔE from the total energies **you** entered and flags inconsistencies (a sign, a missing factor of 2, a unit slip). It never shows you the correct value.

3.6.3 What you will look at

Once your table is complete, the page draws two charts from your own data:

- **Error vs ζ.** The signed error of each method as the basis grows from double- to quadruple-ζ, for each family. This is BSIE made visible: does the error shrink steadily, for every method? Do the two families behave the same way at the same ζ?
- **Cost scaling.** Wall time against the number of basis functions on log–log axes, with a straight-line fit for each method. The slope is the exponent n in time $\propto K^n$. Lecture 2 said the formal scaling of HF and hybrid DFT is K^4, and that practice is usually kinder. You will measure your own n, and decide which points to trust (very short jobs are dominated by fixed overhead).

The questions at the end of the page ask you to describe the trend with ζ for each method, compare the two families, classify each method's behavior, report and interpret your measured exponent, and choose the best combination of method and basis for this reaction, considering both accuracy and cost.

! Lock your predictions first

Before running anything, answer the warm-up quiz and **lock your predictions** in the *Predict first* section of the experiment page. You will predict, among other things, how the error of ωB97M-V and of Hartree–Fock changes from double- to quadruple-ζ, which family gives smaller errors at the same ζ, and which method comes closest to the reference with def2-QZVP. Predictions are not graded on being right: they are part of your submission so you (and we) can see how your intuition changes. Think about them using Lectures 1 and 2, commit, and then let your data answer. Once locked, they cannot be changed.

3.6.4 How to submit

1. Work through the page in order: warm-up quiz, predictions, calculations, results, charts, questions. Your entries are saved only in your own browser, so download your file often.
2. In the *Submit* section, enter your name and your **Florida Tech email** (ending in `@my.fit.edu` or `@fit.edu`), and confirm that you ran the calculations yourself.
3. Download your submission file (a `.json` file).
4. Open the Microsoft Form linked on the page, sign in with your Florida Tech email, and upload the file. You can resubmit until the deadline; the latest upload counts.

The optional Colab notebook at the bottom of the page extrapolates the HCl dimer to the CBS limit with PySCF, using the ideas of Lecture 2.

3.7 Before next lecture

- Install IQmol and run the test job (Guides 1–2) if you have not yet.
- Lock your Experiment 1 predictions, then run the first batch of jobs: for example, all 12 jobs of one method, or all 8 jobs of one basis set. Check one output thoroughly with the five checks above before submitting the rest.
- Lecture 4 opens the second approximation, the Hamiltonian: Hartree–Fock and the self-consistent field. You will find out what the SCF cycles you just counted are actually doing, and why the "HF" row of your Experiment 1 table matters so much.

3.8 Key points

- A calculation is defined by a model chemistry (method/basis), a geometry and an electronic state (charge and multiplicity). Report all of them.
- A Q-Chem single-point input has `$molecule` (charge, multiplicity, coordinates in Å) and `$rem` (`JOBTYPE`, `METHOD`, `BASIS`, `SCF_CONVERGENCE`). Method/basis maps onto `METHOD` and `BASIS`.
- Line 2 of every `.xyz` file in this lab is "charge multiplicity".
- IQmol opens the geometry, writes and submits the input to the free server (10-minute limit), and monitors the job. The experiment pages generate every input and keep a job checklist.
- From each output take the total energy (all digits), the SCF cycle count, the wall time and the number of basis functions; check that the job finished and the SCF converged.
- $\Delta E = \sum_i \nu_i E_i$ with $\nu > 0$ for products and $\nu < 0$ for reactants; multiply by 627.5095 to get kcal/mol. Never round total energies.
- Signed error $\varepsilon = \Delta E_{\text{comp}} - \Delta E_{\text{ref}}$; absolute error $|\varepsilon|$.
- Experiment 1: $(HCl)_2 \longrightarrow 2\ HCl$, reference 2.01 kcal/mol; 4 methods × 6 basis sets × 2 species = 48 jobs; lock predictions first; submit the JSON file through the Microsoft Form with your Florida Tech email.

3.9 Study questions

1. Write the complete Q-Chem input for a single-point B3LYP/def2-TZVP energy of the HCl molecule of Experiment 1. Which lines change if you switch to ωB97M-V/cc-pVQZ?
2. The second line of an `.xyz` file reads `0 2`. What does it tell you about the species? Give an example of a molecule or atom that could have it.
3. An output ends with SCF cycles and no "Thank you very much for using Q-Chem" line. List two possible causes and what you would check.
4. Explain the difference between wall time and CPU time. A job reports 10.0 s (wall) and 38.6 s (cpu). Roughly how many cores did it use effectively?
5. Write $\Delta E = \sum_i \nu_i E_i$ explicitly, with signed coefficients, for: (a) $2\ \mathrm{O_3} \longrightarrow 3\ \mathrm{O_2}$; (b) the atomization of water, $\mathrm{H_2O} \longrightarrow 2\ \mathrm{H} + \mathrm{O}$; (c) the Experiment 1 reaction.
6. With hypothetical energies $E(\mathrm{HCl}) = -460.1000000\ E_\mathrm{h}$ and $E\left((\mathrm{HCl})_2\right) = -920.2040000\ E_\mathrm{h}$, compute ΔE in kcal/mol and its signed and absolute errors with respect to 2.01 kcal/mol. Is the dimer bound in this hypothetical calculation?
7. A classmate computes the Experiment 1 reaction energy as $E\left((\mathrm{HCl})_2\right) - E(\mathrm{HCl})$ and gets about −460 E_h. What went wrong?
8. Why must both species of a reaction be computed with the same method *and* the same basis set?
9. If the wall time grew as K^4, and a job with $K = 50$ took 4 s, how long would a job with $K = 200$ take? Why might a log–log fit of your Experiment 1 timings give a smaller exponent?
10. Without running anything: how do you expect the *total* energy of HCl with HF to change from def2-SVP to def2-QZVP, and why? Does that tell you how the *reaction* energy will change?

3.10 Further reading

- The lab guides: Guide 2 (first job), Guide 3 (reading output) and Guide 5 (units and errors).
- Experiment 1 and its references, in particular the comparison of basis-set families for DFT by Witte, Neaton and Head-Gordon [18].
- E. Epifanovsky et al., the Q-Chem 5 overview [1], for what the program can do beyond this lab.
- P. Morgante and R. Peverati, "The devil in the details" [32]: a tutorial review of often-overlooked practical aspects of DFT calculations, several of which this lab explores one by one.

4 Approximating the Hamiltonian I: Hartree–Fock and the SCF

Learning goals

After studying this chapter you should be able to:

- explain the **mean-field** idea and why it makes the many-electron problem solvable;
- write the energy of a Slater determinant and say what the one-electron, Coulomb and exchange terms mean;
- write the **Fock operator** and explain why exchange has no classical analogue and why Hartree–Fock is free of self-interaction;
- write the **Roothaan–Hall equations** $\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon}$ and describe every step of the **SCF loop**, including what the DIIS error in a Q-Chem output means;
- tell restricted (RHF) from unrestricted (UHF) calculations and know when each is used;
- use **Koopmans' theorem** to estimate an ionization energy, and know its limits;
- define the **correlation energy**, tell dynamic from static correlation, and explain why it matters for reaction energies even though it is only about 1% of the total energy.

Review from P-Chem 2

This chapter builds on the instructor's P-Chem 2 textbook. Review these sections first if they feel rusty:

- Many-electron atoms, §10.1–10.2: many-electron wave functions, the variational method, the helium atom.
- Postulate 6: the Pauli exclusion principle and Spin.
- Solving the electronic eigenvalue problem, §11.3 and the Hartree–Fock method, §11.3.1: LCAO, the secular determinant, the idea of a self-consistent field.
- Hartree–Fock calculation for water, §13.2: a complete minimal-basis calculation, with the energy printed at every SCF cycle.

In P-Chem 2 you saw the *recipe*. Here we open the box and look at the equations behind it.

4.1 Where we are

Every calculation in this course makes two approximations. In Lecture 1 you wrote the electronic Hamiltonian (atomic units, nuclei clamped):

$$\hat{H}_{\mathrm{el}} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A}\frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{25}$$

The three terms are the kinetic energy of the electrons, their attraction to the nuclei, and the electron–electron repulsion. The last term couples every electron to every other, and that is why no analytic solution exists for $N \geq 2$.

In Lecture 2 you approximated the **wave function's flexibility** with a basis set: orbitals are written as a finite sum of basis functions χ_μ. That is the first approximation. This lecture and the next two are about the second one: how we approximate the **Hamiltonian**, or more precisely, how we treat the electron–electron repulsion. Together, the two choices make a *model chemistry*, written method/basis, such as HF/def2-TZVP or B3LYP/def2-SVP.

Hartree–Fock (HF) is the starting point for everything else. It is a method in its own right (you are running it in Experiment 1), it is the reference for the correlated wave-function methods that produce the benchmark numbers of this lab, and its machinery (orbitals, the Fock matrix, the SCF loop) is reused almost unchanged by density functional theory in Lecture 5.

> **i** Experiment 1 check-in
>
> By now you should have submitted your first jobs for Experiment 1. If anything did not work (the server, IQmol, a job that never finished), check the guides on setting up IQmol, your first job and reading the output, and ask in class. We will discuss the results of Experiment 1 in Lecture 7, after everyone has submitted.

4.2 The mean-field idea

4.2.1 Hartree's average field

In 1928 Douglas Hartree proposed a simple way out of the many-electron problem [33]. Instead of following every electron as it dodges every other electron, let each electron move in the **average** electrostatic field created by the nuclei and by the *charge cloud* of all the other electrons. Each electron then obeys its own one-electron Schrödinger equation, with an effective potential that depends on where the other electrons are on average.

This is a **mean-field** approximation. It replaces the instantaneous repulsion $1/|\mathbf{r}_i - \mathbf{r}_j|$ by an interaction with a smeared-out charge density. The catch is that the average field depends on the orbitals of the other electrons, which are themselves the solutions we are looking for. Hartree solved the problem iteratively: guess the orbitals, compute the field, solve for new orbitals, compute a new field, and repeat until the orbitals that come out produce the same field that went in. He called the result a **self-consistent field** (SCF). The name survives in every quantum chemistry program, including the "SCF" lines of your Q-Chem outputs.

Hartree wrote the many-electron wave function as a simple product of orbitals, $\Psi = \varphi_1(\mathbf{r}_1)\varphi_2(\mathbf{r}_2)\cdots\varphi_N(\mathbf{r}_N)$. A product like this is not antisymmetric. It violates the Pauli principle, and it misses an important energy term, as we will see.

4.2.2 The Slater determinant

In 1930 Fock [34] and Slater [5] fixed this by using a determinant of **spin orbitals** instead of a product. A spin orbital $\phi_i(\mathbf{x})$ is a spatial orbital times a spin function, $\phi_i(\mathbf{x}) = \varphi_i(\mathbf{r})\,\sigma(\omega)$ with $\sigma = \alpha$ or β, and $\mathbf{x} = (\mathbf{r}, \omega)$ collects the space and spin coordinates of one electron. The **Slater determinant** is

$$\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{x}_1) & \phi_2(\mathbf{x}_1) & \cdots & \phi_N(\mathbf{x}_1) \\ \phi_1(\mathbf{x}_2) & \phi_2(\mathbf{x}_2) & \cdots & \phi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}_N) & \phi_2(\mathbf{x}_N) & \cdots & \phi_N(\mathbf{x}_N) \end{vmatrix}. \tag{26}$$

Swapping two electrons swaps two rows, which changes the sign of the determinant: the wave function is antisymmetric, as the Pauli principle requires. Putting two electrons in the same spin orbital makes two columns equal, and the determinant vanishes: no two electrons can share all their quantum numbers. You met this determinant in Lecture 1 and in P-Chem 2. We will write it compactly as $\Psi = |\phi_1\phi_2\cdots\phi_N|$.

The Hartree–Fock approximation is a single statement: *the wave function is one Slater determinant, and the best determinant is the one with the lowest energy.* "Best" comes from the variational principle, $E[\Psi_{\text{trial}}] \geq E_0$, which you used for basis sets in Lecture 2. Everything else in this chapter follows from that statement.

4.2.3 The energy of a Slater determinant

Take the expectation value of the Hamiltonian (Equation 25) with a normalized determinant of orthonormal spin orbitals. The algebra is long but standard (see A. Szabo and N. S. Ostlund [6], chapter 2). The result is short:

$$E_{\text{HF}} = \sum_{i=1}^{N} h_i + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(J_{ij} - K_{ij}) + V_{nn}. \tag{27}$$

Here V_{nn} is the constant nuclear repulsion, and the three kinds of integrals are:

- the **one-electron (core) energy** of spin orbital i, its kinetic energy plus its attraction to all nuclei:

$$h_i = \int \phi_i^*(\mathbf{x}) \left[-\frac{1}{2}\nabla^2 - \sum_A \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right] \phi_i(\mathbf{x})\, d\mathbf{x};$$

- the **Coulomb integral**, the classical repulsion between the charge clouds $|\phi_i|^2$ and $|\phi_j|^2$:

$$J_{ij} = \iint \frac{|\phi_i(\mathbf{x}_1)|^2\, |\phi_j(\mathbf{x}_2)|^2}{r_{12}}\, d\mathbf{x}_1\, d\mathbf{x}_2;$$

- the **exchange integral**, which has the two orbitals swapped between the electrons on one side:

$$K_{ij} = \iint \frac{\phi_i^*(\mathbf{x}_1)\phi_j(\mathbf{x}_1)\, \phi_j^*(\mathbf{x}_2)\phi_i(\mathbf{x}_2)}{r_{12}}\, d\mathbf{x}_1\, d\mathbf{x}_2.$$

The first two terms are what you would write for classical charge clouds. The third is new. It comes entirely from the antisymmetry of the determinant: a simple Hartree product gives only h and J. Because the spin parts must match on both sides of K_{ij}, the exchange integral is zero unless electrons i and j have the **same spin**. Exchange is always positive ($K_{ij} \geq 0$), so it always lowers the energy.

> **i** Exchange has no classical analogue
>
> Exchange is not a new force. It is a correction to the Coulomb repulsion that appears because two electrons with the same spin cannot be at the same place (the determinant vanishes when $\mathbf{x}_1 = \mathbf{x}_2$). Around each electron there is a region, the **Fermi hole** or exchange hole, where same-spin electrons are unlikely to be found. Less overlap means less repulsion, and $-K_{ij}$ is that energy saving. Nothing in classical electrostatics gives this term. It is a pure consequence of the Pauli principle.

! Remember this: no self-interaction in Hartree–Fock

Look at the double sum in Equation 27. It includes the terms with $i = j$. The term J_{ii} is the repulsion of electron i with *its own* charge cloud, which is unphysical: an electron does not repel itself. But from the definitions, $K_{ii} = J_{ii}$, so the two cancel exactly: $J_{ii} - K_{ii} = 0$. Hartree–Fock is **free of self-interaction**. Hartree's original product method was not, and, as you will see in Lecture 5, approximate density functionals are not either. This single fact explains why mixing some Hartree–Fock exchange into a functional helps (Lecture 6).

4.2.4 The Hartree–Fock equations and the Fock operator

We want the spin orbitals that minimize Equation 27 while staying orthonormal. This is a constrained minimization, done with Lagrange multipliers. After a unitary rotation that makes the multiplier matrix diagonal (the *canonical* orbitals), the condition for a minimum becomes a set of one-electron eigenvalue equations, the **Hartree–Fock equations**:

$$\hat{f}\,\phi_i = \varepsilon_i\,\phi_i, \qquad i = 1, \ldots, N. \tag{28}$$

The **Fock operator** $\hat{f}$ is an effective one-electron Hamiltonian:

$$\hat{f}(\mathbf{x}_1) = \hat{h}(\mathbf{x}_1) + \sum_{j=1}^{N}\left[\hat{J}_j(\mathbf{x}_1) - \hat{K}_j(\mathbf{x}_1)\right]. \tag{29}$$

- $\hat{h} = -\tfrac{1}{2}\nabla^2 - \sum_A Z_A/|\mathbf{r} - \mathbf{R}_A|$ is the **core Hamiltonian**: kinetic energy plus nuclear attraction of one electron.
- $\hat{J}_j$ is the **Coulomb operator**: it multiplies the orbital by the electrostatic potential of the charge cloud of electron j,

$$\hat{J}_j(\mathbf{x}_1)\,\phi_i(\mathbf{x}_1) = \left[\int \frac{|\phi_j(\mathbf{x}_2)|^2}{r_{12}}\,d\mathbf{x}_2\right]\phi_i(\mathbf{x}_1).$$

- $\hat{K}_j$ is the **exchange operator**: it swaps the orbitals inside the integral,

$$\hat{K}_j(\mathbf{x}_1)\,\phi_i(\mathbf{x}_1) = \left[\int \frac{\phi_j^*(\mathbf{x}_2)\,\phi_i(\mathbf{x}_2)}{r_{12}}\,d\mathbf{x}_2\right]\phi_j(\mathbf{x}_1).$$

  $\hat{K}_j$ is a **nonlocal** operator: the result at point $\mathbf{x}_1$ depends on the value of ϕ_i everywhere in space, not only at $\mathbf{x}_1$. Keep this in mind. Density functional approximations replace this nonlocal operator by a *local* potential, and that is the root of several of their errors.

The sum $\sum_j(\hat{J}_j - \hat{K}_j)$ is the **Hartree–Fock potential**: the average field that each electron feels from all the others. This is the mean-field idea in equation form. The term $j = i$ is included in the sum, but, as above, its Coulomb and exchange parts cancel when acting on ϕ_i.

Each eigenvalue ε_i is an **orbital energy**. It equals $h_i + \sum_j(J_{ij} - K_{ij})$: the kinetic and nuclear-attraction energy of electron i plus its average repulsion with all other electrons. Note that the total energy is **not** the sum of the orbital energies. Summing ε_i over all occupied orbitals counts every electron pair twice, so

$$E_{\mathrm{HF}} = \sum_i \varepsilon_i - \frac{1}{2}\sum_{i,j}\left(J_{ij} - K_{ij}\right) + V_{nn}.$$

This is exactly the bookkeeping check in §13.2 of P-Chem 2: twice the sum of the occupied orbital energies of water, plus the electron–electron repulsion counted once with the opposite sign, plus the nuclear repulsion, gives back the total energy.

4.2.5 Closed shells: restricted Hartree–Fock

Most of the molecules in this course have an even number of electrons, all paired. For these **closed-shell** molecules we let each spatial orbital φ_a hold two electrons, one α and one β. This is **restricted Hartree–Fock** (RHF). Doing the spin integration once and for all, the Fock operator for the $N/2$ doubly occupied spatial orbitals becomes

$$\hat{f}(\mathbf{r}) = \hat{h}(\mathbf{r}) + \sum_{a=1}^{N/2}\Big[2\hat{J}_a(\mathbf{r}) - \hat{K}_a(\mathbf{r})\Big], \tag{30}$$

where the factor 2 counts the two electrons in each orbital, and the exchange term has no factor 2 because exchange acts only between electrons of the same spin. We come back to open shells in Section 4.3.4.

4.3 The Roothaan–Hall equations and the SCF loop

4.3.1 From integro-differential equations to matrices

Equations (Equation 28) are coupled integro-differential equations. For atoms they can be solved numerically on a radial grid, as Hartree did. For molecules this is impractical. In 1951 Roothaan [35] and, independently, Hall [36] used the LCAO idea from Lecture 2: expand each spatial orbital in a finite basis of K functions,

$$\varphi_i(\mathbf{r}) = \sum_{\mu=1}^{K} c_{\mu i}\,\chi_\mu(\mathbf{r}). \tag{31}$$

Insert this into $\hat{f}\varphi_i = \varepsilon_i\varphi_i$, multiply from the left by χ_ν^* and integrate. The unknown functions become unknown coefficients, and the equations become a matrix problem, the **Roothaan–Hall equations**:

$$\mathbf{F}\mathbf{C} = \mathbf{S}\mathbf{C}\boldsymbol{\varepsilon}. \tag{32}$$

All matrices are $K \times K$:

- $\mathbf{C}$ holds the MO coefficients: column i is orbital φ_i.
- $\boldsymbol{\varepsilon}$ is diagonal, with the orbital energies ε_i.
- $\mathbf{S}$ is the **overlap matrix**, $S_{\mu\nu} = \int \chi_\mu^*\chi_\nu\, d\mathbf{r}$. It is not the identity matrix because basis functions on different atoms overlap (the same point is made in P-Chem 2 §11.3.1). This is why Equation 32 is a *generalized* eigenvalue problem.
- $\mathbf{F}$ is the **Fock matrix**, $F_{\mu\nu} = \int \chi_\mu^*\hat{f}\chi_\nu\, d\mathbf{r}$. For RHF it is

$$F_{\mu\nu} = H_{\mu\nu}^{\mathrm{core}} + \sum_{\lambda=1}^{K}\sum_{\sigma=1}^{K} P_{\lambda\sigma}\left[(\mu\nu \mid \lambda\sigma) - \frac{1}{2}(\mu\lambda \mid \nu\sigma)\right], \tag{33}$$

with the core-Hamiltonian matrix $H_{\mu\nu}^{\mathrm{core}} = \int \chi_\mu^*\hat{h}\chi_\nu\, d\mathbf{r}$, the **two-electron integrals**

$$(\mu\nu|\lambda\sigma) = \iint \frac{\chi_\mu^*(\mathbf{r}_1)\chi_\nu(\mathbf{r}_1)\,\chi_\lambda^*(\mathbf{r}_2)\chi_\sigma(\mathbf{r}_2)}{r_{12}}\, d\mathbf{r}_1\, d\mathbf{r}_2,$$

and the **density matrix**

$$P_{\lambda\sigma} = 2\sum_{a=1}^{N/2} c_{\lambda a}\, c_{\sigma a}^{*}. \tag{34}$$

The density matrix is the bridge between orbitals and electron density: $\rho(\mathbf{r}) = \sum_{\lambda\sigma} P_{\lambda\sigma}\chi_\lambda(\mathbf{r})\chi_\sigma^*(\mathbf{r})$. In Equation 33 the first bracket term is the Coulomb part and the second, with the swapped indices, is the exchange part. The notation in P-Chem 2 is slightly different ($h = K + V$, G for the two-electron part, V_{ijkl} for the integrals), but the equation is the same.

Once the equations are solved, the energy is

$$E_{\mathrm{RHF}} = \frac{1}{2}\sum_{\mu\nu} P_{\nu\mu}\big(H_{\mu\nu}^{\mathrm{core}} + F_{\mu\nu}\big) + V_{nn}. \tag{35}$$

> **i** Where the cost comes from
>
> There are about $K^4/8$ unique two-electron integrals $(\mu\nu|\lambda\sigma)$. With 7 basis functions (water in a minimal basis, as in P-Chem 2 §13.2) this is a few hundred. With the 200 or so functions of a medium-sized molecule in def2-TZVP it is about 2×10^8. Computing and contracting these integrals with $\mathbf{P}$ is the formal K^4 (often written N^4) cost of Hartree–Fock that you met in Lecture 2 and are measuring in Experiment 1. In practice programs skip integrals that are negligibly small (screening) or approximate them (density fitting), so the observed scaling is lower.

4.3.2 Why it must be solved iteratively

Equation 32 looks like an ordinary eigenvalue problem, but it is not linear: the Fock matrix depends on the density matrix (Equation 33), which depends on the coefficients (Equation 34), which are the solution of the equation. We need $\mathbf{C}$ to build $\mathbf{F}$, and $\mathbf{F}$ to find $\mathbf{C}$. The way out is Hartree's: iterate until the input and output agree. This is the **SCF procedure** (Figure 8).

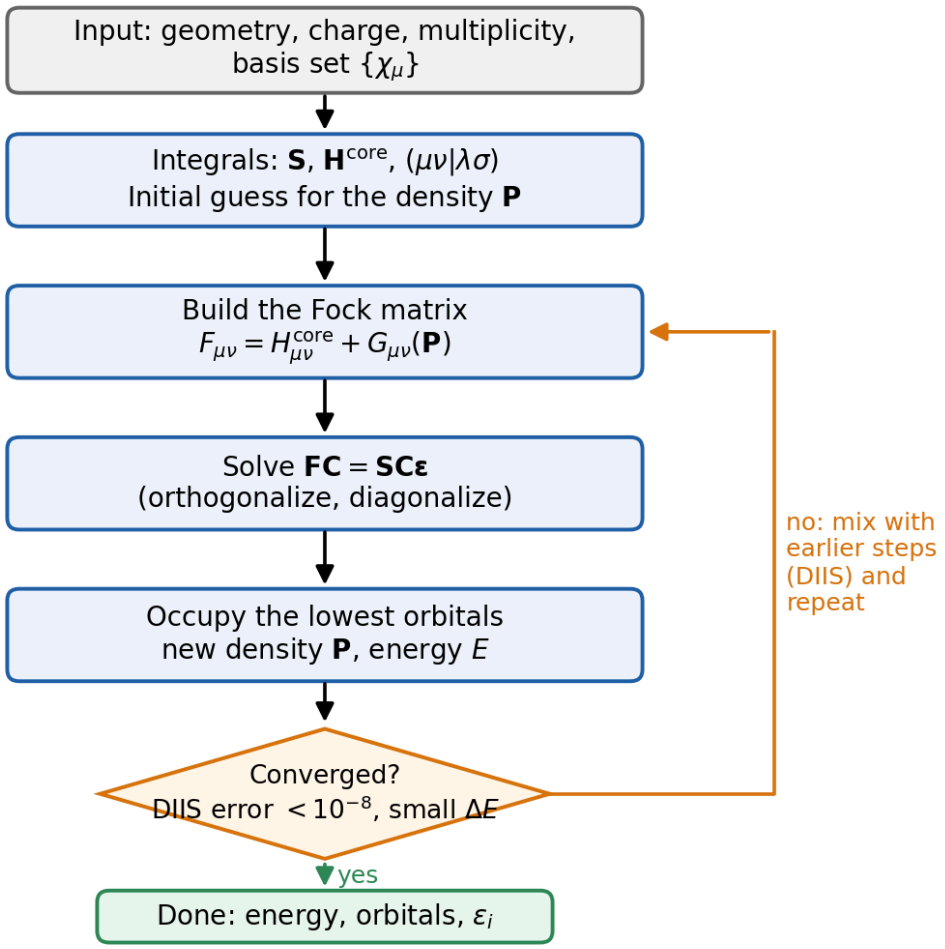


Figure 8: The self-consistent field (SCF) loop used by Hartree–Fock and by Kohn–Sham DFT. Each pass through the loop is one line ("cycle") of the SCF table in a Q-Chem output.

Step by step:

1. **Input.** Geometry, charge, multiplicity and basis set: exactly the `$molecule` and `$rem` sections of a Q-Chem input.
2. **One-time work.** Compute $\mathbf{S}$ and $\mathbf{H}^{\text{core}}$ (cheap), and decide how to handle the two-electron integrals (store them, or recompute them each cycle, which modern programs usually do).
3. **Initial guess.** Make a first density matrix $\mathbf{P}$. A crude choice is to diagonalize $\mathbf{H}^{\text{core}}$ alone (no electron repulsion at all). A much better one, and the default in Q-Chem, is a superposition of precomputed atomic densities (SAD). A good guess saves cycles and makes it more likely that the SCF lands on the right solution.
4. **Build the Fock matrix** from the current $\mathbf{P}$ with Equation 33. This is the expensive step.
5. **Solve $\mathbf{FC} = \mathbf{SC}\varepsilon$.** First transform to an orthonormal basis (for example with $\mathbf{S}^{-1/2}$), so the problem becomes an ordinary eigenvalue problem $\mathbf{F}'\mathbf{C}' = \mathbf{C}'\varepsilon$. Diagonalize, then transform the coefficients back.
6. **Occupy** the $N/2$ orbitals with the lowest ε_i (the *aufbau* principle), build the new $\mathbf{P}$ with Equation 34, and compute the energy with Equation 35.
7. **Check convergence.** If the new density is (almost) the same as the old one, stop. If not, go back to step 4 with the new density.

4.3.3 Convergence, DIIS and the SCF table in your output

What does "the same" mean in step 7? At self-consistency the Fock matrix and the density matrix commute in the metric of the overlap matrix:

$$\mathbf{FPS} - \mathbf{SPF} = \mathbf{0}. \tag{36}$$

Away from convergence this commutator is not zero, and its size is a natural error measure. Pulay's **DIIS** method (direct inversion in the iterative subspace) [37] uses it twice. First, the largest element of the commutator is the **DIIS error** that Q-Chem prints in the third column of the SCF table. Second, instead of using only the newest Fock matrix, DIIS builds the next one as a linear combination of the Fock matrices from the last several cycles, with coefficients chosen to make the combined error as small as possible. This extrapolation is what makes modern SCF calculations converge in 10–20 cycles instead of hundreds.

Here is what this looks like in a Q-Chem output (see the guide on reading the output):

```
---------------------------------------
 Cycle       Energy         DIIS error
---------------------------------------
   1    -920.4153225231      4.31e-02
   2    -920.6017766913      7.45e-03
   ...
  11    -920.6431907518      1.12e-07
  12    -920.6431907520      4.05e-09  Convergence criterion met
---------------------------------------
```

Each line is one pass through Figure 8. The energy goes down and settles. The DIIS error shrinks by orders of magnitude. The job stops when the DIIS error falls below 10^{-n}, where n is the value of `SCF_CONVERGENCE` in the `$rem` section. All experiments in this course use `SCF_CONVERGENCE 8`, a tight threshold: the energy is then converged far beyond the digits you need for a reaction energy, because the energy error is roughly the *square* of the density error. That is a consequence of the variational principle: at the minimum, first-order errors in the orbitals give only second-order errors in the energy.

⚠ When the SCF goes wrong

The SCF is a numerical procedure, and it can fail in three ways. Learn to recognize them now. You will meet all three later in the course.

- **Oscillation.** The energy jumps up and down from cycle to cycle and never settles. This happens when two or more orbitals are close in energy near the HOMO–LUMO gap and swap occupation back and forth. Typical for transition-metal compounds and stretched bonds.
- **Slow convergence.** The DIIS error decreases, but very slowly, and the job hits the maximum number of cycles (Q-Chem prints a warning that the SCF failed to converge). The energy of a job that did not converge is meaningless. Never use it.
- **Convergence to the wrong solution.** The SCF converges perfectly, but to a stationary point that is not the lowest-energy solution: an excited configuration, or a saddle point that is a minimum only within the constraints you imposed (for example, RHF for a molecule that wants unpaired electrons). The output looks normal. Only a **stability analysis** can tell. This is the subject of Lecture 12 and Experiment 8.

Programs have many tools against the first two: better initial guesses, damping, level shifting, and second-order algorithms such as geometric direct minimization. You will not need them for the organic molecules of Experiments 1–7.

4.3.4 Restricted and unrestricted Hartree–Fock

Molecules with unpaired electrons (radicals, the O and S atoms, triplet states, many transition-metal complexes) cannot use Equation 30, because their α and β electrons are not paired up in the same spatial orbitals. There are two common ways to handle them:

- **Unrestricted Hartree–Fock (UHF)** [38] gives the α and β electrons their own sets of spatial orbitals, $\{\varphi_i^\alpha\}$ and $\left\{\varphi_i^\beta\right\}$. There are two Fock matrices and two density matrices, coupled through the Coulomb term (the α electrons repel the β electrons and vice versa). In a basis these are the Pople–Nesbet equations, $\mathbf{F}^\alpha\mathbf{C}^\alpha = \mathbf{S}\mathbf{C}^\alpha\boldsymbol{\varepsilon}^\alpha$ and the same for β. UHF has more freedom than RHF, so its energy is lower or equal. The price is that the UHF determinant is not, in general, an eigenfunction of $\hat{S}^2$: it is contaminated by higher spin states. Q-Chem prints $\langle\hat{S}^2\rangle$ for unrestricted jobs so you can check (a pure doublet has 0.75, a pure triplet 2.0).
- **Restricted open-shell Hartree–Fock (ROHF)** keeps doubly occupied orbitals for the paired electrons and singly occupied orbitals for the unpaired ones. It gives a pure spin state, at a slightly higher energy than UHF.

In this course, open-shell species are run **unrestricted** (Q-Chem does this automatically when the multiplicity is larger than 1). The same choice exists in DFT, where we speak of restricted and unrestricted Kohn–Sham. You will use unrestricted calculations for atoms and radicals in Experiment 5 and later.

💡 Multiplicity reminder

The multiplicity is $2S + 1$, where S is the total spin: 1 for a closed shell (singlet), 2 for one unpaired electron (doublet), 3 for two parallel unpaired electrons (triplet). It is the second number on line 2 of every `.xyz` file of this lab ("charge multiplicity"). A triplet O atom run as a singlet gives a perfectly converged, perfectly wrong energy.

4.4 Orbital energies and Koopmans' theorem

Diagonalizing the Fock matrix gives K orbitals, but only the lowest $N/2$ (for RHF) are **occupied**. The remaining $K - N/2$ are **virtual** (unoccupied) orbitals. The highest occupied one is the **HOMO** and the lowest unoccupied one the **LUMO**. The number of virtual orbitals grows with the basis set, while the number of occupied orbitals is fixed by the number of electrons.

Q-Chem prints the orbital energies after the SCF table, under a heading "Orbital Energies (a.u.)", split into "Occupied" and "Virtual" lists (for unrestricted jobs, separately for α and β). Occupied orbital energies are negative: these electrons are bound. Their pattern follows chemical intuition: very low core levels (about -20.5 Eh for the O 1s orbital in water), then valence levels between about -1.5 and -0.3 Eh. Compare the minimal-basis water numbers in P-Chem 2 §13.2.

Koopmans' theorem [39] gives the occupied orbital energies a physical meaning. If we remove an electron from orbital i and keep all the other orbitals frozen, the energy of the resulting cation, minus the energy of the neutral molecule, is exactly $-\varepsilon_i$. So the first ionization energy is approximately

$$\mathrm{IE} \approx -\varepsilon_{\mathrm{HOMO}}. \tag{37}$$

Two errors are hidden in this approximation. **Orbital relaxation**: in the real cation the remaining electrons rearrange to the new field, which lowers the cation energy, so Koopmans' value is too large. **Correlation**: the neutral molecule has one more electron pair to correlate than the cation, so including correlation increases the true ionization energy, and Koopmans' value is too small. The two errors have opposite signs and partly cancel, so Koopmans IEs of closed-shell molecules are often within about 1 eV of experiment. The cancellation is fortunate, not guaranteed. The analogous estimate of the electron affinity from the LUMO, $\mathrm{EA} \approx -\varepsilon_{\mathrm{LUMO}}$, is much worse, because the two errors now add up.

> **i** Worked example: water
>
> An HF calculation on water with a large basis gives $\varepsilon_{\mathrm{HOMO}} \approx -0.507$ Eh. Then $\mathrm{IE} \approx 0.507 \times 27.211 \approx 13.8$ eV. The experimental vertical ionization energy of water is 12.6 eV. Koopmans' estimate is about 1.2 eV too high, as expected when relaxation wins over correlation. Try it yourself: run water with HF and a triple-ζ basis and look for the last "Occupied" number in the output. Your value will depend slightly on the geometry and basis.

> **What the orbital pictures in IQmol show**
>
> IQmol can plot any orbital from a Q-Chem job as an isosurface (a surface of constant $|\varphi_i|$, with two colors for the two signs). These pictures are useful for chemical intuition: the HOMO of ethene is the C=C π orbital, the LUMO is the π*. Keep two things in mind. First, the canonical orbitals of Equation 28 are delocalized over the whole molecule and do not look like Lewis bonds and lone pairs (water's two "lone pairs" are two quite different orbitals). Any unitary rotation of the occupied orbitals gives the *same* determinant and the same energy, so localized orbitals are equally valid. Second, the virtual orbitals depend strongly on the basis set, and adding diffuse functions changes them a lot.

4.5 What Hartree–Fock misses: electron correlation

4.5.1 The correlation energy

The Hartree–Fock wave function is the best single determinant, but it is still only one determinant. In the mean-field picture, each electron sees the *average* position of the others. Real electrons react to each other's *instantaneous* positions: they avoid each other more than the average field suggests. Their motions are **correlated**. Exchange already keeps same-spin electrons apart (the Fermi hole), but nothing in HF keeps opposite-spin electrons apart.

Löwdin defined the **correlation energy** as the difference between the exact energy and the Hartree–Fock energy [40]:

$$E_{\mathrm{corr}} = E_{\mathrm{exact}} - E_{\mathrm{HF}}. \tag{38}$$

Because of the variational principle, E_{corr} is always negative. Two details matter. "Exact" means the exact nonrelativistic Born–Oppenheimer energy. And both energies should be in the same basis set, or, ideally, both at the complete-basis-set (CBS) limit. Otherwise the basis-set error gets mixed into the correlation energy.

How big is it? For the helium atom, $E_{\mathrm{HF}} = -2.8617$ Eh at the basis-set limit, while the exact energy is -2.9037 Eh. The correlation energy is -0.042 Eh, or -26 kcal/mol: about 1.4% of the total energy. For the water molecule it is roughly -0.37 Eh (about -230 kcal/mol), about 0.5% of the total. As a rule of thumb, the correlation energy is around 1% of the total energy, or very roughly 25 kcal/mol per electron pair.

One percent sounds small. It is not. **Chemistry happens in energy differences.** A typical reaction energy is 1–100 kcal/mol, and the conformational energies of Experiment 3 are 1–3 kcal/mol. The total energy of a small organic molecule is about 100,000–500,000 kcal/mol. If the correlation energy changed by just 1% between reactants and products, it would shift a reaction energy by a few kcal/mol. When bonds are broken or formed, or when the number of electron pairs changes, the change in correlation energy is much larger than that.

> **i** Hartree–Fock has no dispersion at all
>
> London dispersion, the attraction between two neutral, nonpolar molecules (two argon atoms, two methane molecules, the stacked layers of graphite), comes from correlated, instantaneous fluctuations of the electron clouds. A mean-field method cannot describe it: in HF each molecule sees only the *static* average charge of the other. At long range, HF gives no dispersion attraction at all. Weakly bound complexes held together mainly by dispersion are unbound, or far too weakly bound, at the HF level, no matter how large the basis set. Keep this in mind when you look at your Experiment 1 results. Dispersion returns in Lecture 8.

4.5.2 Dynamic and static correlation

It is useful to split correlation into two kinds, although there is no sharp boundary between them.

Dynamic correlation is the short-range dodging of electrons: each electron creates a small "Coulomb hole" around itself where other electrons, of either spin, are less likely to be. It is present in every molecule with more than one electron. Describing it requires mixing in many determinants, each with a small weight. It converges slowly with the basis set, because the wave function has a cusp where two electrons meet.

Static (nondynamic) correlation appears when one determinant is not even qualitatively right: when two or more electron configurations have nearly the same energy and must be mixed with comparable weights. Typical examples are stretched or breaking bonds, diradicals, ozone, and many transition-metal compounds.

The textbook example is the dissociation of H_2 (Figure 9). Near the equilibrium distance, one doubly occupied bonding orbital σ_g is a good description, and RHF is qualitatively right. The minimum is a little too shallow, and the difference from the exact curve is the dynamic correlation energy. But as the bond is stretched, the RHF determinant $|\sigma_g\alpha\,\sigma_g\beta|$ still forces both electrons into the same spatial orbital. Expanded in atomic orbitals, it is 50% covalent (H· + ·H) and 50% ionic (H^+ + H^-), at any distance. At infinite separation it should be 100% covalent. The ionic half costs a lot of energy, so the RHF curve rises far above the energy of two H atoms. The exact wave function fixes this by mixing in a second determinant, $|\sigma_u\alpha\,\sigma_u\beta|$, with a weight that grows until both have equal weight at dissociation. That second determinant is the static correlation.

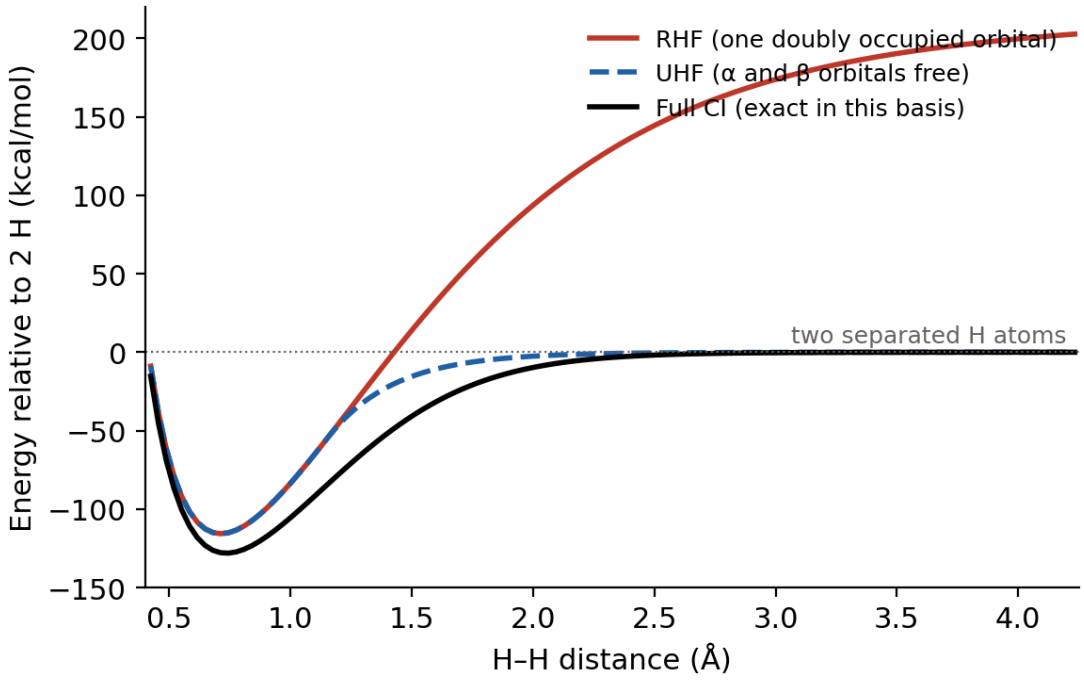


Figure 9: Dissociation of H_2 computed with a minimal basis set (STO-3G), where the full configuration interaction (full CI) curve is the exact answer within this basis. RHF is good near equilibrium but goes to an energy far above two H atoms. UHF breaks the spin symmetry beyond about 1.2 Å and dissociates correctly, at the cost of a spin-contaminated wave function. The gap between RHF and full CI at the minimum is the dynamic correlation energy; the growing gap at long distance is static correlation. Computed with `images/src/l04_l06_figures.py`.

UHF also "fixes" the dissociation: past a certain distance (the Coulson–Fischer point) the α electron localizes on one atom and the β electron on the other, and the energy goes correctly to two H atoms. But the UHF wave function is then a 50/50 mixture of a singlet and a triplet, and the curve has a kink. UHF solutions that are lower than RHF are an example of an **instability** of the RHF solution: the topic of Lecture 12.

⚠ Common mistake: "a converged SCF means the method is right"

A converged SCF tells you that the equations were solved, not that the approximation behind them is good. RHF for H_2 at 3 Å converges without trouble to a number that is 200 kcal/mol too high. Convergence is about numerics. Accuracy is about the method.

4.6 Ways forward

There are two broad families of methods that go beyond Hartree–Fock.

Correlated wave-function methods keep the HF determinant as a starting point (the *reference*) and add excited determinants to it.

- **Møller–Plesset perturbation theory** treats the difference between the real Hamiltonian and the sum of Fock operators as a perturbation [41]. Its second-order version, **MP2**, recovers a large part of the dynamic correlation and describes dispersion, at a cost that scales as N^5.
- **Coupled-cluster theory** includes excitations to all orders in an exponential form. **CCSD(T)** (singles, doubles and a perturbative estimate of triples) [42] is called the "gold standard" of quantum chemistry: for molecules without strong static correlation, CCSD(T) near the CBS limit is typically within about 1 kcal/mol of experiment for reaction energies. Its cost scales as N^7: doubling the size of the molecule makes the calculation about 128 times slower. The reference values of most reactions in this lab come from CCSD(T)/CBS or similar high-level calculations collected in benchmark databases [43], as you will see in Lecture 7.
- **Multireference methods** (CASSCF, MRCI, CASPT2) start from several determinants and are needed when static correlation is strong. They are outside the scope of this course.

Density functional theory takes a different path. It keeps a one-determinant, mean-field machinery, with the same SCF loop and the same N^3–N^4 cost as HF, but it changes the *Hamiltonian* of that machinery so that it includes, approximately, both exchange and correlation. How that is possible, and why the hard part is a single term called the exchange–correlation energy, is the subject of Lecture 5.

Table 12: Hartree–Fock and the methods that go beyond it. "Formal cost" is how the time grows with system size N for a fixed basis-set quality.

Method	Correlation	Formal cost	Dispersion	Typical use
HF	none	N^4	none	reference, starting point
DFT (GGA, hybrid)	approximate, through E_{xc}	N^3–N^4	only if added (-D4, VV10) or fitted	everyday chemistry
MP2	dynamic, second order	N^5	yes (often overestimated)	medium molecules
CCSD(T)	dynamic, nearly complete	N^7	yes	benchmarks, small molecules

Key points

- Hartree–Fock approximates the wave function by **one Slater determinant** and finds the best one with the variational principle. Each electron moves in the **average field** of the others.

- The HF energy contains one-electron terms, classical **Coulomb** terms J_{ij}, and **exchange** terms K_{ij} between same-spin electrons. Exchange comes from antisymmetry and has no classical analogue.
- Because $J_{ii} = K_{ii}$, HF is **free of self-interaction**. Approximate density functionals are not.
- The **Fock operator** is $\hat{f} = \hat{h} + \sum_j (\hat{J}_j - \hat{K}_j)$. In a basis, the HF equations become the **Roothaan–Hall equations** $\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon}$.
- **F** depends on its own solution through the density matrix, so the equations are solved **iteratively** (the SCF loop). DIIS accelerates the loop, and the **DIIS error** in the output measures how far from self-consistency you are. `SCF_CONVERGENCE 8` means a DIIS error below 10^{-8}.
- The SCF can oscillate, converge slowly, or converge to the **wrong solution**. A converged SCF is not proof of a correct answer.
- RHF pairs electrons in the same spatial orbitals; **UHF** gives α and β electrons their own orbitals and is used for open shells.
- **Koopmans' theorem**: IE $\approx -\varepsilon_{\mathrm{HOMO}}$, thanks to a partial cancellation of relaxation and correlation errors.
- The **correlation energy** $E_{\mathrm{exact}} - E_{\mathrm{HF}}$ is about 1% of the total energy but as large as or larger than most reaction energies. It has **dynamic** and **static** parts. HF has **no dispersion**.
- Correlated wave-function methods (MP2, CCSD(T)) are accurate but scale as N^5 to N^7. DFT is the cheaper alternative.

Study questions

1. Write the Hartree product and the Slater determinant for a two-electron system. Show that only the determinant changes sign when the two electrons are exchanged, and that it vanishes if both electrons occupy the same spin orbital.
2. In Equation 27, the double sum runs over all i and j, including $i = j$. Why is this harmless in Hartree–Fock? What would go wrong if the exchange term were approximated by something that does not exactly cancel J_{ii}?
3. Explain in words why $K_{ij} = 0$ when electrons i and j have opposite spins. What does this imply about how HF treats opposite-spin electrons?
4. The Roothaan–Hall equations look like an eigenvalue problem, but they must be solved iteratively. Explain why, in two or three sentences, using the words "Fock matrix" and "density matrix".
5. In the SCF table of a Q-Chem output, what is printed in the "DIIS error" column, and what does the job need to see there to stop with `SCF_CONVERGENCE 8`? Why can the energy be converged to more digits than the density?
6. A colleague's job printed a final energy, but no "Convergence criterion met" line. Can the energy be used in a reaction energy? Why or why not?
7. Use Koopmans' theorem to estimate the ionization energy (in eV) of a molecule with $\varepsilon_{\mathrm{HOMO}} = -0.40$ Eh. Name the two approximations hidden in this estimate and the direction of each error.
8. Using Figure 9, explain why RHF fails to dissociate H_2 into two H atoms. What does the exact wave function contain that the RHF wave function lacks? Is this dynamic or static correlation?
9. The correlation energy of a molecule is "only" about 1% of its total energy. Explain why it is still essential for computing reaction energies, using an order-of-magnitude argument.
10. Argon dimer is bound by about 0.3 kcal/mol. What would an HF calculation with a very large basis set predict for it, and why?

Further reading

- P-Chem 2 textbook: Introduction to Molecules, §11.3 and Hartree–Fock calculation for water, §13.2.

- A. Szabo and N. S. Ostlund [6], chapters 2–3: the classic, step-by-step derivation of everything in this chapter, including a complete minimal-basis H_2 calculation like the one in Figure 9.
- T. Helgaker, P. Jørgensen, and J. Olsen [7]: the advanced reference on wave-function methods, for those who want the full theory.
- P. Pulay [37]: the original DIIS paper, short and readable.
- W. Koch and M. C. Holthausen [8], chapter 1: a compact review of Hartree–Fock and correlation written as a lead-in to DFT.

5 Approximating the Hamiltonian II: density functional theory

Learning goals

After studying this chapter you should be able to:

- define the electron density $\rho(\mathbf{r})$, state its main properties, and explain why it is a much simpler object than the wave function;
- state the two **Hohenberg–Kohn theorems** and explain what they do and do not promise;
- explain why **Thomas–Fermi** theory, which uses only the density, fails for chemistry;
- describe the **Kohn–Sham** scheme: the non-interacting reference system, the energy decomposition $E = T_s + E_{ne} + J + E_{xc}$, and the Kohn–Sham equations;
- explain what is inside the **exchange–correlation energy** E_{xc} and why everything hard about DFT is in that one term;
- write the **LSDA** exchange energy and explain where LSDA correlation comes from;
- list what LSDA gets right and wrong, and name the three main error types of approximate functionals.

Review from P-Chem 2 and from Lecture 4

- The probability density $|\Psi|^2$ and the Born interpretation: Postulate 1.
- Radial distribution functions of the hydrogen atom: The hydrogen atom, §5.1. The density of the 1s orbital, $\rho(r) = e^{-2r}/\pi$, is used in a worked example below.
- The variational method: Many-electron atoms, §10.2.1.
- From Lecture 4: the Fock operator, Coulomb and exchange, self-interaction, the SCF loop, and the definition of the correlation energy.

5.1 Why look beyond the wave function?

The many-electron wave function $\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ depends on $3N$ spatial coordinates (plus N spin coordinates). For benzene, with 42 electrons, that is a function of 126 variables. Suppose you wanted to store it on a coarse grid of only 10 points per coordinate: you would need 10^{126} numbers, far more than the number of atoms in the observable universe. Walter Kohn called this growth the "exponential wall" of wave-function methods [44].

Correlated wave-function methods avoid storing Ψ on a grid, but they pay in a different currency: computer time. As you saw in Lecture 4, MP2 scales as N^5 and CCSD(T) as N^7 with system size. CCSD(T) on a molecule with 30 heavy atoms is a major undertaking, and on the 50–100 atoms of a typical catalyst or drug molecule it is out of reach for routine work.

Yet most questions in chemistry are about one simple function: where the electrons are. The **electron density** $\rho(\mathbf{r})$ depends on only **three** coordinates, whatever the number of electrons. Density functional theory (DFT) is the theory that says this function is enough, in principle, to get the exact ground-state energy. In practice it is the method behind most of the quantum chemistry done today, and the method of almost every experiment in this course.

5.2 The electron density

5.2.1 Definition and properties

The electron density is the probability of finding *any* electron at point $\mathbf{r}$, with any spin, whatever the other electrons are doing:

$$\rho(\mathbf{r}_1) = N \int \cdots \int |\Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)|^2 d\omega_1 \, d\mathbf{x}_2 \cdots d\mathbf{x}_N. \tag{39}$$

We integrate $|\Psi|^2$ over the spin of electron 1 and over all coordinates of the other $N-1$ electrons. The factor N appears because the electrons are indistinguishable: any of them could be the one at $\mathbf{r}_1$. Its main properties are:

- **It integrates to the number of electrons**: $\int \rho(\mathbf{r})\, d\mathbf{r} = N$.
- **It is positive** everywhere and decays to zero far from the molecule. The decay is exponential, $\rho \propto e^{-2\sqrt{2I}\,r}$, where I is the first ionization energy (in atomic units).
- **It has cusps at the nuclei.** At each nucleus the density has a sharp peak whose slope is fixed by the nuclear charge (Kato's cusp condition). So the density "knows" where the nuclei are and what their charges are.
- **It is an observable.** X-ray diffraction measures the electron density of crystals, and electron-density maps are how crystallographers see molecules.
- For a single determinant (Hartree–Fock or Kohn–Sham), it is simply the sum of the occupied orbital densities: $\rho(\mathbf{r}) = \sum_i^{\text{occ}} |\varphi_i(\mathbf{r})|^2$ (over spin orbitals).

For open-shell systems we also use the **spin densities** $\rho_\alpha(\mathbf{r})$ and $\rho_\beta(\mathbf{r})$, with $\rho = \rho_\alpha + \rho_\beta$. Their difference, the spin density $\rho_\alpha - \rho_\beta$, shows where the unpaired electrons are.

5.2.2 The cusp argument: a hint that the density is enough

Here is an informal argument, due to E. Bright Wilson, that the density should contain everything. From ρ you can read off N (integrate it), the positions of the nuclei (the cusps), and their charges (the slope at each cusp). But N, the nuclear positions and the nuclear charges are all you need to write the electronic Hamiltonian (Lecture 1). With the Hamiltonian you can, in principle, solve for Ψ and everything else. So the ground-state density determines the ground-state wave function. The argument works only for Coulomb potentials. The general proof came in 1964.

5.3 The Hohenberg–Kohn theorems

In the electronic Hamiltonian, the only part that depends on the molecule is the attraction to the nuclei. We collect it in the **external potential** $v_{\text{ext}}(\mathbf{r}) = -\sum_A Z_A/|\mathbf{r} - \mathbf{R}_A|$. The kinetic energy $\hat{T}$ and the electron–electron repulsion $\hat{V}_{ee}$ are the same for every molecule with N electrons. In 1964 Pierre Hohenberg and Walter Kohn proved two theorems about this situation [45].

First theorem (existence). *The ground-state density determines the external potential, up to an additive constant.* Since v_{ext} and N fix the Hamiltonian, the density determines the Hamiltonian, and therefore the ground-state wave function, the energy, and every other ground-state property.

The proof is short. Suppose two different external potentials v and v' (with Hamiltonians $\hat{H}$ and $\hat{H}'$, and nondegenerate ground states Ψ and Ψ') gave the same ground-state density ρ. Use Ψ' as a trial function for $\hat{H}$. By the variational principle,

$$E_0 < \langle \Psi' | \hat{H} | \Psi' \rangle = \langle \Psi' | \hat{H}' | \Psi' \rangle + \langle \Psi' | \hat{H} - \hat{H}' | \Psi' \rangle = E_0' + \int [v(\mathbf{r}) - v'(\mathbf{r})] \rho(\mathbf{r})\, d\mathbf{r}.$$

Swapping the roles of the two systems gives $E_0' < E_0 + \int [v' - v]\rho \, d\mathbf{r}$. Adding the two inequalities gives $E_0 + E_0' < E_0' + E_0$, which is impossible. So two different potentials cannot give the same ground-state density.

Second theorem (variational principle). The ground-state energy is a **functional** of the density,

$$E_v[\rho] = F[\rho] + \int v_{\text{ext}}(\mathbf{r})\, \rho(\mathbf{r})\, d\mathbf{r}, \tag{40}$$

and it is minimized by the exact ground-state density. For any trial density $\tilde{\rho}$ that is positive and integrates to N (with some technical conditions), $E_v[\tilde{\rho}] \geq E_0$.

A *functional* is a rule that takes a whole function and returns a number, just as $E[\Psi] = \langle \Psi | \hat{H} | \Psi \rangle$ takes a wave function and returns an energy. We write functionals with square brackets.

The term $F[\rho] = T[\rho] + V_{ee}[\rho]$ is the **universal functional**: the kinetic and electron–electron energy of the electrons. "Universal" means it is the same for every atom, molecule or solid. Only v_{ext} changes from one system to another. Levy later gave a cleaner definition that removes most of the technical conditions: $F[\rho]$ is the lowest value of $\langle \Psi | \hat{T} + \hat{V}_{ee} | \Psi \rangle$ over all wave functions that give the density ρ [46].

> ⚠ What the theorems do not say
>
> The Hohenberg–Kohn theorems are existence proofs. They say that $F[\rho]$ exists, not what it is. Nobody knows the exact $F[\rho]$, and it is very likely that nobody ever will, in a form that is cheap to evaluate. All of practical DFT is about approximating it. The theorems are about the **ground state**: excited states need extensions (time-dependent DFT), which we do not cover. And "DFT is exact" is true only for the exact functional. Every calculation you will run uses an approximation.

5.4 A cautionary tale: Thomas–Fermi theory

The idea of using only the density is older than the theorems. In 1927, Thomas [47] and Fermi [48] wrote the energy of an atom as a functional of the density alone. They took the electrostatic terms exactly,

$$E_{ne}[\rho] = \int v_{\text{ext}}(\mathbf{r})\, \rho(\mathbf{r})\, d\mathbf{r}, \qquad J[\rho] = \frac{1}{2} \iint \frac{\rho(\mathbf{r})\, \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}\, d\mathbf{r}', \tag{41}$$

and approximated the kinetic energy by that of a **uniform electron gas** (electrons in a box with a uniform positive background) evaluated at the local density:

$$T_{\text{TF}}[\rho] = C_F \int \rho(\mathbf{r})^{5/3}\, d\mathbf{r}, \qquad C_F = \frac{3}{10}(3\pi^2)^{2/3} \approx 2.871.$$

$J[\rho]$ is the **Hartree energy**: the classical self-repulsion of the charge cloud ρ. It is the same as $\frac{1}{2}\sum_{ij} J_{ij}$ of Lecture 4, including the unphysical $i = j$ terms. Dirac soon added an exchange term of the same local form [49], which we will meet again below.

Thomas–Fermi theory gives rough total energies of heavy atoms, but it fails for chemistry in a spectacular way: it has no shell structure (the density decreases smoothly from the nucleus, with no K, L, M shells), and Teller proved in 1962 that **no molecule is bound** in Thomas–Fermi theory [50]. Any molecule is always higher in energy than its separated atoms. The culprit is the kinetic energy. It is the largest energy term after the nuclear attraction, and the local formula above misses most of the subtle changes in kinetic energy that

make chemical bonds. Orbital-free kinetic-energy functionals are still an active research topic, but they are not accurate enough for chemistry.

The lesson for what follows: **do not approximate the kinetic energy with a simple density functional.**

5.5 The Kohn–Sham scheme

5.5.1 A fictitious system with the real density

In 1965, Kohn and Lu Jeu Sham found a way around the kinetic-energy problem [51]. The idea is to bring orbitals back, but only as a tool.

Imagine a fictitious system of N **non-interacting** electrons moving in some effective local potential $v_s(\mathbf{r})$, chosen so that the density of this fictitious system is exactly equal to the density of the real, interacting molecule (Figure 10). A system of non-interacting electrons is easy: its exact wave function is a single Slater determinant of orbitals φ_i, each obeying a one-electron Schrödinger equation. And its kinetic energy can be computed **exactly** from the orbitals:

$$T_s[\rho] = \sum_{i=1}^{N} \left\langle \varphi_i \middle| -\frac{1}{2}\nabla^2 \middle| \varphi_i \right\rangle . \tag{42}$$

T_s is not the true kinetic energy T of the interacting electrons, but it is close: typically it accounts for all but about 1% of it. It also contains the shell structure that Thomas–Fermi theory misses, because it is computed from orbitals.

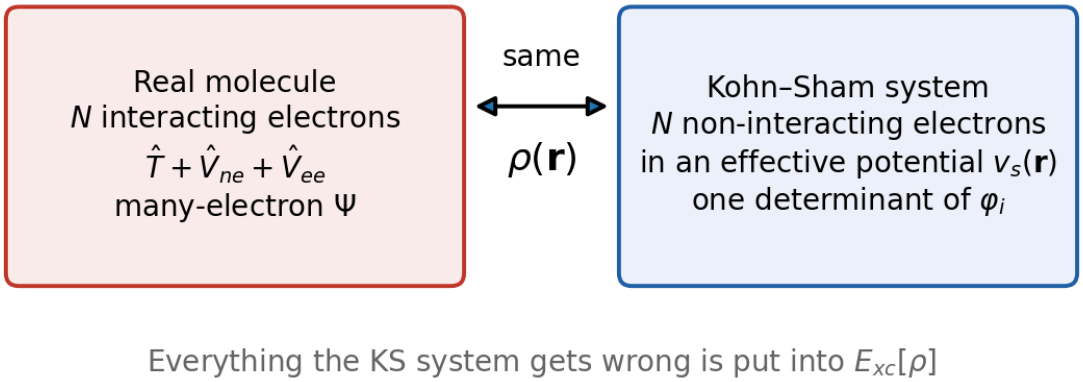


Figure 10: The Kohn–Sham idea. The real, interacting molecule and a fictitious system of non-interacting electrons are linked by one requirement: they have the same electron density. The fictitious system is easy to solve (one Slater determinant). Everything it gets wrong compared with the real system is put into the exchange–correlation energy.

5.5.2 The Kohn–Sham energy decomposition

Kohn and Sham then rewrote the exact energy functional (Equation 40) by separating out the pieces we know how to compute:

$$E[\rho] = T_s[\rho] + \int v_{\text{ext}}(\mathbf{r})\,\rho(\mathbf{r})\,d\mathbf{r} + J[\rho] + E_{xc}[\rho]. \tag{43}$$

The four terms are:

Table 13: The Kohn–Sham energy decomposition. Only the last term must be approximated.

Term	Name	What it is	Known exactly?
$T_s[\rho]$	non-interacting kinetic energy	Equation 42, from the KS orbitals	yes
$E_{ne}[\rho] = \int v_{\text{ext}}\rho$	electron–nuclear attraction	classical attraction of ρ to the nuclei	yes
$J[\rho]$	Hartree (Coulomb) energy	classical self-repulsion of ρ (Equation 41)	yes
$E_{xc}[\rho]$	exchange–correlation energy	everything else	**no**

The **exchange–correlation energy** is *defined* as whatever is needed to make Equation 43 exact. Comparing with $F[\rho] = T[\rho] + V_{ee}[\rho]$:

$$E_{xc}[\rho] = \Big(T[\rho] - T_s[\rho] \Big) + \Big(V_{ee}[\rho] - J[\rho] \Big) . \tag{44}$$

So E_{xc} collects three things:

1. **Exchange**: the Pauli-principle correction to the classical repulsion (the $-K$ terms of Hartree–Fock), including the cancellation of the self-interaction contained in $J[\rho]$.
2. **Correlation**: the extra lowering of the repulsion because electrons avoid each other dynamically.
3. **The kinetic correlation energy** $T - T_s$: the difference between the true kinetic energy of interacting electrons and that of the non-interacting reference.

It is customary to split $E_{xc} = E_x + E_c$. The **exchange** part is defined as the Hartree–Fock-like exchange energy evaluated with the Kohn–Sham orbitals; everything else is **correlation**. Exchange is by far the larger piece. For the helium atom, $E_x \approx -1.03$ Eh and $E_c \approx -0.042$ Eh. For the hydrogen atom, with a single electron, $E_x = -J = -0.3125$ Eh and $E_c = 0$ exactly. Although it is a modest fraction of the total energy (about 10% for the neon atom, and less for heavier atoms), E_{xc} is the "glue" that holds molecules together. Without it, as in Hartree's original method, chemical bonding is very poorly described.

5.5.3 The Kohn–Sham equations

Minimizing Equation 43 with respect to the orbitals, under the constraint that they stay orthonormal, gives the **Kohn–Sham equations**:

$$\left[-\frac{1}{2}\nabla^2 + v_{\text{ext}}(\mathbf{r}) + v_J(\mathbf{r}) + v_{xc}(\mathbf{r}) \right] \varphi_i(\mathbf{r}) = \varepsilon_i \, \varphi_i(\mathbf{r}). \tag{45}$$

The operator in brackets is the **Kohn–Sham operator**. Its potential $v_s = v_{\text{ext}} + v_J + v_{xc}$ has three parts:

- v_{ext}, the attraction to the nuclei;
- the **Hartree potential** $v_J(\mathbf{r}) = \int \rho(\mathbf{r}')/|\mathbf{r} - \mathbf{r}'| \, d\mathbf{r}'$, the electrostatic potential of the whole electron cloud (the same as $\sum_j \hat{J}_j$ in Hartree–Fock);
- the **exchange–correlation potential** $v_{xc}(\mathbf{r}) = \delta E_{xc}/\delta\rho(\mathbf{r})$, the functional derivative of E_{xc}. For the approximations of this chapter, it is simply a function of $\mathbf{r}$ computed from the density at (and near) that point.

i What is a functional derivative?

For an ordinary function $f(x)$, the derivative tells you how much f changes when you change x a little. For a functional $E[\rho]$, the **functional derivative** $\delta E/\delta\rho(\mathbf{r})$ tells you how much E changes when you add a tiny bit of density at the point $\mathbf{r}$. It is therefore itself a function of $\mathbf{r}$: a potential. For a functional of the simple local form $E[\rho] = \int g(\rho(\mathbf{r}))\, d\mathbf{r}$, the functional derivative is just the ordinary derivative of the integrand, $\delta E/\delta\rho(\mathbf{r}) = g'(\rho(\mathbf{r}))$. For example, the LDA exchange energy introduced below, $E_x = -C_x \int \rho^{4/3}\, d\mathbf{r}$, gives the exchange potential $v_x(\mathbf{r}) = -\frac{4}{3} C_x\, \rho(\mathbf{r})^{1/3}$. The Hartree energy $J[\rho]$ gives the Hartree potential v_J in the same way. This is how every term of the energy produces a term of the Kohn–Sham potential.

Compare Equation 45 with the Hartree–Fock equations of Lecture 4:

Table 14: Hartree–Fock and Kohn–Sham side by side.

	Hartree–Fock	Kohn–Sham DFT
Wave function	one Slater determinant, an approximation to Ψ	one Slater determinant of a fictitious system; only its density is "real"
One-electron operator	$\hat{h} + \sum_j (\hat{J}_j - \hat{K}_j)$	$\hat{h} + v_J + v_{xc}$
Exchange	exact, nonlocal ($\hat{K}$)	approximate, inside the local v_{xc}
Correlation	none	approximate, inside v_{xc}
Exact if…	never (misses E_{corr})	the exact E_{xc} were known
Equations in a basis	$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon}$	$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon}$, with $F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + J_{\mu\nu} + V_{\mu\nu}^{xc}$
Solved by	SCF loop	the **same** SCF loop

The two methods share almost all of their machinery. Kohn–Sham DFT expands the orbitals in the same basis sets, builds a Fock-like matrix, and solves the same generalized eigenvalue problem by the same SCF loop, with the same DIIS acceleration. That is why the SCF table of a DFT job in Q-Chem looks exactly like that of an HF job: cycle number, energy, DIIS error, and "Convergence criterion met" at the end. Everything you learned about convergence in Lecture 4 applies unchanged.

⚠ Common mistake: adding up orbital energies

As in Hartree–Fock, the Kohn–Sham total energy is **not** the sum of the occupied orbital energies. Each ε_i contains the full Hartree potential and the full exchange–correlation potential felt by electron i. Summing them counts the electron–electron repulsion twice, and it uses $\int v_{xc}\rho$ in place of E_{xc}, which are different numbers. Programs therefore compute the energy from Equation 43 directly, with the density of the last SCF cycle. The number to use in your reaction energies is always the total energy printed at the end of the SCF ("Total energy in the final basis set", see the guide on reading the output).

! The one sentence to remember

Kohn–Sham DFT is **exact in principle**: with the exact exchange–correlation functional, it would give the exact ground-state energy and density with the cost of a mean-field calculation. All the difficulty, and all the variety among the hundreds of functionals you will meet, is in **approximating $E_{xc}[\rho]$**.

i What do the Kohn–Sham orbitals mean?

Strictly speaking, the KS orbitals and their energies belong to the fictitious non-interacting system. Only the density, and the total energy, are guaranteed to be physical. In practice, KS orbitals look very similar to Hartree–Fock orbitals, and chemists use them in the same way (MO diagrams, HOMO–LUMO pictures in IQmol). For the exact functional, $-\varepsilon_{\mathrm{HOMO}}$ equals the first ionization energy exactly. With the approximate functionals of this chapter and the next, $-\varepsilon_{\mathrm{HOMO}}$ is usually several eV *smaller* than the ionization energy, so Koopmans-style estimates work worse in DFT than in HF. The reason is the self-interaction error discussed in Section 5.8.

5.6 The first approximation: the local spin-density approximation

5.6.1 The uniform electron gas

To use Kohn–Sham DFT we need an approximate $E_{xc}[\rho]$. The first and simplest approximation, already proposed by Kohn and Sham in 1965, uses the one system for which the exchange–correlation energy is accurately known: the **uniform electron gas** (also called *jellium*). It is an infinite system of electrons with constant density ρ, neutralized by a uniform positive background. Its exchange–correlation energy per electron, $\varepsilon_{xc}^{\mathrm{unif}}(\rho)$, depends only on the value of ρ.

The **local density approximation** (LDA) treats every small volume element of a molecule as if it were a piece of uniform gas with the local density $\rho(\mathbf{r})$, and adds up the contributions:

$$E_{xc}^{\mathrm{LDA}}[\rho] = \int \rho(\mathbf{r})\, \varepsilon_{xc}^{\mathrm{unif}}\big(\rho(\mathbf{r})\big)\, d\mathbf{r}. \tag{46}$$

For molecules we always use the spin-resolved version, the **local spin-density approximation** (LSDA), in which $\varepsilon_{xc}^{\mathrm{unif}}$ depends on both spin densities ρ_α and ρ_β. LSDA is the first rung of Jacob's ladder (Lecture 6).

5.6.2 LSDA exchange

For the uniform gas, the exchange energy can be computed analytically from plane-wave orbitals (Dirac, 1930 [49]). The result is

$$E_x^{\mathrm{LDA}}[\rho] = -C_x \int \rho(\mathbf{r})^{4/3}\, d\mathbf{r}, \qquad C_x = \frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \approx 0.7386. \tag{47}$$

In spin-polarized form, exchange acts only between same-spin electrons, and each spin contributes separately:

$$E_x^{\mathrm{LSDA}}[\rho_\alpha, \rho_\beta] = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3} \sum_{\sigma=\alpha,\beta} \int \rho_\sigma(\mathbf{r})^{4/3}\, d\mathbf{r}. \tag{48}$$

For a closed shell ($\rho_\alpha = \rho_\beta = \rho/2$), Equation 48 reduces to Equation 47. Notice the power $4/3$. You can check it by dimensional analysis: an energy in atomic units has dimension 1/length, a density 1/length³, and $\rho^{4/3}\, d\mathbf{r}$ has dimension $(1/\text{length})^4 \times \text{length}^3 = 1/\text{length}$.

In 1951 Slater arrived at the same $\rho^{4/3}$ form independently, by approximating the Hartree–Fock exchange potential [52]. His constant was $3/2$ times larger. Programs later introduced the $X\alpha$ method, with an adjustable prefactor α: Slater's value is $\alpha = 1$, and the uniform-gas (Kohn–Sham) value is $\alpha = 2/3$. Today, "Slater exchange" in a functional name (the "S" in SVWN or SPW92) means Equation 48 with the uniform-gas constant.

5.6.3 LSDA correlation

The correlation energy of the uniform gas has no closed form. It is known analytically only in the limits of very high and very low density. In 1980, Ceperley and Alder computed it accurately at intermediate densities with quantum Monte Carlo simulations [53]. Others then fitted analytic formulas that reproduce the Monte Carlo numbers and the known limits, as functions of the **Wigner–Seitz radius** $r_s = (3/4\pi\rho)^{1/3}$ (the radius of a sphere that contains, on average, one electron) and of the spin polarization $\zeta = (\rho_\alpha - \rho_\beta)/\rho$. The two parametrizations you will see are:

- **VWN** (Vosko, Wilk and Nusair, 1980) [54], which comes in several variants (the paper gives more than one fit). Different programs use different VWN variants inside the same named functional, which is one reason why "the same" functional can give slightly different energies in different codes.
- **PW92** (Perdew and Wang, 1992) [55], a later and more careful fit.

In this course, LSDA means **SPW92**: Slater exchange plus PW92 correlation. Its close relative SVWN5 (Slater plus VWN correlation, form V) gives very similar results. Both are "pure" LSDA: they use nothing but the local spin densities.

i Worked example: LSDA for the hydrogen atom

The hydrogen atom is a good test because we know the exact answer. Its exact density is $\rho(r) = e^{-2r}/\pi$, fully spin polarized ($\rho_\alpha = \rho$, $\rho_\beta = 0$). The exact energy terms are $T = 0.5$, $E_{ne} = -1$ Eh and, since there is only one electron, $V_{ee} = 0$. In Kohn–Sham language, $J[\rho] = 5/16 = 0.3125$ Eh, so the exact E_{xc} must be -0.3125 Eh: $E_x = -0.3125$ Eh cancels the self-repulsion exactly, and $E_c = 0$.

Evaluate the LSDA terms with the exact density (computed with the script in `images/src/l04_l06_figures.py`):

	Exact	LSDA (SPW92)
E_x (Eh)	-0.3125	-0.268
E_c (Eh)	0	-0.022
$J + E_{xc}$ (Eh)	0	$+0.022$
Total energy (Eh)	-0.500	-0.478

LSDA exchange recovers only 86% of the exchange energy needed to cancel the self-repulsion, and LSDA correlation adds a spurious correlation energy to a system with one electron. The net result is an electron that repels itself by 0.022 Eh (14 kcal/mol). This is the **self-interaction error** in its simplest form. In Hartree–Fock this error is exactly zero (Lecture 4).

5.6.4 What LSDA gets right, and what it gets wrong

It is surprising that a model built on a uniform gas works at all for molecules, where the density varies by orders of magnitude between the nuclei and the tails. It works better than it should, for a deep reason: the exchange–correlation *hole* of the uniform gas obeys the same exact sum rules as the true hole (it contains exactly one electron, removed from around each electron), and the energy depends only on a spherical average of the hole. Errors in exchange and in correlation also partly cancel.

What LSDA does well:

- **Geometries**: bond lengths are usually within a few hundredths of an ångström (often slightly too short), and bond angles are good.
- **Vibrational frequencies** of reasonable quality, and good properties of simple metals and many solids, which is why LSDA was the workhorse of solid-state physics for decades.

What LSDA does badly:

- **Overbinding.** Atomization energies are much too large, typically by tens of kcal/mol for small molecules, and the error grows with molecular size. Bond energies, hydrogen bonds and weak complexes are all too strong.
- **Barrier heights** are too low, often far too low.
- **Ionization energies from orbital energies** are much too small, and anions are often not bound.

These errors are the reason LSDA was never popular in chemistry. Chemists started to use DFT only after the gradient corrections of the late 1980s and the hybrids of 1993 (the subject of Lecture 6). In Experiment 3 you will compute SPW92 alongside fifteen other methods and see where it stands.

5.7 How a DFT program evaluates E_{xc}: the integration grid

In Hartree–Fock, every integral in the energy is a product of Gaussian basis functions that can be done analytically. In DFT, the terms T_s, E_{ne} and J are computed in the same way. But E_{xc} is not. Integrals like $\int \rho^{4/3}\, d\mathbf{r}$, and the more complicated ones of the next chapter, have **no analytic solution** for a density built from Gaussians. Programs compute them numerically, as a weighted sum over a **grid of points** placed around each atom:

$$E_{xc} \approx \sum_k w_k \, \rho(\mathbf{r}_k)\, \varepsilon_{xc}\left[\, \rho(\mathbf{r}_k), \ldots \,\right] .$$

The grid is an extra approximation that Hartree–Fock does not have. A grid that is too coarse gives energies with numerical noise, which can be larger than you expect. You will meet grids in detail in Lecture 11 and test them in Experiment 7. For now, remember that the grid is the second reason, after the choice of E_{xc}, why two DFT calculations with "the same" functional and basis can disagree.

One more preview. Nothing forces E_{xc} to be a pure density functional. **Hybrid functionals** replace part of the approximate exchange by exact (Hartree–Fock) exchange, computed from the Kohn–Sham orbitals with the $\hat{K}$ operator of Lecture 4. This partly cures the self-interaction error. Hybrids are the fourth rung of Jacob's ladder, in Lecture 6.

5.8 Known errors of approximate functionals

Every functional you will use in this course is an approximation, and approximations fail in characteristic ways. Three of them are worth knowing by name from the start. Each returns in later classes, with experiments designed to expose it.

1. Self-interaction and delocalization error. In approximate functionals, the approximate E_x does not exactly cancel the self-repulsion contained in $J[\rho]$ (as in the hydrogen atom above). Perdew and Zunger made this precise in 1981 and proposed a correction [56]. Because an electron that repels itself prefers to be spread out, approximate functionals systematically over-stabilize **delocalized** charge. This is called **delocalization error** [57]. Its symptoms are striking. In stretched H_2^+, a one-electron system, LSDA and GGAs predict an energy far too low, with the electron shared between two distant protons. Reaction barriers come out too low, because transition states have stretched, delocalized bonds. Charge-transfer complexes are overbound, and orbital energies are too high. Mixing in Hartree–Fock exchange, which has no self-interaction, reduces this error (Lecture 6). It returns in Lecture 9.

2. Missing long-range dispersion. LSDA depends only on the density at each point. GGAs and meta-GGAs (Lecture 6) also look at its derivatives at that point, but still only locally. Dispersion, however, is a correlation between density fluctuations at two points that can be far apart, on two molecules whose densities do not overlap. A local or semilocal functional cannot see it: as two argon atoms are pulled apart, the energy of any such functional approaches the separated-atom value too quickly. The functional may give some spurious binding where the densities overlap (LSDA often does), but for the wrong reason. Dispersion must be added, by an empirical correction like **-D4** [58], [59] or by a nonlocal correlation functional (VV10) [60]. This is the subject of Lecture 8 and Experiment 4.

3. Static correlation. The Kohn–Sham reference is a single determinant. When the true wave function needs two or more determinants with similar weights (stretched bonds, diradicals, ozone, many transition-metal compounds), approximate functionals struggle. In principle the exact E_{xc} would handle it. In practice, approximate functionals do it badly or not at all, and often the unrestricted calculation breaks spin symmetry, just as UHF does for stretched H_2 (Lecture 4). Static correlation and self-interaction are related, and they pull in

opposite directions: adding Hartree–Fock exchange reduces self-interaction error but makes static-correlation problems worse. This tension is one reason why no single functional is best for everything. It returns in Lecture 9 and Lecture 12.

> ⚠ Common mistake: "DFT includes correlation, so it is always better than HF"
>
> Approximate DFT includes *approximate* exchange and correlation. For most chemistry it is far better than HF. But it can also be worse, and in ways that HF is not: HF has no self-interaction error, so for some properties dominated by that error (for example, the dissociation of one-electron bonds) HF is qualitatively right and LSDA or GGA functionals are qualitatively wrong. Always ask which errors your problem is sensitive to.

Key points

- The electron density $\rho(\mathbf{r})$ depends on 3 coordinates, while the wave function depends on $3N$. It integrates to N, has cusps at the nuclei, and is measurable.
- **Hohenberg–Kohn (1964)**: the ground-state density determines the external potential and therefore everything; the energy is a functional of the density, minimized by the true density. The theorems prove existence but do not give the functional.
- **Thomas–Fermi** theory approximates the kinetic energy as a local density functional and binds no molecules. Lesson: don't approximate T that way.
- **Kohn–Sham (1965)**: a fictitious non-interacting system with the same density. Its kinetic energy T_s is computed exactly from orbitals. $E[\rho] = T_s + E_{ne} + J + E_{xc}$.
- E_{xc} contains exchange, correlation, and the kinetic correction $T - T_s$. It is the only unknown term, and **KS-DFT would be exact with the exact E_{xc}**.
- The KS equations look like the HF equations, with the exchange operator replaced by a local potential v_{xc}. They are solved by the **same SCF machinery**, and the output looks the same.
- **LSDA** takes E_{xc} from the uniform electron gas: exchange $\propto \rho^{4/3}$ (Dirac/Slater), correlation from quantum Monte Carlo data fitted by VWN or PW92. SPW92 is the LSDA of this course.
- LSDA gives good geometries but **overbinds**: atomization energies far too large, barriers too low.
- E_{xc} is evaluated numerically on a **grid** (Lecture 11).
- Three error types to remember: **self-interaction/delocalization error**, **missing dispersion**, and **static correlation**.

Study questions

1. How many spatial coordinates does the wave function of a water molecule depend on? And its electron density? Why does this matter for the cost of a calculation?
2. State the first Hohenberg–Kohn theorem in your own words. Then reproduce the proof by contradiction for a nondegenerate ground state.
3. What is a functional? Give two examples from this chapter or from Lecture 4, and say what each takes as input and what it returns.
4. Thomas–Fermi theory and Kohn–Sham DFT both write the energy as a functional of the density. What is the key difference that makes Kohn–Sham useful for chemistry while Thomas–Fermi binds no molecules?
5. Write the Kohn–Sham energy decomposition, name each term, and say which terms are computed exactly. List the three physical effects collected in E_{xc}.

6. Compare the Hartree–Fock and Kohn–Sham one-electron operators term by term. Which term is nonlocal in HF and local in LSDA?
7. Show that Equation 48 reduces to Equation 47 for a closed-shell density. Then use dimensional analysis to explain why exchange in the LDA must scale as $\rho^{4/3}$.
8. For the hydrogen atom, the exact exchange–correlation energy must equal $-J[\rho]$. Explain why, and use the worked-example numbers to compute the self-interaction error of LSDA in kcal/mol.
9. A DFT calculation and an HF calculation on the same molecule with the same basis set both print an SCF table. Name two numerical approximations that are present in the DFT job but absent from the HF job.
10. Name the three main error types of approximate functionals. For each, give one chemical situation where you would expect it to matter.

Further reading

- W. Koch and M. C. Holthausen [8], chapters 3–6: the standard introduction to DFT for chemists, from Hohenberg–Kohn to the LSDA and beyond. Written for exactly the audience of this course.
- W. Kohn [44]: Walter Kohn's Nobel lecture, an elegant and readable account of why the density is enough.
- K. Burke [61]: a short perspective on what DFT is, what it is not, and where approximations come from.
- R. G. Parr and W. Yang [62]: the classic, more mathematical textbook.
- A. J. Cohen, P. Mori-Sánchez, and W. Yang [57]: a review of the main errors of approximate functionals (self-interaction, delocalization, static correlation), for after Lecture 9.
- R. Peverati and D. G. Truhlar [30] and N. Mardirossian and M. Head-Gordon [63]: reviews of modern functionals, useful for Experiment 2.

6 Approximating the Hamiltonian III: Jacob's ladder

Learning goals

After studying this chapter you should be able to:

- describe Perdew's **Jacob's ladder** and name the ingredient that each rung adds;
- define the **reduced density gradient** s and the **exchange enhancement factor** $F_x(s)$, and explain what B88 and PBE do;
- define the **kinetic-energy density** τ and explain what a meta-GGA can do that a GGA cannot;
- write the formula of a **global hybrid** (B3LYP, PBE0) and explain, with the adiabatic connection, why mixing in Hartree–Fock exchange helps;
- explain **range separation** with $\mathrm{erf}(\omega r)/r$ and read the percentages of a range-separated hybrid;
- say what a **double hybrid** adds;
- tell how **dispersion** is included in a functional (-D4, VV10, fitted, or not at all) and whether a functional is **non-empirical** or **empirical**;
- place any functional on the ladder from its ingredients, using the original paper or a review: the task of Experiment 2.

Review from Lectures 4 and 5

This chapter builds directly on the previous two. Before reading it, make sure you can:

- write the Hartree–Fock exchange energy and explain why it is free of self-interaction (Lecture 4);
- write the Kohn–Sham energy $E = T_s + E_{ne} + J + E_{xc}$ (Lecture 5);
- write the LSDA exchange energy, $E_x^{\mathrm{LDA}} = -C_x \int \rho^{4/3}\, d\mathbf{r}$, and list what LSDA gets right and wrong (Lecture 5).

6.1 Jacob's ladder

In Lecture 5 we saw that Kohn–Sham DFT is exact except for one term, the exchange–correlation energy $E_{xc}[\rho]$, and that the simplest approximation to it, LSDA, overbinds molecules badly. Since the late 1980s, hundreds of better approximations have been proposed. Their names can look like a random collection of letters and numbers: BLYP, PBE0, M06-2X, ωB97M-V, MN15-L. There is order behind them, and the most useful way to see it is a metaphor introduced by John Perdew and Karla Schmidt in 2001 [64], [65].

In the Book of Genesis, Jacob dreams of a ladder that connects earth to heaven. In Perdew's version, the ladder rises from the "Hartree world" (no exchange or correlation at all, as in Hartree's 1928 method) to the "heaven of chemical accuracy", meaning errors below about 1 kcal/mol for energy differences. Each rung adds a new **ingredient**: a new kind of local information that the functional can use (Figure 11).

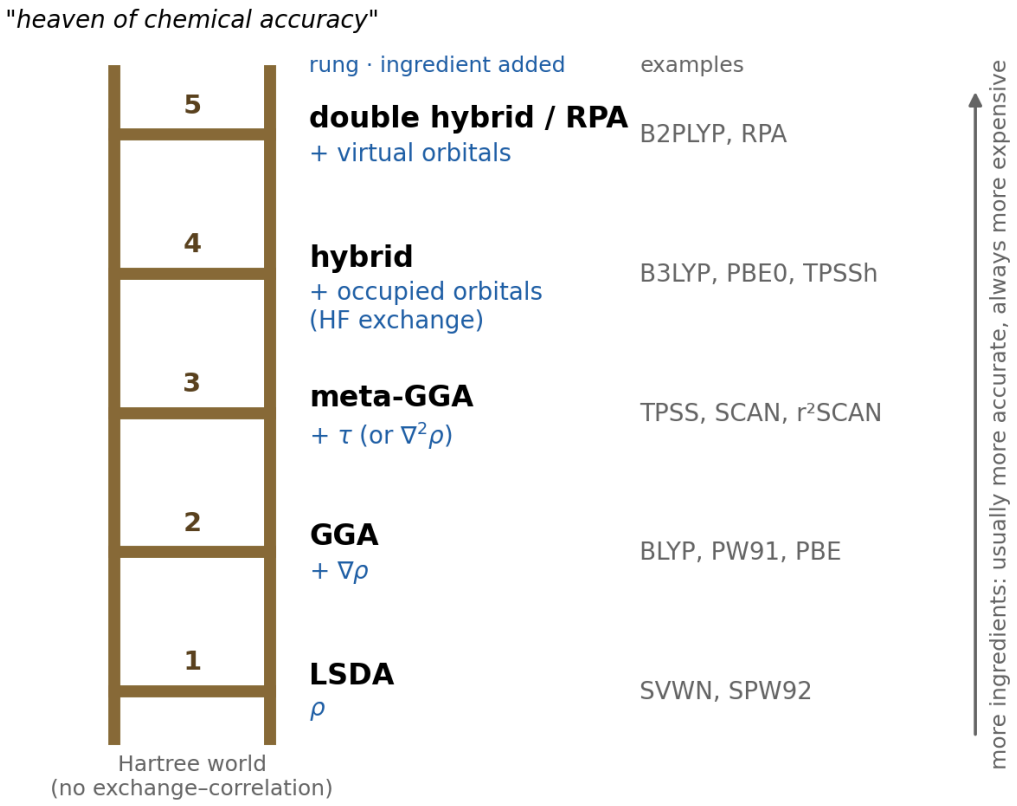


Figure 11: Perdew's Jacob's ladder of density functional approximations. Each rung keeps the ingredients of the rungs below it and adds a new one. Higher rungs are usually more accurate and always more expensive. Examples are typical functionals for each rung.

Table 15: The five rungs of Jacob's ladder.

Rung	Name	New ingredient	Typical examples
1	LSDA	spin densities ρ_α, ρ_β	SVWN, SPW92
2	GGA	density gradient $\nabla\rho_\sigma$	BLYP, PW91, PBE
3	meta-GGA	kinetic-energy density τ_σ (or the Laplacian $\nabla^2\rho_\sigma$)	TPSS, SCAN, r²SCAN
4	hybrid	occupied orbitals, through Hartree–Fock exchange	B3LYP, PBE0, TPSSh, ωB97X-V
5	double hybrid, RPA	virtual (unoccupied) orbitals	B2PLYP, RPA

Three facts about the ladder are worth stating clearly.

- **Each rung includes the lower ones.** A GGA still uses the density, a meta-GGA still uses the gradient, a hybrid is built on top of a GGA or a meta-GGA.

- **Higher is usually better, but not always.** A well-designed GGA can beat a poorly designed hybrid for some properties. The ladder ranks *ingredients*, not individual functionals.
- **Higher is always more expensive.** Rungs 1–3 are **semilocal**: the energy density at a point depends only on information at that point, and the cost is small. Rung 4 needs the exchange matrix of Hartree–Fock, which is more expensive, and rung 5 needs a correlated wave-function step like MP2.

The rest of this chapter climbs the ladder one rung at a time. Rung 1 (LSDA) was the subject of Lecture 5.

6.2 Rung 2: the generalized gradient approximation (GGA)

6.2.1 The reduced gradient and the enhancement factor

LSDA uses only the value of the density at each point. In a molecule, the density changes rapidly: near the nuclei, in bonds, and in the tails. The natural next ingredient is the **gradient** of the density, $\nabla\rho$, which tells the functional how fast the density is changing. The early attempts to add a simple gradient correction (the gradient expansion) made things worse, because the expansion behaves badly where the gradient is large. The solution, developed in the 1980s, was the **generalized gradient approximation** (GGA): a more flexible function of ρ and $\nabla\rho$ that is built to behave well everywhere.

To compare gradients in regions of very different density, GGAs use a dimensionless **reduced density gradient**:

$$s(\mathbf{r}) = \frac{|\nabla\rho(\mathbf{r})|}{2(3\pi^2)^{1/3}\rho(\mathbf{r})^{4/3}}. \tag{49}$$

$s = 0$ in the uniform gas. In molecules, s is small (below about 1) in the valence and bonding regions, and it grows without limit in the exponential tails of the density, far from the nuclei.

Any GGA exchange functional can then be written as the LDA exchange energy density multiplied by an **enhancement factor** $F_x(s)$:

$$E_x^{\mathrm{GGA}}[\rho] = -C_x \int \rho(\mathbf{r})^{4/3}\, F_x\big(\, s(\mathbf{r})\big)\, d\mathbf{r}, \qquad C_x = \frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}. \tag{50}$$

(For spin-polarized systems the same formula is applied to each spin density separately.) $F_x = 1$ gives back LSDA. Different GGA exchange functionals are simply different choices of $F_x(s)$. Because $F_x > 1$, GGAs make the exchange energy more negative than LSDA, which, as we saw for the hydrogen atom in Lecture 5, is what is needed.

6.2.2 B88 exchange

In 1988 Axel Becke proposed the exchange functional that made DFT popular in chemistry [66]. In its original form, written for each spin density,

$$E_x^{\mathrm{B88}} = E_x^{\mathrm{LSDA}} - \beta \sum_\sigma \int \rho_\sigma^{4/3}\, \frac{x_\sigma^2}{1 + 6\beta\, x_\sigma \sinh^{-1} x_\sigma}\, d\mathbf{r}, \qquad x_\sigma = \frac{|\nabla\rho_\sigma|}{\rho_\sigma^{4/3}}, \tag{51}$$

with a single parameter $\beta = 0.0042$ a.u. Becke chose the form so that the exchange energy density has the correct $-1/(2r)$ behavior far from an atom, and fitted β to the exact (Hartree–Fock) exchange energies of the noble-gas atoms. B88 is therefore a slightly **empirical** functional: one parameter, fitted to data. It is the "B" in BLYP, BP86 and B3LYP.

6.2.3 PBE exchange and the idea of exact constraints

In 1996 Perdew, Burke and Ernzerhof proposed a GGA with a different philosophy [67]. Instead of fitting to data, they chose the simplest form that satisfies known **exact constraints**, properties that the exact functional is proven to have. The PBE enhancement factor is

$$F_x^{\mathrm{PBE}}(s) = 1 + \kappa - \frac{\kappa}{1 + \mu s^2/\kappa}, \qquad \kappa = 0.804, \quad \mu \approx 0.2195. \tag{52}$$

Each number has a physical origin. μ is chosen so that, for small gradients, the gradient correction to exchange cancels that of PBE correlation, to recover the accurate linear response of the uniform gas. κ is chosen so that F_x never exceeds 1.804, a local version of the **Lieb–Oxford bound**, a proven lower limit on how negative the exchange–correlation energy can be. PBE is **non-empirical**: no parameter was fitted to molecular data. This makes it a safe, general-purpose functional for chemistry and especially for solids, where it is the default in many codes.

Figure 12 compares the two enhancement factors. They are almost identical for $s < 1$, where most of the chemistry happens. They differ in the tails: B88 keeps growing, while PBE levels off toward its bound.

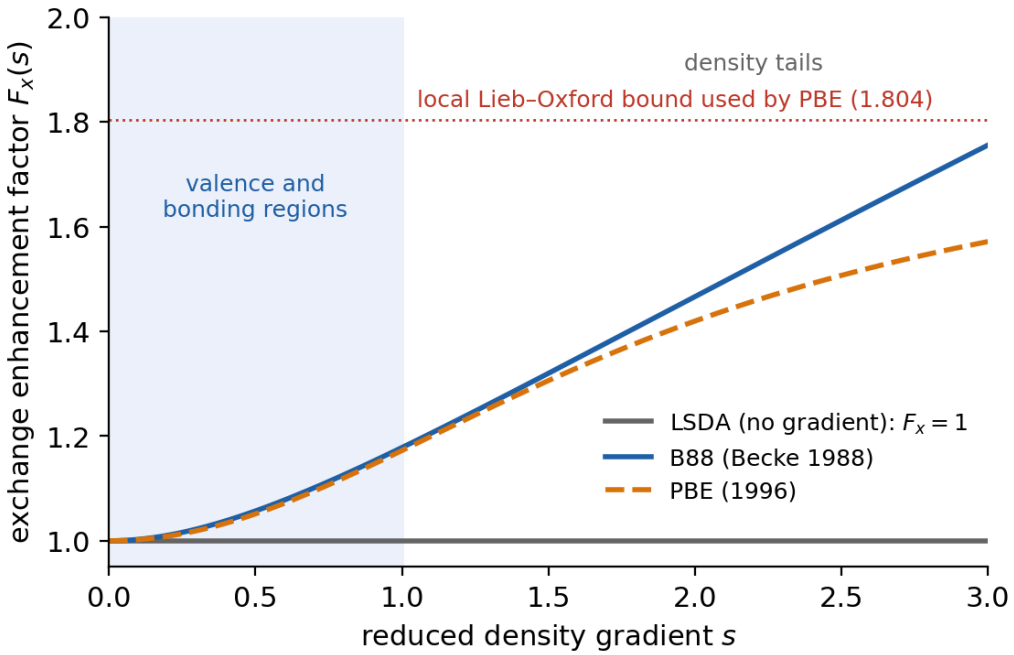


Figure 12: Exchange enhancement factors of B88 and PBE as functions of the reduced gradient s, computed from Equation 51 and Equation 52 for a spin-unpolarized density. LSDA corresponds to $F_x = 1$. The two GGAs agree closely in the valence region ($s < 1$) and differ in the low-density tails. Computed with `images/src/l04_l06_figures.py`.

i Worked example: GGA exchange for the hydrogen atom

In Lecture 5 we evaluated LSDA exchange for the exact density of the hydrogen atom and found -0.268 Eh, against the exact -0.3125 Eh. With the same density, the two GGAs give (computed with the course script):

	LSDA	PBE	B88	Exact
E_x (Eh)	-0.268	-0.306	-0.310	-0.3125
error (kcal/mol)	$+28$	$+4$	$+2$	0

The gradient correction removes most of the exchange error. The rest is a remaining self-interaction error, which a semilocal functional cannot remove completely.

6.2.4 GGA correlation and the names of GGA functionals

A GGA functional needs a correlation part too. The common ones are:

- **LYP** (Lee, Yang and Parr, 1988) [68], derived from a correlation formula fitted to the helium atom. It is not built on the uniform gas at all, and it gives zero correlation for any one-electron system, as it should.
- **P86** (Perdew, 1986) [69], [70], an early gradient correction to LSDA correlation.
- **PW91** (Perdew, Wang and co-workers, 1991–1992) [71], a non-empirical GGA for both exchange and correlation, the predecessor of PBE.
- **PBE** correlation [67], the non-empirical partner of PBE exchange.

GGA names usually combine the exchange part and the correlation part: **BLYP** is B88 exchange + LYP correlation, **BP86** is B88 + P86, **PBE** is PBE exchange + PBE correlation. GGAs reduced the errors of LSDA in atomization energies by a large factor, from tens of kcal/mol to typically 5–10 kcal/mol for small molecules, and gave much better hydrogen bonds and barriers. This is why the 1990s were the decade in which chemists adopted DFT.

A third family of GGAs is built with Becke's 1997 power-series form [72], in which F_x and the correlation enhancement factors are written as short power series with coefficients fitted to data. Many modern **empirical** functionals use this form, including the B97 series and the ωB97 family that we will meet below.

6.3 Rung 3: meta-GGAs

6.3.1 The kinetic-energy density

The third rung adds an ingredient that depends on the Kohn–Sham orbitals, but only in a semilocal way: the **kinetic-energy density** of each spin,

$$\tau_\sigma(\mathbf{r}) = \frac{1}{2}\sum_{i}^{\text{occ}} |\nabla \varphi_{i\sigma}(\mathbf{r})|^2. \tag{53}$$

Integrated over space, τ gives the Kohn–Sham kinetic energy T_s of Lecture 5. Some meta-GGAs use the **Laplacian** of the density, $\nabla^2\rho$, instead of or in addition to τ. Both carry information about how the density is curved, which the gradient alone does not.

Why is τ useful? Compare it with the **von Weizsäcker** kinetic-energy density, $\tau_W = |\nabla\rho|^2/(8\rho)$, which is the exact kinetic-energy density whenever only one orbital is occupied. The ratio τ_W/τ equals 1 in regions

dominated by a single orbital (for example, any one-electron system, or a region dominated by a single electron pair, like a covalent single bond) and is smaller elsewhere. Modern meta-GGAs use a closely related variable, $\alpha = (\tau - \tau_W)/\tau^{\text{unif}}$, which is

- $\alpha = 0$ in one-orbital regions,
- $\alpha = 1$ in the uniform electron gas,
- $\alpha \gg 1$ in the region between two closed shells that do not bond (for example, between two weakly interacting molecules).

So a meta-GGA can **recognize** the kind of chemical region it is in, which a GGA cannot. This lets it, for instance, make the correlation energy vanish for any one-electron density (removing the one-electron self-correlation error of LSDA and most GGAs), and treat single bonds, metallic-like regions and weak interactions differently.

6.3.2 Examples

- **TPSS** (Tao, Perdew, Staroverov and Scuseria, 2003) [73], a non-empirical meta-GGA.
- **SCAN** ("strongly constrained and appropriately normed", 2015) [74], a non-empirical meta-GGA that satisfies all 17 exact constraints known to be appropriate for a meta-GGA, and is "normed" to reproduce exactly a few systems (such as the hydrogen atom and the uniform gas). SCAN is accurate but numerically delicate, and needs fine integration grids. Its revision **r²SCAN** (2020) [75] keeps most of its accuracy with much better numerical behavior.
- **Empirical meta-GGAs**, fitted to large databases of reference energies. An example is **B97M-rV** [76], [77].

Meta-GGAs cost only a little more than GGAs, because τ is computed from the same orbitals that the program already has. They need finer integration grids, though, because τ and α vary rapidly in space and many fitted meta-GGAs have oscillating enhancement factors. You will see this directly in Experiment 7 and in Lecture 11.

> ⚠ "Local" means "no Hartree–Fock exchange", not "LSDA"
>
> Rungs 1–3 together are called **semilocal** or simply **local** functionals: they use only information at each point (density, gradient, τ), and no Hartree–Fock exchange. In the Minnesota family, the suffix **-L** marks such a local functional. The suffix does not tell you whether the functional is a GGA or a meta-GGA: for that, you need to know which ingredients it uses.

6.4 Rung 4: hybrid functionals

6.4.1 Why mix in Hartree–Fock exchange? The adiabatic connection

So far, every functional has been a function of local quantities. The fourth rung adds a completely different ingredient: the occupied orbitals themselves, used to compute **exact exchange**, the Hartree–Fock exchange energy of Lecture 4 evaluated with the Kohn–Sham orbitals:

$$E_x^{\text{HF}} = -\frac{1}{2}\sum_{i,j}^{\text{occ}} K_{ij}. \tag{54}$$

Why would that help? A rigorous argument comes from the **adiabatic connection**. Imagine scaling the electron–electron repulsion by a factor λ that goes from 0 to 1, while adjusting the external potential to keep the density fixed. At $\lambda = 0$ the electrons do not interact: this is the Kohn–Sham system, and its exchange–

correlation energy is pure exchange, exactly E_x^{HF} with KS orbitals. At $\lambda = 1$ we have the real molecule. The exact E_{xc} is the average of the exchange–correlation energy over this path:

$$E_{xc} = \int_0^1 W_\lambda \, d\lambda, \tag{55}$$

where W_λ is the potential-energy part of exchange–correlation at coupling strength λ. We know the $\lambda = 0$ end exactly ($W_0 = E_x^{\mathrm{HF}}$). Semilocal functionals are best near the $\lambda = 1$ end. A simple approximation to the integral therefore mixes the two. Becke tried exactly this in 1993, with a "half-and-half" mixture of HF exchange and LSDA [78]. A few months later he proposed a more flexible, fitted version [79], the ancestor of B3LYP.

The chemical argument is simpler. Hartree–Fock is free of self-interaction (Lecture 4), while semilocal functionals are not (Lecture 5). Their errors often have opposite signs: HF underbinds and gives barriers that are too high, semilocal DFT overbinds and gives barriers that are too low. Mixing the two cancels part of both errors.

6.4.2 Global hybrids: B3LYP and PBE0

A **global hybrid** uses the same fraction a_0 of HF exchange everywhere in space:

$$E_{xc}^{\mathrm{hybrid}} = a_0 \, E_x^{\mathrm{HF}} + (1 - a_0) \, E_x^{\mathrm{DFA}} + E_c^{\mathrm{DFA}}, \tag{56}$$

where DFA stands for the underlying semilocal density functional approximation (a GGA or a meta-GGA).

B3LYP is the most cited functional in the history of chemistry. Becke's 1993 three-parameter form [79], with LYP correlation substituted by Stephens and co-workers in 1994 [80], is

$$E_{xc}^{\mathrm{B3LYP}} = (1 - a_0) \, E_x^{\mathrm{LSDA}} + a_0 \, E_x^{\mathrm{HF}} + a_x \, \Delta E_x^{\mathrm{B88}} + (1 - a_c) \, E_c^{\mathrm{VWN}} + a_c \, E_c^{\mathrm{LYP}}, \tag{57}$$

with $a_0 = 0.20$, $a_x = 0.72$ and $a_c = 0.81$. Here $\Delta E_x^{\mathrm{B88}}$ is the gradient correction part of B88 (Equation 51, minus its LSDA part). So B3LYP contains **20% HF exchange**. The three parameters were fitted by Becke to a small set of thermochemical data (with PW91 correlation in his original functional, today called **B3PW91**), and the same values were kept when LYP replaced PW91. B3LYP is an empirical hybrid GGA. Because different programs use different forms of VWN correlation (Lecture 5), "B3LYP" energies can differ slightly from code to code.

PBE0 [81], [82] is the non-empirical counterpart:

$$E_{xc}^{\mathrm{PBE0}} = \frac{1}{4} E_x^{\mathrm{HF}} + \frac{3}{4} E_x^{\mathrm{PBE}} + E_c^{\mathrm{PBE}}. \tag{58}$$

The **25% HF exchange** is not fitted. Perdew, Ernzerhof and Burke argued from perturbation theory that 1/4 is a good general choice for molecules [83].

Hybrids built on meta-GGAs are **hybrid meta-GGAs**. Examples are TPSSh (TPSS with 10% HF exchange) [84] and several Minnesota functionals: **M06-2X** [85] has **54%** HF exchange (the "2X" stands for doubled HF exchange relative to its parent functional), and **MN15** [31] has **44%**. Notice how much the HF fraction varies between functionals. Fitted functionals designed for main-group thermochemistry and barrier heights tend to use more HF exchange (40–60%), while functionals meant also for transition metals use less, because HF exchange makes the static-correlation problems of Lecture 5 worse.

i The cost of exact exchange

Exact exchange is computed exactly as in Hartree–Fock: the program must build the exchange matrix from the two-electron integrals, which costs more than the Coulomb part and much more than evaluating a semilocal functional on a grid. In practice, a hybrid calculation takes noticeably longer per SCF cycle than a GGA in the same basis. You will measure this in Experiment 3.

6.4.3 Range-separated hybrids

A global hybrid uses the same fraction of HF exchange for electrons that are close together and for electrons that are far apart. That is a compromise. At **short range**, semilocal exchange works reasonably well, and its errors partly cancel those of semilocal correlation. At **long range**, semilocal exchange is qualitatively wrong: the exchange–correlation potential of a semilocal functional decays exponentially far from a molecule, instead of as $-1/r$. This makes orbital energies, charge-transfer excitations, Rydberg states and stretched bonds poorly described, and leaves residual self-interaction where it hurts most.

Range-separated hybrids use different amounts of HF exchange at different electron–electron distances. They split the Coulomb operator with the error function:

$$\frac{1}{r_{12}} = \underbrace{\frac{\operatorname{erfc}(\omega r_{12})}{r_{12}}}_{\text{short range}} + \underbrace{\frac{\operatorname{erf}(\omega r_{12})}{r_{12}}}_{\text{long range}}, \tag{59}$$

where erf is the error function, $\operatorname{erfc} = 1 - \operatorname{erf}$, and the **range-separation parameter** ω (in bohr^{-1}) sets where the switch happens: $1/\omega$ is roughly the distance at which long range takes over (Figure 13, left). The exchange energy is then written as

$$E_x = c_{\text{SR}}\, E_x^{\text{HF,SR}} + (1 - c_{\text{SR}})\, E_x^{\text{DFA,SR}} + c_{\text{LR}}\, E_x^{\text{HF,LR}} + (1 - c_{\text{LR}})\, E_x^{\text{DFA,LR}}, \tag{60}$$

where c_{SR} and c_{LR} are the fractions of HF exchange at short and long range. The effective fraction of HF exchange grows smoothly from c_{SR} at $r_{12} = 0$ to c_{LR} at large r_{12} (Figure 13, right).

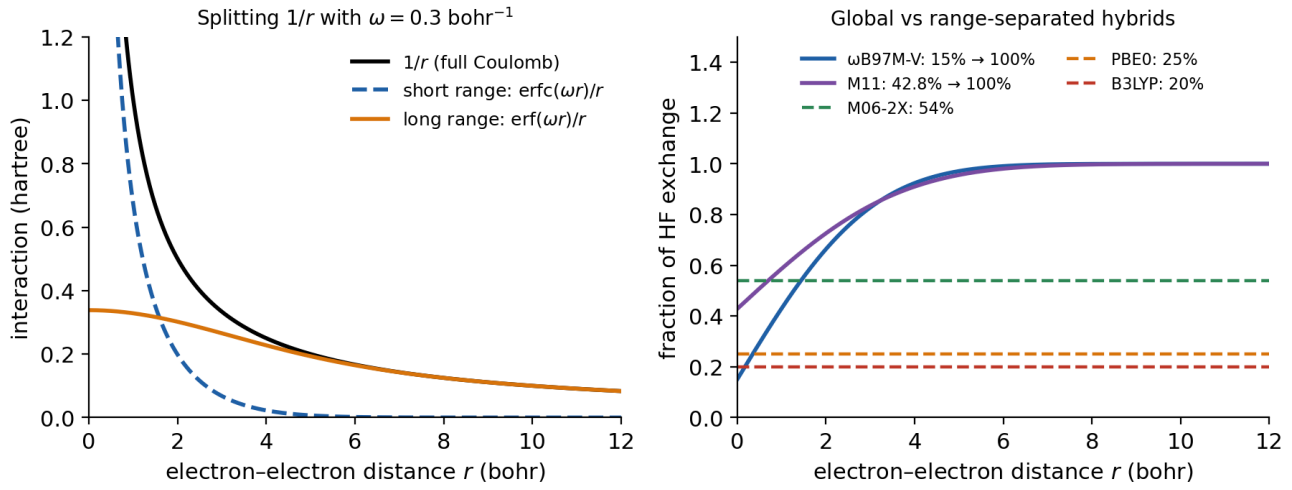


Figure 13: Left: the Coulomb interaction $1/r$ split into a short-range part, $\mathrm{erfc}(\omega r)/r$, and a long-range part, $\mathrm{erf}(\omega r)/r$, for $\omega = 0.3$ bohr^{-1}. Right: the effective fraction of Hartree–Fock exchange as a function of the electron–electron distance for two range-separated hybrids (ωB97M-V with $\omega = 0.3$ bohr^{-1}, M11 with $\omega = 0.25$ bohr^{-1}) and three global hybrids. Computed with `images/src/l04_l06_figures.py`.

The most common choice in chemistry is a **long-range corrected** hybrid, with $c_{\mathrm{LR}} = 1$ (100% HF exchange at long range, which restores the correct $-1/r$ potential). The idea goes back to Iikura, Tsuneda, Yanai and Hirao [86]. Examples:

Table 16: Some range-separated hybrids. CAM-B3LYP uses a slightly more general partition than Equation 60. HSE does the opposite of a long-range correction: it keeps HF exchange only at short range, which makes calculations on metals and solids cheaper.

Functional	c_{SR}	c_{LR}	ω (bohr^{-1})	Base rung
CAM-B3LYP [87]	19%	65%	0.33	GGA
ωB97X-V [88]	16.7%	100%	0.30	GGA
M11 [89]	42.8%	100%	0.25	meta-GGA
ωB97M-V [90]	15%	100%	0.30	meta-GGA
HSE (screened, for solids) [91]	25%	0%	about 0.1	GGA

The "ω" at the beginning of a name (ωB97X-V, ωB97M-V, ωPBE) means range separation. Not every range-separated functional has it in its name, though: M11 and CAM-B3LYP are both range-separated. Range separation is also used in some local functionals, with two *different semilocal* exchange forms at short and long range and no HF exchange at all. So "range-separated" does not automatically mean "hybrid".

6.5 Rung 5: double hybrids and RPA

The fifth rung uses the **virtual** (unoccupied) Kohn–Sham orbitals as well. **Double hybrids** add a fraction of second-order perturbation (MP2-like) correlation, computed from the virtual orbitals, on top of a hybrid. The first widely used one, **B2PLYP** (Grimme, 2006) [92], has 53% HF exchange and 27% MP2-like correlation, with

B88 and LYP for the rest. The **random-phase approximation** (RPA) is a different fifth-rung method, which treats correlation through all virtual orbitals and naturally includes dispersion.

Fifth-rung methods are among the most accurate DFT approaches for main-group thermochemistry [43]. The price is MP2-like cost (N^5) and slow basis-set convergence (like any correlated wave-function method, Lecture 4). We will not use them in the experiments, and there are no double hybrids in Experiment 2.

6.6 Two design questions that cut across the ladder

The rung tells you *which ingredients* a functional uses. Two other questions are independent of the rung, and you need both to understand a functional's name and behavior.

6.6.1 How is dispersion included?

As we saw in Lecture 5, no semilocal functional describes long-range dispersion, and adding HF exchange does not help (HF has no dispersion at all, Lecture 4). There are four possibilities:

1. **Not at all.** Many classic functionals, including plain B3LYP, PBE and PBE0, have no long-range dispersion. They underbind dispersion-dominated complexes and get conformational energies of large flexible molecules wrong.
2. **An empirical pairwise correction, -D.** Grimme's corrections add an atom–atom term $-\sum_{AB} C_6^{AB}/R_{AB}^6$ (plus higher terms), switched off at short range by a damping function, on top of the DFT energy. The history: D2 (2006) [93], D3 (2010) [94], and the current **D4** (2017–2019) [58], [59], whose C_6 coefficients depend on the atomic partial charges and so on the chemical environment. In this course we always use **-D4**: B3LYP-D4 is B3LYP plus the D4 correction. The correction costs almost nothing.
3. **A nonlocal correlation functional, -V.** Vydrov and Van Voorhis's **VV10** [60] adds a correlation term that depends on the density at *two* points at once, $\iint \rho(\mathbf{r})\,\Phi(\mathbf{r},\mathbf{r}')\,\rho(\mathbf{r}')\,d\mathbf{r}\,d\mathbf{r}'$, so it can see two separate molecules. A "-V" at the end of a name (ωB97X-V, ωB97M-V, B97M-V) means VV10; "-rV" means the revised, faster rVV10 [95], as in B97M-rV. Note that VV10 is a correlation ingredient; it does not change the rung of the functional.
4. **Built into a fitted functional.** Highly parametrized functionals such as the Minnesota family were fitted to data that include noncovalent complexes, so they capture **medium-range** dispersion (at van der Waals contact distances) through their parameters, without any explicit term. At long range they behave like any semilocal functional. In the course data files this is labeled "fitted".

Dispersion is the topic of Lecture 8, and Experiment 4 compares these options directly.

6.6.2 Non-empirical or empirical?

The second question is how the parameters were chosen.

- **Non-empirical** (constraint-satisfying) functionals, like PBE, PBE0, TPSS and SCAN, are built to obey exact constraints of the true functional, with no fitting to molecular data [65]. Their strength is that they are expected to be reasonable for *any* system, including systems unlike anything their designers tested.
- **Empirical** (fitted) functionals, like B3LYP (3 parameters), the B97 family, the Minnesota functionals and ωB97M-V, have parameters fitted to databases of reference energies. ωB97M-V was designed by a "combinatorial" search: its authors fitted thousands of candidate forms and kept the one with the best balance between training and test sets [90]. Highly parametrized functionals can be very accurate for the kinds of systems they were trained on.

The danger of fitting is **overfitting**: a functional with many parameters can reproduce its **training set** very well and still fail for systems unlike those in the training set. A fair test of a functional uses data it was not fitted

to. This is not a reason to avoid empirical functionals (several of the best functionals available are empirical), but it is a reason to check what a functional was trained on before trusting a good result. You will meet a concrete case in Experiment 5 and in Lecture 9. In reality the line between the two groups is blurred: many "non-empirical" functionals use exact data for a few simple systems (norms), and most empirical functionals also impose some exact constraints.

6.6.3 Decoding functional names

There is no universal naming rule, but the following conventions cover most cases:

Table 17: How to read functional names. Always confirm with the original paper: names are conventions, not definitions.

Piece of the name	Meaning	Example
Exchange + correlation authors	B = Becke 88, LYP = Lee–Yang–Parr, P86 = Perdew 86, PW91 = Perdew–Wang	BLYP, BP86
A number before the correlation part	number of fitted mixing parameters of a hybrid	B3LYP, B3PW91
"0" at the end	a parameter-free hybrid	PBE0
"M" in B97M, ωB97M	meta-GGA	ωB97M-V
ω or "LC-" at the start	range-separated	ωB97M-V, LC-ωPBE
"-V", "-rV"	VV10 or rVV10 nonlocal correlation	ωB97M-V, B97M-rV
"-D4" (or -D3, -D2)	empirical dispersion correction added	B3LYP-D4, PBE-D4
Minnesota: M + year (+ suffix)	M = Minnesota, MN = Minnesota nonseparable; 06, 11, 15 = year	M06-2X, M11, MN15
"–2X" (Minnesota)	doubled HF exchange relative to the parent functional	M06-2X
"-L" (Minnesota)	local: no HF exchange	MN15-L

Four worked examples:

- **B3LYP-D4**: Becke's 3-parameter hybrid with LYP correlation (a hybrid GGA, 20% HF exchange), plus the D4 dispersion correction.
- **ωB97M-V**: range-separated (ω) hybrid built on Becke's 1997 power-series form (B97), meta-GGA (M), with VV10 nonlocal correlation (V). 15% HF exchange at short range, 100% at long range.
- **M06-2X**: Minnesota 2006, with "double" HF exchange (54%). Which rung? The name alone does not say; the paper does (it uses τ).
- **MN15-L**: Minnesota 2015, "N" for *nonseparable* (exchange and correlation are not fitted as separate pieces), "L" for local, so no HF exchange. Whether it is a GGA-type or a meta-GGA-type functional, the name does not say. You need the paper.

The last two examples carry the main message for Experiment 2: **the name is a hint, the paper is the answer.**

6.7 Experiment 2: climbing the ladder yourself

Experiment 2 has no calculations. You will classify 20 functionals into five categories: **LSDA**, **GGA**, **meta-GGA**, **hybrid GGA** and **hybrid meta-GGA** (there are no double hybrids). The point is to learn how to find this information for *any* functional, which is something you will need every time you read a computational paper or choose a method.

A reliable procedure, for each functional:

1. **Does it contain any Hartree–Fock exchange?** If yes, it is a hybrid (global or range-separated). Look for a percentage of "exact" or "HF" exchange in the paper or the Q-Chem manual. If no, it is local (rungs 1–3).
2. **Does it use the kinetic-energy density τ or the Laplacian $\nabla^2\rho$?** If yes, it is a meta-GGA (or a hybrid meta-GGA). Look for τ in the equations, or for "meta" in the abstract.
3. **Does it use the density gradient?** If yes (and not τ), it is a GGA. If it uses only the density, it is LSDA.

Note separately, for yourself, whether the functional is range-separated and how it handles dispersion. The form does not ask for these, but you will need them in Experiments 3–5.

Where to look. The best source is the original paper (the references for every functional appear on the pages of Experiments 3–5). Reviews are faster: N. Mardirossian and M. Head-Gordon [63] and L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme [43] classify hundreds of functionals with consistent notation, and R. Peverati and D. G. Truhlar [30] and H. S. Yu, S. L. Li, and D. G. Truhlar [96] explain the Minnesota family and the general landscape. The Q-Chem manual lists the functionals available in the program, with their references [1]. Avoid unsourced web pages.

Common traps.

- A functional named after its *parameters* or its *year* tells you nothing about its rung.
- "Range-separated" does not imply "hybrid" (see above).
- "-V" adds nonlocal correlation, not HF exchange, and does not move a functional up the ladder.
- "-L" means no HF exchange, but it does not tell you GGA or meta-GGA.
- MN-type ("nonseparable") functionals are classified, on this simplified ladder, with the GGAs or meta-GGAs of the same ingredients.

> ! Before next week: lock your Experiment 3 predictions
>
> Experiment 3 asks you to predict, before running anything, how methods on different rungs will perform on four "easy" organic reactions. Make those predictions with what you learned in Lectures 4–6, and lock them in before you start the calculations. Predictions are graded for being made, not for being right.

6.8 The 16 methods of Experiments 3–5 on the ladder

In Experiments 3, 4 and 5 you will run the same 16 methods: fifteen density functionals and Hartree–Fock, all with the def2-TZVP basis. Every chart in those experiments colors the methods by rung. Several of them are also in the Experiment 2 exercise, so place them on the ladder yourself first, using the procedure above. The complete table is in Lecture 7, which you should read after you submit Experiment 2.

Key points

- **Jacob's ladder** (Perdew, 2001) orders functionals by their ingredients: LSDA (ρ), GGA ($+\nabla\rho$), meta-GGA ($+\tau$ or $\nabla^2\rho$), hybrid (+ occupied orbitals via HF exchange), double hybrid/RPA (+ virtual orbitals). Higher rungs are usually more accurate and always more expensive.

- GGA exchange is LSDA exchange times an **enhancement factor** $F_x(s)$ of the reduced gradient s. **B88** (1988) has one fitted parameter; **PBE** (1996) is non-empirical, with parameters fixed by exact constraints.
- **Meta-GGAs** use the kinetic-energy density τ, which lets them recognize one-orbital regions, uniform-gas-like regions and weak-interaction regions. Examples: TPSS, SCAN, r²SCAN. They need finer grids.
- **Global hybrids** mix a fixed fraction of HF exchange: B3LYP 20%, PBE0 25%, M06-2X 54%, MN15 44%. The adiabatic connection justifies mixing; HF exchange reduces self-interaction error.
- **Range-separated hybrids** split $1/r_{12}$ with $\mathrm{erf}(\omega r_{12})$. Long-range corrected hybrids reach 100% HF exchange at long range: ωB97M-V 15% $\longrightarrow$ 100% ($\omega = 0.3$), M11 42.8% $\longrightarrow$ 100% ($\omega = 0.25$).
- **Double hybrids** (B2PLYP) add MP2-like correlation from virtual orbitals, at MP2-like cost.
- **Dispersion** is independent of the rung: none, **-D4**, nonlocal VV10 ("-V", "-rV"), or fitted (medium range only).
- Functionals are **non-empirical** (constraints) or **empirical** (fitted). Fitted functionals can be excellent, but beware of overfitting and of judging a functional on its own training set.
- **The name is a hint, the paper is the answer.**

Study questions

1. List the five rungs of Jacob's ladder and the ingredient each one adds. Why is each rung more expensive than the one below it?
2. Calculate the reduced gradient s for the hydrogen-atom density $\rho = e^{-2r}/\pi$ as a function of r. (Hint: $|\nabla\rho| = 2\rho$.) Where is s small and where is it large? Which part of Figure 12 matters most for the hydrogen atom's valence region?
3. B88 and PBE exchange agree closely for $s < 1$. Why do they still give somewhat different energies for atoms and molecules?
4. What does the variable $\alpha = (\tau - \tau_W)/\tau^{\mathrm{unif}}$ tell a meta-GGA? Give its value in a one-electron system and in the uniform gas.
5. Write the B3LYP and PBE0 formulas. Which parameters were fitted and which were derived? What fraction of HF exchange does each contain?
6. Using the adiabatic connection, explain in a few sentences why a hybrid functional should contain *some* but not 100% HF exchange.
7. Sketch the fraction of HF exchange as a function of electron–electron distance for B3LYP, for ωB97M-V and for a screened hybrid like HSE. Explain why long-range correction restores the correct asymptotic potential.
8. Decode the names B3PW91, ωB97X-V, PBE-D4 and B97M-rV: what does each part of the name mean? Which of them are hybrids, and which include dispersion, and how?
9. A colleague says: "M11 is range-separated, so it must be a hybrid; and B97M-V has a V, so it must be better at dispersion than B97M-rV." Comment on both statements.
10. Why is overfitting a concern for functionals with many parameters? How would you test whether a fitted functional really is better than a non-empirical one?

Further reading

- J. P. Perdew and K. Schmidt [64] and J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka [65]: the original Jacob's ladder papers, and a clear statement of the constraint-satisfaction philosophy.
- N. Mardirossian and M. Head-Gordon [63]: a long, careful review of 200 functionals tested on a large database, with a classification of each. The best single reference for Experiment 2.

- L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme [43]: the GMTKN55 database paper, which also ranks functionals on each rung of the ladder.
- R. Peverati and D. G. Truhlar [30]: a review of the quest for a universal density functional, with the history of the Minnesota functionals.
- W. Koch and M. C. Holthausen [8], chapters 6–7: GGAs and hybrids explained for chemists.
- A. D. Becke [97]: Becke's personal perspective on fifty years of DFT, from the man behind B88 and B3LYP.

7 How we test a functional

i Learning goals

After this lecture you should be able to:

- explain why the error of a DFT calculation on $(HCl)_2$ shrinks as the basis grows, and why the Hartree–Fock error does not;
- explain why the measured cost exponent of a calculation is smaller than the formal N^4;
- place any of the 16 methods of Experiments 3–5 on Jacob's ladder;
- describe how a benchmark database is built, where its reference values come from, and what "CCSD(T)/CBS" means;
- compute and interpret the signed error, mean signed error (MSE), mean absolute error (MAE), maximum error and relative error;
- say what "chemical accuracy" means and why a single statistic never tells the whole story;
- start Experiment 3 knowing what to run, what to record and what to look for.

This lecture has two parts. Part 1 looks back at the first two experiments. Part 2 introduces the tools we use from now on to judge a method: benchmark databases and error statistics. It ends with the briefing for Experiment 3.

7.1 Part 1 · Looking back at Experiments 1 and 2

! Read this after you submit Experiment 2

This part discusses the trends you probably saw in Experiment 1 and the classification of Experiment 2. If you have not submitted them yet, stop here and come back later.

7.1.1 Experiment 1: the basis-set error

In Experiment 1 you computed the dissociation energy of the HCl dimer,

$$\Delta E = 2\,E(\mathrm{HCl}) - E\left((\mathrm{HCl})_2\right), \tag{61}$$

with four methods (HF, PBE, B3LYP, ωB97M-V) and six basis sets from two families (def2-SVP/TZVP/QZVP and cc-pVDZ/TZ/QZ). The reference value is 2.01 kcal/mol. Because the reaction is a *dissociation*, ΔE is positive, and a **positive signed error** $\varepsilon = \Delta E_{\mathrm{comp}} - \Delta E_{\mathrm{ref}}$ means the dimer is **bound too strongly**.

7.1.1.1 What you probably saw for the density functionals

The chart *How the error changes as the basis grows* puts ζ (DZ, TZ, QZ) on the horizontal axis and |ε| on the vertical axis, one panel per family. For the three functionals the typical picture is:

- the error is largest with the double-ζ basis sets, and the dimer is usually **too strongly bound**;
- the error drops sharply from double-ζ to triple-ζ;
- the change from triple-ζ to quadruple-ζ is small. The curve flattens out.

Two effects make the double-ζ number too large. The first is the **basis-set incompleteness error** (BSIE): a small basis cannot describe the orbitals well, and the error does not cancel between the dimer and the monomers. The second, and the dominant one for a weakly bound dimer, is the **basis-set superposition error** (BSSE). In the dimer, each HCl molecule can "borrow" the basis functions centered on its partner to improve its

own description. The isolated monomers have no partner to borrow from. So the dimer is artificially stabilized, and the binding energy comes out too large. BSSE shrinks quickly as the basis grows, because a molecule that already has a nearly complete basis gains little from its neighbor's functions. We return to BSSE and the counterpoise correction in Lecture 10.

When the curve flattens, you are close to the **complete-basis-set (CBS) limit** for that functional. What is left is the error of the functional itself. This is the most important lesson of Experiment 1: **the basis-set error goes to zero as the basis grows, the functional error does not.** With a quadruple-ζ basis, the remaining error tells you how good the *method* is. For $(HCl)_2$ the functional with built-in nonlocal dispersion (ωB97M-V) usually ends up closest to the reference. Functionals with no dispersion term can converge to a value that is slightly too small (underbinding), so their $|\varepsilon|$ can go down and then up again as BSSE disappears and their real underbinding is revealed. That is why "no specific pattern" is a legitimate answer for some methods.

7.1.1.2 What you probably saw for Hartree–Fock

Hartree–Fock behaves very differently, and this is the first big surprise of the course. The HF error often does *not* decrease with the basis. It may even grow. The reason is physical. $(HCl)_2$ is held together by a mixture of dipole–dipole electrostatics and **London dispersion**. Hartree–Fock has no electron correlation at all, and dispersion is a correlation effect (Lecture 8 explains why). So at the CBS limit, HF underbinds the dimer. With a small basis, BSSE adds a spurious attraction that partly hides this missing physics. As the basis grows, the spurious attraction disappears and the true HF answer (too little binding) emerges.

In other words, HF/def2-SVP can look better than HF/def2-QZVP *for the wrong reason*. Keep this example in mind: a small basis can make a bad method look good. It is the theme of Experiment 6 and Lecture 10.

> ⚠ Common mistake: "the error got smaller, so the method got better"
>
> A smaller error is not proof of a better description. Ask *why* it got smaller. If the improvement comes from an artifact (BSSE) that happens to compensate for missing physics (dispersion), it will not transfer to another molecule.

7.1.1.3 def2 or cc-pVnZ?

The two families were built for different purposes. The Dunning correlation-consistent sets, cc-pV*n*Z [19], were designed so that the **correlation energy** of wave-function methods (MP2, CCSD(T)) converges smoothly with *n*. Each step adds a full shell of polarization functions chosen for correlation. The Karlsruhe def2 sets [17] were optimized mostly at the Hartree–Fock and DFT level, with the goal of a balanced error across the whole periodic table. Head-Gordon and co-workers found the def2 family to be the most cost-effective for DFT [18].

What you probably saw is that, at the same ζ, the two families give errors of similar size, with def2 often a bit better for the functionals at double- and triple-ζ. The differences shrink at quadruple-ζ, where both are close to the CBS limit. Note also that "triple-ζ" does not mean "the same number of functions": look at the number of basis functions you recorded for each set. Comparing basis sets by name alone is risky, and we will see this again in Experiment 6.

7.1.1.4 The cost: measured exponents

The second chart plots wall time against the number of basis functions N on a log–log scale. If the cost follows $t = a\,N^n$, then

$$\log t = \log a + n \log N, \tag{62}$$

and the slope of the straight line is the exponent *n*. In Lecture 2 you learned that the formal cost of HF and hybrid DFT grows as N^4, because the number of two-electron integrals $(\mu\nu|\lambda\sigma)$ grows as N^4. Most of you measured a smaller exponent, often between 2 and 3. Why?

1. **Integral screening.** Most of the N^4 integrals are tiny because the Gaussians are far apart. Programs estimate integrals cheaply (Schwarz inequality) and skip the negligible ones. For large enough molecules the number of significant integrals grows closer to N^2.
2. **Fixed overhead.** Small jobs spend a large share of their time on start-up, reading the basis, building the grid and writing output. This cost does not depend much on N, and it flattens the curve at small N.
3. **Other steps with different scaling.** For functionals, the exchange–correlation integration on the grid scales roughly linearly with the size of the molecule. Linear algebra (diagonalization) scales as N^3 but has a small prefactor.
4. **A narrow range.** Your N values span less than one decade. A fit over a short range is sensitive to noise and to the choice of points.

Figure 14 shows the effect of a fixed overhead on an ideal N^4 law: at small N the curve is almost flat, and only at large N does it approach the true slope. Which of your points would you trust most? The largest ones.

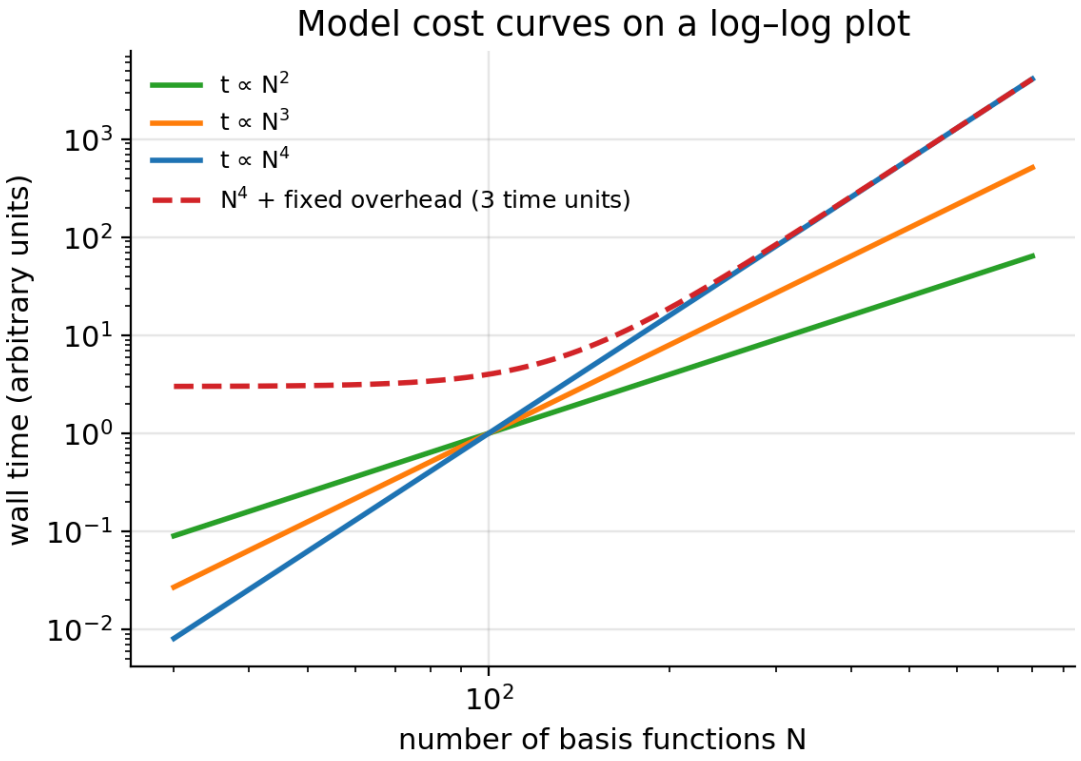


Figure 14: Model cost curves on a log–log plot. Straight lines are ideal power laws $t \propto N^n$ for n = 2, 3, 4. The dashed curve is an N^4 law plus a constant overhead: at small N it is almost flat, so a fit over a short range of N gives an exponent well below 4. All curves are analytic models, not measured timings.

7.1.1.5 The best compromise

Putting accuracy and cost together, the usual conclusion is that **triple-ζ** is the sweet spot for DFT. Double-ζ is cheap but not converged (BSIE and BSSE are still large). Quadruple-ζ is close to the CBS limit but costs several times more than triple-ζ for a small gain. This is why Experiments 3–5 use def2-TZVP. If you checked the class results (when released), the heatmap shows the same picture for the whole class.

7.1.2 Experiment 2: climbing the ladder

Experiment 2 asked you to classify 20 functionals into LSDA, GGA, meta-GGA, hybrid GGA and hybrid meta-GGA. The rung depends only on the **ingredients** of the functional (see Lecture 6):

Rung	Ingredients
LSDA	$\rho(\mathbf{r})$ only
GGA	ρ and $\nabla\rho$
meta-GGA	ρ, $\nabla\rho$ and the kinetic-energy density τ (or $\nabla^2\rho$)
hybrid GGA	GGA + a fraction of HF exchange
hybrid meta-GGA	meta-GGA + a fraction of HF exchange

The most common mistakes were these:

- **Range separation is not the same as "hybrid".** M11-L is range separated, but both ranges use *local* (semilocal) exchange. It contains no HF exchange, so it is a meta-GGA. M11 and ωB97M-V, by contrast, use HF exchange at long range (and some at short range), so they are hybrid meta-GGAs.
- **"-V" is not a rung.** The V in B97M-V and ωB97M-V stands for the VV10 nonlocal correlation term, which adds dispersion. It does not add HF exchange. B97M-V is a meta-GGA; ωB97M-V is a hybrid meta-GGA because of the ω (range-separated exchange), not because of the V.
- **"-L" means local.** M06-L, M11-L and MN15-L have no HF exchange. MN15-L is formally a meta-NGA, which sits on the meta-GGA rung in our simplified ladder.
- **Names that count parameters or exchange.** The "3" in B3LYP and B3PW91 refers to three mixing parameters. The "2X" in M06-2X means twice the HF exchange of M06 (54% vs 27%). M06-HF has 100% HF exchange and is still a hybrid meta-GGA, because it also contains a meta-GGA part.

The 16 methods you use in Experiments 3–5 sit on the ladder like this:

Table 18: The methods of Experiments 3–5. "Fitted" means the functional was parametrized on data that include noncovalent complexes.

Rung	Methods	Dispersion treatment
LSDA	SPW92	none
GGA	BP86, BLYP, PBE, PBE-D4	none; -D4 for PBE-D4
meta-GGA	SCAN, B97M-V	none; VV10 for B97M-V
hybrid GGA	B3LYP, B3LYP-D4, B3PW91, PBE0	none; -D4 for B3LYP-D4
hybrid meta-GGA	M06-2X, M11, MN15, ωB97M-V	fitted (Minnesota); VV10 for ωB97M-V
(no DFT)	HF	none

Every chart in Experiments 3–5 colors the methods by rung, so you can see whether climbing the ladder really pays off.

7.2 Part 2 · How we test a functional

7.2.1 Why functionals must be tested

The exact exchange–correlation functional exists, but nobody knows its form. Every functional you use is an approximation, and unlike wave-function methods (HF → MP2 → CCSD → CCSD(T) → …), there is no systematic path that guarantees improvement. Climbing Jacob's ladder *usually* helps, but a meta-GGA is not guaranteed to beat every GGA on every problem.

So we test. A functional is judged by how well it reproduces trusted reference values for many chemical problems. This is called **benchmarking**, and it is how the whole field decides which functionals to recommend [43], [63], [98]. In Experiments 3–5 you will do a small benchmark yourself.

7.2.2 Reference values

A benchmark is only as good as its reference values. There are two sources.

7.2.2.1 Experiment

Experimental thermochemistry (enthalpies of formation, bond energies, ionization potentials) has the advantage of being "real". But experimental numbers cannot be compared directly with a single-point electronic energy:

- experiments measure enthalpies or free energies at a finite temperature, which include zero-point vibrational energy and thermal corrections (see Lecture 13);
- some experimental values have uncertainties of 1 kcal/mol or more;
- many interesting quantities (a barrier height, the energy of an unstable conformer, a dimer at a fixed geometry) cannot be measured cleanly at all.

To compare with experiment you must back-correct the experimental value to an electronic energy difference, or add the corrections to the computed one. Both introduce extra uncertainty.

7.2.2.2 High-level theory: CCSD(T)/CBS

The alternative is to compute the reference with a method far more accurate than DFT, at the same geometries. The workhorse is **CCSD(T)**, coupled cluster with single and double excitations and a perturbative estimate of triple excitations [42]. It is called the "gold standard" of quantum chemistry because, for molecules dominated by a single electron configuration, it reproduces experiment to a few tenths of a kcal/mol. Its cost scales as N^7, so it is affordable only for small molecules.

CCSD(T) also needs a very large basis set, because the correlation energy converges slowly with the basis. The trick is to compute the energy with two or three basis sets of increasing cardinal number n (e.g., cc-pVTZ and cc-pVQZ) and **extrapolate** to the complete-basis-set (CBS) limit. A common form for the correlation energy is

$$E_{\mathrm{corr}}(n) = E_{\mathrm{corr}}^{\mathrm{CBS}} + \frac{A}{n^3}, \tag{63}$$

where A is fitted from two values of n [24]. With two points, Equation 63 can be solved for the limit:

$$E_{\mathrm{corr}}^{\mathrm{CBS}} = \frac{n^3 E_{\mathrm{corr}}(n) - (n-1)^3 E_{\mathrm{corr}}(n-1)}{n^3 - (n-1)^3}. \tag{64}$$

The HF part converges much faster and is usually taken from the largest basis or extrapolated exponentially. The combination is written **CCSD(T)/CBS**. For the most demanding thermochemistry, composite protocols

such as W4 add even higher excitations and small corrections, reaching uncertainties around 0.1–0.2 kcal/mol; the W4-17 database used for reaction 5d was built this way [99].

> Worked example: a two-point CBS extrapolation
>
> Suppose the CCSD(T) correlation energy of a small molecule is −0.27500 E_h with cc-pVTZ (n = 3) and −0.28500 E_h with cc-pVQZ (n = 4). These are made-up numbers for practice. Using Equation 64:
>
> $$E_{\text{corr}}^{\text{CBS}} = \frac{64(-0.28500) - 27(-0.27500)}{64 - 27} = \frac{-18.24000 + 7.42500}{37} = \frac{-10.81500}{37} = -0.29230\ E_h.$$
>
> The extrapolation adds −0.00730 E_h beyond the cc-pVQZ value, which is −0.00730 × 627.5095 = −4.58 kcal/mol. The basis-set error of the correlation energy is far from negligible even with a quadruple-ζ basis. For reaction energies much of it cancels, but not all.

The important point for you: the reference values in Experiments 3–5 come from CCSD(T)/CBS or better, computed at the *same* geometries you use. So the difference between your number and the reference is the error of your method (functional + basis set), with no thermal corrections or experimental uncertainty mixed in.

7.2.3 Benchmark databases

A single reaction says little about a functional. A **benchmark database** collects hundreds or thousands of reference values, grouped in **subsets** by chemical type. Three families of databases matter for this course.

7.2.3.1 GMTKN55

GMTKN55 (General Main-group Thermochemistry, Kinetics and Noncovalent interactions, 55 subsets) was published by Goerigk, Grimme and co-workers in 2017 [43]. It contains about 1500 relative energies, organized in five categories:

1. basic properties and reaction energies of small systems (atomization energies, ionization potentials, proton affinities, ...);
2. reaction energies of larger systems and isomerizations;
3. reaction barrier heights;
4. intermolecular noncovalent interactions;
5. intramolecular noncovalent interactions (conformers).

Reactions 3a and 3b come from its BHROT27 subset (rotational barriers), 3c from BUT14DIOL (butane-1,4-diol conformers), 3d from FH51 (reaction energies), and in Experiment 4 you will use DARC (Diels–Alder reactions) and ADIM6 (alkane dimers). The paper tested more than 200 functional/dispersion combinations, and it is one of the best places to look up how a functional performs.

Because the energies in GMTKN55 range from tenths of a kcal/mol (conformers) to hundreds (atomizations), a plain average would be dominated by the large ones. GMTKN55 therefore uses a **weighted** mean absolute deviation, WTMAD-2, in which the MAD of each subset i is scaled by the ratio between 56.84 kcal/mol (the average absolute reference energy over the whole database) and the average absolute reference energy of that subset, $\overline{|\Delta E|}_i$:

$$\text{WTMAD-2} = \frac{1}{\sum_i N_i} \sum_i N_i \frac{56.84\ \text{kcal/mol}}{\overline{|\Delta E|}_i} \text{MAD}_i. \tag{65}$$

Take two subsets whose average reference energies are 1 kcal/mol (conformers) and 50 kcal/mol (reaction energies). An MAD of 0.5 kcal/mol in the first then weighs as much as an MAD of 25 kcal/mol in the second, because 0.5 × 56.84/1 = 25 × 56.84/50 = 28.4. A half-kcal/mol error on a 1 kcal/mol conformer energy really is as bad, chemically, as a 50% error on a larger reaction energy.

7.2.3.2 MGCDB84

MGCDB84 (Main-Group Chemistry DataBase, 84 subsets), assembled by Mardirossian and Head-Gordon [63], contains close to 5000 data points. It grew out of the data used to develop the ωB97X-V, B97M-V and ωB97M-V functionals, and the same paper uses it to evaluate 200 functionals. Its categories distinguish "easy" from "difficult" problems: noncovalent dimers and clusters, isomerization energies, thermochemistry and barrier heights, each split into easy and difficult groups. The AlkIsomer11 and S66 subsets of Experiment 4 come from it.

7.2.3.3 The Minnesota databases

Truhlar's group has built a long series of databases to develop and test the Minnesota functionals (M06, M11, MN15 and others). The latest, Minnesota Database 2015B, was used to parametrize MN15 [30], [31]. It includes many inorganic and transition-metal systems, and multireference cases. Reactions 5b (CaO) and 5c (H_2SO_4) come from it, and Experiment 1 came from its NCCE30 subset.

7.2.3.4 Training sets and test sets

Many modern functionals contain parameters fitted to a database. The data used for fitting form the **training set**. Data never used for fitting form the **test set**. A functional's performance on its own training set is not an independent test: it is guaranteed to be good there, by construction. A fair benchmark looks at the test set, or at a database built by someone else. Keep this in mind for Lecture 9: one reaction in Experiment 5 belongs to the training set of one of your functionals.

7.2.4 Error statistics

You will compute these quantities in every experiment from now on. They are also defined in the units and errors guide.

7.2.4.1 The signed error

For one reaction k,

$$\varepsilon_k = \Delta E_k^{\mathrm{comp}} - \Delta E_k^{\mathrm{ref}}. \tag{66}$$

The sign carries meaning. If ε > 0, the computed reaction energy is *more positive* (less negative) than the reference. For an exothermic reaction (ΔE < 0), a positive ε means the method underestimates how favorable the reaction is. For a dissociation (ΔE > 0), a positive ε means the method overbinds. **Always write down the direction of the reaction before interpreting the sign.**

7.2.4.2 Statistics over a set of N reactions

$$\mathrm{MSE} = \frac{1}{N}\sum_{k=1}^{N} \varepsilon_k, \qquad \mathrm{MAE} = \frac{1}{N}\sum_{k=1}^{N} |\varepsilon_k|, \tag{67}$$

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{k=1}^{N} \varepsilon_k^2}, \qquad \mathrm{MaxE} = \max_k |\varepsilon_k|. \tag{68}$$

- The **mean signed error** (MSE) measures *systematic bias*. A method with MSE ≈ 0 is not necessarily good: large positive and negative errors can cancel.
- The **mean absolute error** (MAE, sometimes MAD or MUE) measures *typical accuracy*. It is the number most benchmark papers report.
- The **root-mean-square deviation** (RMSD) weights large errors more heavily. RMSD ≥ MAE always; a RMSD much larger than the MAE warns you of outliers.
- The **maximum error** tells you the worst case. If you design a drug or a catalyst, the worst case may matter more than the average.

7.2.4.3 The relative error

$$\varepsilon_k^{\mathrm{rel}} = \frac{\varepsilon_k}{|\Delta E_k^{\mathrm{ref}}|} \times 100\%. \tag{69}$$

The relative error puts reactions of very different size on the same footing. But chemistry is decided by *absolute* energies: an equilibrium constant depends on ΔG in kcal/mol, not on a percentage. A 1% error on an atomization energy of 600 kcal/mol is 6 kcal/mol, which changes a rate or an equilibrium by orders of magnitude. And a relative error is meaningless when the reference is close to zero. Experiment 5 has a chart for exactly this comparison.

> Worked example: statistics for a hypothetical method
>
> A hypothetical method gives these reaction energies (kcal/mol) for four reactions. The numbers are invented for practice; they are not results for any real functional.
>
> | Reaction | ΔE_{ref} | ΔE_{comp} | ε | $\lvert\varepsilon\rvert$ |
> |---|---|---|---|---|
> | A | 2.50 | 2.10 | −0.40 | 0.40 |
> | B | 1.00 | 1.30 | +0.30 | 0.30 |
> | C | −10.00 | −8.80 | +1.20 | 1.20 |
> | D | 50.00 | 47.60 | −2.40 | 2.40 |
>
> - MSE = (−0.40 + 0.30 + 1.20 − 2.40)/4 = −1.30/4 = −0.33 kcal/mol (−0.325 before rounding).
> - MAE = (0.40 + 0.30 + 1.20 + 2.40)/4 = 4.30/4 = 1.08 kcal/mol (1.075).
> - RMSD = √[(0.16 + 0.09 + 1.44 + 5.76)/4] = √(7.45/4) = √1.8625 = 1.36 kcal/mol.
> - MaxE = 2.40 kcal/mol (reaction D).
> - Relative errors: A 16%, B 30%, C 12%, D 4.8%.
>
> Reaction D has the largest absolute error but the smallest relative error. Reaction B has the smallest absolute error but the largest relative error. Which one is “worst” depends on the question you are asking.

⚠ Common mistakes with error statistics

- Averaging signed errors and calling the result "the accuracy". That is the MSE, a measure of bias.
- Rounding the total energies before taking differences. Keep all the digits of the total energies (in E_h), convert the *difference* to kcal/mol with 1 E_h = 627.5095 kcal/mol, and round only at the end.
- Forgetting the stoichiometric coefficients. In $\Delta E = \Sigma\, \nu_i E_i$ the coefficients are signed: $\nu > 0$ for products, $\nu < 0$ for reactants. "2 ethane ⟶ dimer" has $\nu = -2$ for ethane.
- Comparing MAEs computed over different sets of reactions.

7.2.5 Chemical accuracy

The traditional target for a computational method is **chemical accuracy**, an error of about **1 kcal/mol** (4.2 kJ/mol). The number is not arbitrary. At 298 K, RT = 0.5925 kcal/mol, and an error δ in a free-energy difference changes an equilibrium constant (or a rate constant, through the barrier) by a factor exp(δ/RT):

$$\frac{K_{\text{comp}}}{K_{\text{true}}} = e^{\delta/RT} = e^{1.000/0.5925} = e^{1.688} \approx 5.4. \tag{70}$$

So a 1 kcal/mol error means a factor of about 5 in K or in a rate. That is often good enough to tell which product dominates. A 3 kcal/mol error means a factor of $e^{5.06} \approx 160$, which is often not.

For reference: on GMTKN55, the best hybrid and double-hybrid functionals reach a WTMAD-2 of a few kcal/mol, and the best functionals have MAEs well below 1 kcal/mol for easy noncovalent and conformational data, but several kcal/mol for difficult thermochemistry [43], [63]. Chemical accuracy is not reached uniformly by any functional.

7.2.6 Why a single number never tells the whole story

Benchmark rankings are useful, but they must be read with care:

1. **The composition of the database matters.** A ranking depends on which reactions are included and how they are weighted. A functional that is excellent for organic chemistry may be mediocre for transition metals, and an overall average hides this.
2. **Training sets.** A functional fitted to a database will look good on that database (see above).
3. **Error cancellation.** A good number can come from a lucky cancellation between two errors (Experiment 1 already showed one). Lucky cancellations don't transfer.
4. **Outliers.** A small MAE can coexist with a catastrophic failure on one system. Always look at the maximum error and at the distribution.
5. **The reference itself.** CCSD(T)/CBS is excellent for single-reference molecules, but less reliable for systems with strong static correlation (Lecture 9).

The practical advice: before using a functional for a new problem, look up its performance on the *subsets* most similar to your problem, not just its overall rank.

7.3 Experiment 3 briefing

7.3.1 The goal

Experiment 3 is your first benchmark. You will test 16 methods on four reactions that appear in every organic chemistry textbook:

Reaction	Description	ΔE_{ref} (kcal/mol)
3a	ethane, staggered ⟶ eclipsed (rotational barrier)	2.73
3b	methanol, staggered ⟶ eclipsed (rotational barrier)	1.01
3c	butane-1,4-diol, conformer B1 ⟶ B43 (breaks an intramolecular H-bond)	3.29
3d	*cis*-pent-2-ene ⟶ *trans*-pent-2-ene	−1.15

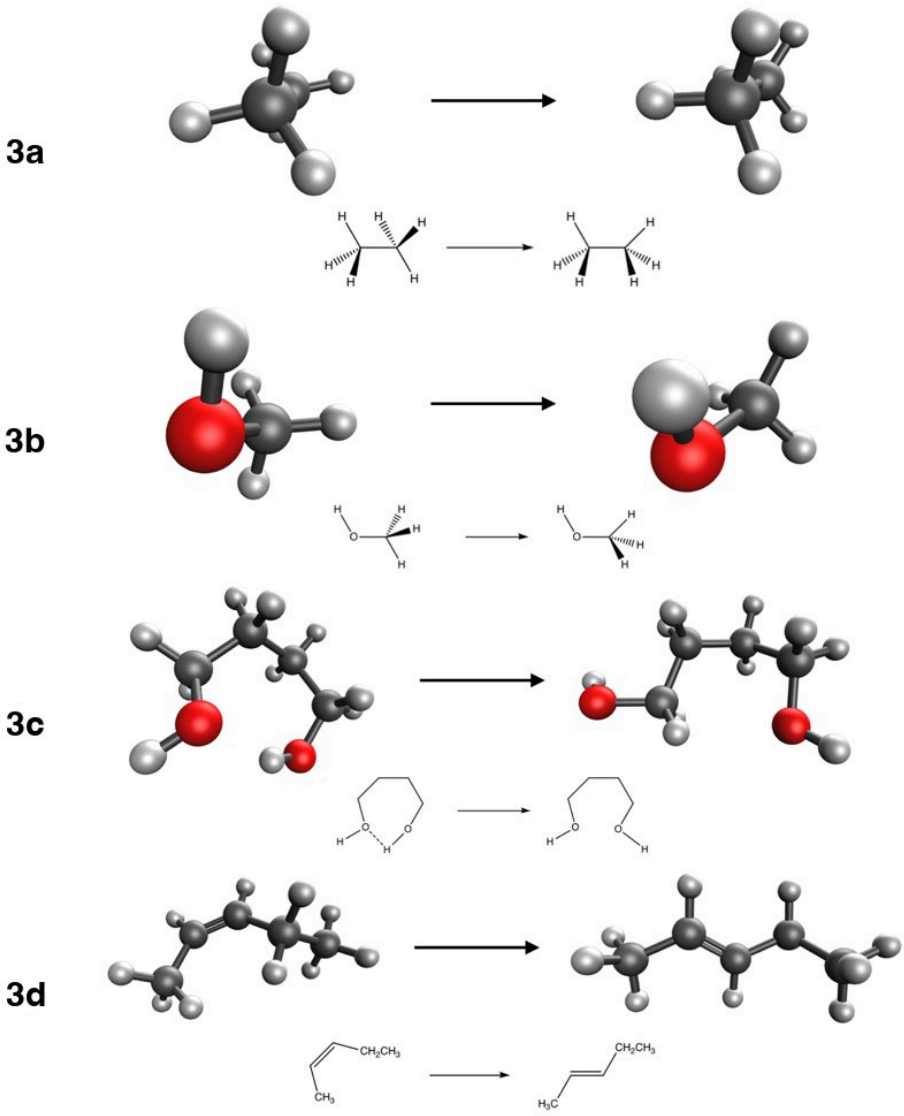


Figure 15: Reactions of Experiment 3: rotational barriers of ethane (3a) and methanol (3b), a conformational change of butane-1,4-diol (3c), and the cis–trans isomerization of pent-2-ene (3d).

7.3.2 Methods and jobs

You will use the 16 methods of Table 18 with the def2-TZVP basis set. Eight molecules × 16 methods = **128 jobs**. Each one is small and fast, but there are many, so submit them in batches (see working on a shared machine). For every job record the total energy, the number of SCF cycles and the wall time.

7.3.3 Why these reactions are "easy"

All four reactions compare two forms of the same molecule (two conformers, or a cis and a trans isomer). The atoms, the bonds and the number of electron pairs are the same on both sides. Only the arrangement in space changes. Most errors of an approximate functional are tied to specific features of the electron density: the core, the lone pairs, each type of bond. When the same features appear on both sides of the reaction, their errors largely **cancel** when you take products minus reactants.

This idea has a name in thermochemistry. An **isodesmic reaction** conserves the number and type of each bond, and a **homodesmotic** reaction also conserves the hybridization of the atoms. Chemists design such reactions on purpose, because they make the computed reaction energy much less sensitive to the method. Conformational changes are the extreme case: nothing changes but the geometry.

Error cancellation is not guaranteed, though. Ask yourself, for each reaction, what *does* change. In 3c an intramolecular hydrogen bond is broken. In 3a and 3b the barrier comes from a subtle balance of steric and hyperconjugative effects. How well does each rung describe a hydrogen bond? How does a method with no electron correlation at all (HF) do?

7.3.4 Cost per SCF cycle

In Experiment 3 you also measure cost. Different functionals need different numbers of SCF cycles to converge, so the total time mixes "how expensive is one iteration" with "how hard was it to converge". Dividing the wall time by the number of SCF cycles isolates the cost of one iteration. Think about what each rung adds to that cost:

- a GGA or meta-GGA needs the density (and its gradient, and τ) on a numerical grid;
- a hybrid also needs the HF exchange matrix K, built from the two-electron integrals, which is usually the most expensive step;
- a range-separated hybrid needs extra integrals for the long-range part;
- a nonlocal VV10 term needs a double integral over the grid.

The site computes the time per cycle for you and plots it against the MAE in the *cost vs accuracy* chart. The ideal method sits in the bottom-left corner: cheap and accurate.

7.3.5 What to watch out for

- Read the reaction direction in the table: ΔE = E(product) − E(reactant). 3d is negative.
- These reaction energies are small (1–3 kcal/mol). Keep every digit of the total energies.
- Check that each job finished ("Thank you very much for using Q-Chem") and that the SCF converged. Re-run any job that failed.
- Lock in your predictions on the experiment page *before* you run anything, including the "functional draft": the method you bet will win over Experiments 3, 4 and 5.

7.3.6 Key points

- In Experiment 1 the basis-set error (BSIE + BSSE) of DFT shrinks quickly with ζ; what remains at quadruple-ζ is the error of the functional.

- Hartree–Fock has no dispersion. Its apparent accuracy with small basis sets for $(HCl)_2$ comes from BSSE, an artifact that disappears as the basis grows.
- Measured cost exponents are below the formal N^4 because of integral screening, fixed overhead and the short range of N. Triple-ζ is the usual sweet spot for DFT.
- The rung of a functional depends only on its ingredients. Range separation, "-V" and "-D4" are not rungs.
- Functionals are judged by benchmarking against reference data, usually CCSD(T)/CBS, collected in databases such as GMTKN55, MGCDB84 and the Minnesota databases.
- Signed error, MSE, MAE, RMSD, maximum error and relative error each answer a different question. Use more than one.
- Chemical accuracy (1 kcal/mol) corresponds to a factor of about 5 in an equilibrium or rate constant at room temperature.
- Performance on a functional's own training set is not an independent test.

7.3.7 Study questions

1. In Experiment 1, the reaction is a dissociation. Explain in one sentence what a positive signed error means for the binding of $(HCl)_2$.
2. Why does BSSE make dimers too stable and not too unstable? Why does it shrink as the basis grows?
3. Explain why the HF error for $(HCl)_2$ can grow when going from def2-SVP to def2-QZVP. What does this tell you about judging a method by a single small-basis calculation?
4. A student measures time $\propto N^{2.3}$ for a hybrid functional. Give three reasons why the exponent is smaller than 4.
5. Classify M11-L, B97M-V, ωB97M-V and M06-HF on Jacob's ladder, and explain the trap in each name.
6. Why are CCSD(T)/CBS values preferred over experimental values as references for benchmarking functionals? Give two reasons.
7. Using Equation 64, extrapolate correlation energies of $-0.2000\ E_h$ ($n = 2$) and $-0.2300\ E_h$ ($n = 3$). How large is the extrapolation correction, in kcal/mol?
8. A method has MSE = 0.0 kcal/mol and MAE = 3.0 kcal/mol on a set of reactions. Describe what its errors look like.
9. Show that an error of 2 kcal/mol in ΔG changes an equilibrium constant at 298 K by a factor of about 29.
10. Why might a functional rank first on one database and tenth on another?

7.3.8 Further reading

- L. Goerigk et al., "A look at the density functional theory zoo with the advanced GMTKN55 database" [43]. The database, its categories and WTMAD-2, with a ranking of more than 200 methods.
- N. Mardirossian and M. Head-Gordon, "Thirty years of density functional theory in computational chemistry" [63]. MGCDB84, and a review of 200 functionals.
- L. Goerigk and N. Mehta, "A trip to the density functional theory zoo" [98]. A readable guide to choosing functionals.
- P. Morgante and R. Peverati, "The devil in the details" [32]. The tutorial review on which this lab is based.
- The course guide on units and errors.

8 Dispersion and noncovalent interactions

ℹ Learning goals

After this lecture you should be able to:

- explain why most methods give small errors for the reactions of Experiment 3, and why LSDA and Hartree–Fock are the usual exceptions;
- read the cost-vs-accuracy chart and explain why hybrids cost more per SCF cycle than (meta-)GGAs;
- derive the $-C_6/R^6$ form of the London dispersion energy from a model of two coupled oscillators;
- explain why semilocal functionals and Hartree–Fock cannot describe dispersion;
- describe how the -D4 correction works (pairwise C_6 terms, damping, charge dependence, three-body term) and how nonlocal VV10/rVV10 correlation differs from it;
- explain why intramolecular dispersion matters for branched alkanes;
- start Experiment 4 knowing what each reaction tests.

8.1 Part 1 · Looking back at Experiment 3

! Read this after you submit Experiment 3

This part discusses the trends you probably saw in Experiment 3. If you have not submitted it yet, stop here and come back later.

8.1.1 The "easy" reactions really were easy

In Experiment 3 you computed four small energy differences: two rotational barriers (3a ethane, 2.73 kcal/mol; 3b methanol, 1.01 kcal/mol), a conformational change of butane-1,4-diol that breaks an intramolecular hydrogen bond (3c, 3.29 kcal/mol), and the *cis*⟶*trans* isomerization of pent-2-ene (3d, −1.15 kcal/mol). The 16 methods range from LSDA to hybrid meta-GGAs, plus Hartree–Fock.

What you probably saw on the *signed error* chart is a sea of very short bars. For 3a, 3b and 3d, nearly every method lands within a few tenths of a kcal/mol of the reference, whatever its rung. The ranking among the best methods changes from reaction to reaction, and the differences between them are often smaller than 0.1–0.2 kcal/mol. Differences that small are close to the uncertainty of the reference values and of your own numerical settings, so they should not be over-interpreted.

8.1.1.1 Why: error cancellation

The explanation is the one we anticipated in Lecture 7. Each reaction compares two forms of the *same* molecule. The number of electrons, the bonds, the lone pairs and the core orbitals are identical on both sides. An approximate functional makes errors in describing each of these features, but it makes nearly the same errors in the reactant and in the product. When you compute ΔE = E(product) − E(reactant), those errors cancel. What remains is the error in describing the small *change* in the density, and that change is small and local.

This is the most powerful idea in practical quantum chemistry. It is also fragile, because it depends on the reaction, not on the method. Nobody can tell in advance, from the reaction type alone, whether a conformational change will be easy. That is why we benchmark.

8.1.1.2 The usual outliers: butane-1,4-diol, LSDA and HF

Reaction 3c is the one where the sea of short bars breaks up. In conformer B1 the two OH groups form an intramolecular hydrogen bond; in B43 they do not. So 3c measures, roughly, the energy of one hydrogen bond plus some conformational strain. A change in a noncovalent interaction does not cancel.

The two methods that most often stand out are at the two ends of the ladder:

- **LSDA (SPW92)** is known to **overbind** hydrogen bonds and, more generally, to give too much attraction wherever two densities overlap weakly. The uniform-electron-gas model has no way to tell a weak, stretched interaction from a real bond. For 3c this makes the hydrogen-bonded conformer too stable, so the energy to break the H-bond comes out too large.
- **Hartree–Fock** has no correlation. A hydrogen bond is mostly electrostatics, which HF describes reasonably, but the correlation and dispersion parts are missing, so HF tends to **underbind**. Its error on 3c has the opposite sign to LSDA's.

The modern hybrid and meta-GGA functionals usually describe the hydrogen bond well. You may also have noticed that -D4 changed the numbers in Experiment 3 very little. The dispersion energy depends on interatomic distances, and the distances barely change in a rotation or in a cis–trans isomerization, so the -D4 contribution nearly cancels too. In 3c, where the molecule changes shape more, -D4 makes a slightly larger difference.

8.1.2 Cost vs accuracy

The *cost vs accuracy* chart plots the MAE against the time per SCF cycle on a log axis. You probably saw something like this:

- **SPW92 and the GGAs** are the cheapest per cycle. They need only the density and its gradient on the integration grid.
- **Meta-GGAs** cost slightly more: they need τ on the grid too, and are often run on finer grids (Lecture 11).
- **Hybrids** are clearly more expensive per cycle. They need the HF exchange matrix K, built from the four-index two-electron integrals, and that step dominates.
- **Range-separated hybrids** (ωB97M-V, M11) need extra integrals for the long-range exchange, and **VV10** (B97M-V, ωB97M-V) adds a double integral over grid points. These are usually the most expensive per cycle.
- **HF** needs K but no grid at all, so it costs about the same as a hybrid.
- **B3LYP is not special.** It costs about the same as B3PW91 and PBE0, because cost is set by the rung (and by the HF exchange), not by the name.

The ideal method sits bottom-left. For Experiment 3, cheap methods are almost as accurate as expensive ones, because error cancellation does most of the work. The ranking will look very different in Experiments 4 and 5. Keep your chart: comparing it with the next two is the real lesson.

> ⚠ Common mistake: generalizing from easy reactions
>
> "SPW92 was fine for the ethane barrier, so LSDA is good enough" is wrong. Experiment 3 tells you that *conformational energies with similar bonding* are easy. It says almost nothing about reactions where bonds or noncovalent contacts change.

8.2 Part 2 · Dispersion and noncovalent interactions

8.2.1 Noncovalent interactions

Noncovalent interactions are the forces between molecules (and between distant parts of the same molecule) that do not involve sharing electrons in a bond. They are weak, typically 0.1–10 kcal/mol per contact, but they decide the structure of DNA, the folding of proteins, the packing of crystals, and the binding of a drug to its target. The total interaction energy is often split into four physical contributions:

1. **electrostatics**: the classical interaction between the charge distributions of the two monomers (charges, dipoles, quadrupoles);
2. **exchange (Pauli) repulsion**: electrons with the same spin cannot occupy the same region; it grows roughly exponentially as the densities overlap;
3. **induction (polarization)**: each monomer's density is distorted by the field of the other;
4. **dispersion**: the attraction between correlated, instantaneous fluctuations of the electron clouds.

A hydrogen bond, like the one in the methanol dimer, is dominated by electrostatics, with important induction and dispersion parts. The interaction between two alkanes, like the ethane dimer, is almost purely dispersion balanced by Pauli repulsion: ethane has no dipole moment and no hydrogen-bond donors or acceptors.

8.2.2 The physics of London dispersion

8.2.2.1 Fluctuating dipoles

Even a perfectly spherical atom, like argon, has an instantaneous dipole moment at every moment, because its electrons move. On average the dipole is zero. But an instantaneous dipole on atom A creates an electric field at atom B, which induces a dipole on B. The two dipoles are aligned so that they attract. The motions of the electrons on A and B become **correlated**, and the correlation lowers the energy. This is the picture Fritz London gave in 1930 [100]: dispersion is an electron correlation effect between two distant regions of space.

8.2.2.2 A model: two coupled oscillators

The simplest model that captures the physics is the **Drude model**. Represent each atom by a charge −q bound to a nucleus by a spring of force constant $k = m\omega_0^2$, so that it oscillates with frequency ω_0. The static polarizability of such an oscillator is $\alpha = q^2/k$. Put two of these oscillators a distance R apart, and let the charges move only along the line that joins them (a one-dimensional model). The dipole–dipole interaction between the displacements x_1 and x_2 is

$$V = -\frac{2q^2 x_1 x_2}{R^3}. \tag{71}$$

The Hamiltonian of the two coupled oscillators can be separated into two normal modes, $x_1 \pm x_2$, with frequencies

$$\omega_\pm = \omega_0 \sqrt{1 \mp \frac{2\alpha}{R^3}}. \tag{72}$$

Quantum mechanics says each oscillator keeps a zero-point energy $\hbar\omega/2$ even in its ground state. The total zero-point energy of the pair is $(\hbar/2)(\omega_+ + \omega_-)$. Expanding the square roots for large R with $\sqrt{(1 + x)} \approx 1 + x/2 - x^2/8$, the first-order terms cancel and the second-order terms do not:

$$\frac{\hbar}{2}(\omega_+ + \omega_-) \approx \hbar\omega_0 - \frac{\hbar\omega_0}{2}\,\frac{\alpha^2}{R^6}. \tag{73}$$

The first term is just the energy of two separate oscillators. The second is the interaction:

$$E_{\text{disp}} = -\frac{\hbar\omega_0\,\alpha^2}{2R^6} \quad \text{(1D model)}, \qquad E_{\text{disp}} = -\frac{3\hbar\omega_0\,\alpha^2}{4R^6} \quad \text{(3D model)}\ . \tag{74}$$

Three lessons come out of this simple calculation.

1. Dispersion is **always attractive**.
2. It decays as $\mathbf{R^{-6}}$. We write $E_{\text{disp}} = -C_6/R^6$, with a **dispersion coefficient** C_6.
3. It is a **quantum** effect: it comes from the change in zero-point energy. Classical static charge distributions with no dipole would not interact at all.

The model also tells us what C_6 depends on: the **polarizability** α (how easily the electron cloud is distorted) and a characteristic excitation energy $\hbar\omega_0$. London approximated $\hbar\omega_0$ by the ionization energy I, which gives the famous **London formula** for two different atoms:

$$C_6^{AB} \approx \frac{3}{2}\,\frac{I_A I_B}{I_A + I_B}\,\alpha_A \alpha_B. \tag{75}$$

The exact expression, due to Casimir and Polder [101], integrates the frequency-dependent polarizabilities at imaginary frequency:

$$C_6^{AB} = \frac{3}{\pi}\int_0^\infty \alpha_A(i\omega)\,\alpha_B(i\omega)\,d\omega. \tag{76}$$

Large, soft, polarizable atoms (Xe, I, S) have large C_6; small, hard ones (He, Ne, H) have small C_6. That is why boiling points rise down a group, and why long alkanes are liquids or solids while methane is a gas.

> **Worked example: the size of dispersion between two argon atoms**
>
> The C_6 coefficient of the argon dimer is about 64.3 $E_h\,a_0^6$. Convert it to kcal/mol Å^6 using 1 E_h = 627.5095 kcal/mol and 1 a_0 = 0.529177 Å, so 1 a_0^6 = 0.529177^6 Å^6 = 0.021958 Å^6:
>
> $$C_6 = 64.3 \times 627.5095 \times 0.021958\ \text{kcal mol}^{-1}\,\text{Å}^6 = 886\ \text{kcal mol}^{-1}\,\text{Å}^6.$$
>
> At the equilibrium distance of Ar_2, R ≈ 3.76 Å, $R^6 = 3.76^6 = 2.83 \times 10^3$ Å^6, and
>
> $$-\frac{C_6}{R^6} = -\frac{886}{2.83\times 10^3} = -0.31\ \text{kcal/mol}\ .$$
>
> The full well depth of Ar_2 is about 0.28 kcal/mol (Experiment 7 studies this curve). The dispersion attraction is slightly larger than the well depth, because Pauli repulsion already cancels part of it at that distance.

Figure 16 (a) shows a Lennard-Jones model of a rare-gas dimer. The minimum arises from the balance of a steep repulsion and the R^{-6} attraction. Without the attraction there is no minimum at all: the curve is purely repulsive.

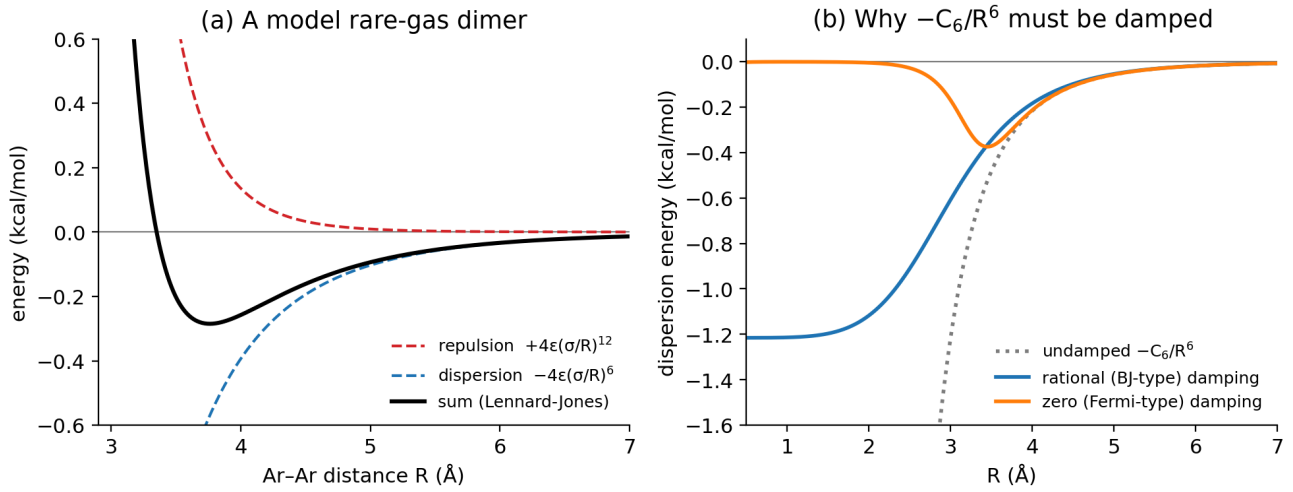


Figure 16: (a) Lennard-Jones model of Ar_2 (ε = 0.285 kcal/mol, σ = 3.35 Å), split into its repulsive R^{-12} and attractive R^{-6} terms. (b) The dispersion energy $-C_6/R^6$ for Ar–Ar (C_6 = 64.3 E_h a_0^6) without damping, with rational (Becke–Johnson-type) damping and with zero (Fermi-type) damping. The damping radius is illustrative, not the -D4 value. All curves are analytic models.

8.2.3 Why semilocal DFT and Hartree–Fock miss dispersion

8.2.3.1 Semilocal functionals

A GGA writes the exchange–correlation energy as a single integral over space:

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[\rho] = \int f\left(\rho(\mathbf{r}), \nabla\rho(\mathbf{r})\right) d\mathbf{r}. \tag{77}$$

The energy density at each point depends only on the density and its gradient *at that point*. A meta-GGA adds $\tau(\mathbf{r})$, still at the same point. Now consider two argon atoms far apart, so that their densities do not overlap. At every point in space, the density is either that of atom A or that of atom B (or zero). The integral splits into a sum of two atomic integrals, and the interaction energy is exactly zero. As the atoms approach and the densities start to overlap, the semilocal functional "sees" the overlap, but this effect decays **exponentially** with distance, like the density tails, not as R^{-6}.

Dispersion is a correlation between fluctuations at $\mathbf{r}$ in atom A and at $\mathbf{r}'$ in atom B, which may be far apart. A functional that looks only at one point at a time cannot know about it. This is a limitation of the *form* of the functional, not of its parameters.

Different semilocal functionals fail in different ways in the overlap region, where they can produce *spurious* binding or *spurious* repulsion. LSDA tends to overbind weakly interacting systems. PBE gives a little binding for some dispersion-bound complexes. B88 exchange (in BLYP, B3LYP) is quite repulsive, so BLYP and B3LYP often predict little or no binding for rare-gas and alkane dimers. None of these behaviors comes from real dispersion: they are side effects of how the exchange functional behaves where the density is small and varies quickly. Getting binding from LSDA is "right for the wrong reason".

8.2.3.2 Hybrids

Hybrid functionals add a fraction of HF exchange, which is nonlocal. But dispersion is a *correlation* effect, and the correlation part of a hybrid is still semilocal. Hybrids miss dispersion just as GGAs do.

8.2.3.3 Hartree–Fock

Hartree–Fock describes each electron in the average field of the others. Instantaneous, correlated fluctuations are exactly what the mean field leaves out, so HF has **no dispersion at all**. HF describes electrostatics, induction and Pauli repulsion, so it binds hydrogen-bonded or ionic complexes partially, but rare-gas and alkane dimers come out unbound. Experiment 1 already showed this for $(HCl)_2$.

8.2.3.4 Wave-function methods

Correlated wave-function methods include dispersion naturally, because they include excitations on both monomers at once. MP2 includes it at second order (and tends to overestimate it for π stacks). CCSD(T) describes it accurately, which is why it provides the reference values for noncovalent databases such as S66 [102], [103].

8.2.4 How to put dispersion back into DFT

Three strategies are in use [104], [105]. Experiment 4 compares all three.

8.2.4.1 Strategy 1: pairwise corrections (-D)

The simplest idea is to add the London formula, atom pair by atom pair, to the DFT energy:

$$E_{\mathrm{tot}} = E_{\mathrm{DFT}} + E_{\mathrm{disp}}, \qquad E_{\mathrm{disp}} = -\sum_{A<B}\sum_{n=6,8} s_n \frac{C_n^{AB}}{R_{AB}^n} f_{\mathrm{damp},n}(R_{AB}). \tag{78}$$

Here R_{AB} is the distance between atoms A and B, C_6^{AB} and C_8^{AB} are dispersion coefficients (the C_8/R^8 term describes dipole–quadrupole fluctuations and matters at medium range), s_n are scaling factors fitted for each functional, and f_{damp} is a **damping function**. The correction depends only on the geometry, so it costs almost nothing and does not change the density or the orbitals. It adds energy (and forces, for geometry optimizations).

Why damping? The bare $-C_6/R^6$ diverges at short distance (Figure 16 b). It would make molecules collapse. Moreover, at short and medium range the functional already describes some correlation; adding the full $-C_6/R^6$ there would count it twice. The damping function switches the correction off where the functional is already doing the job. Two forms are common:

- **zero damping** sends the correction to zero at short range. D2 used a Fermi-type function of R/R_0, and the original D3 a similar "zero-damping" function.
- **rational (Becke–Johnson, BJ) damping** replaces R^6 by $R^6 + R_0^6$, so the correction tends to a finite constant at short range [105]:

$$f_{\mathrm{damp,n}}^{\mathrm{BJ}}(R) = \frac{R^n}{R^n + (a_1 R_0^{AB} + a_2)^n}, \qquad R_0^{AB} = \sqrt{C_8^{AB}/C_6^{AB}}. \tag{79}$$

The parameters a_1, a_2 and s_8 are fitted for each functional, because each functional has its own amount of spurious short-range binding or repulsion. That is why the same -D4 correction gives different numbers on PBE and on B3LYP.

A short history. Grimme's first corrections (D, 2004, and D2, 2006 [93]) used fixed C_6 for each element. But the C_6 of carbon depends on its environment: an sp^3 carbon in an alkane is less polarizable than an sp^2 carbon in benzene. D3 (2010) made C_6 depend on the **coordination number** of each atom, computed from the geometry [94]. D3(BJ) added rational damping [105].

-D4, the version used in this course [58], [59], goes one step further. Its main ingredients are:

- **reference polarizabilities.** For each element, D4 stores frequency-dependent polarizabilities $\alpha(i\omega)$ of reference systems with different coordination numbers.
- **coordination and charge.** For each atom in your molecule, D4 interpolates between these references using the atom's coordination number *and its partial charge*. The charges come from a fast electronegativity-equilibration model that uses only the geometry. An anion-like atom is more polarizable (larger C_6) than a cation-like one, which D3 could not capture.
- **Casimir–Polder integration.** The atom-in-molecule $\alpha(i\omega)$ are combined with Equation 76 to give C_6^{AB} for every pair.
- **a three-body term.** The Axilrod–Teller–Muto term describes the dispersion among three atoms at once. It is usually repulsive for compact arrangements and matters for large molecules and molecular crystals.

In Q-Chem, -D4 is requested with `DFT_D = D4` on top of the functional (e.g., `METHOD = B3LYP`). The site's input generator does it for you.

8.2.4.2 Strategy 2: nonlocal correlation functionals (VV10, rVV10)

The second strategy is to fix the functional itself, by adding a correlation term that couples two points of space:

$$E_c^{nl} = \frac{1}{2}\iint \rho(\mathbf{r})\,\Phi(\mathbf{r},\mathbf{r}')\,\rho(\mathbf{r}')\,d\mathbf{r}\,d\mathbf{r}' + \beta N. \tag{80}$$

The **kernel** Φ depends on the distance $|\mathbf{r}-\mathbf{r}'|$ and on the local density and its gradient at both points. The first functional of this kind was vdW-DF by Dion, Lundqvist and co-workers [106], [107]. Vydrov and Van Voorhis built the simpler and more accurate **VV10** [60]. Its kernel has the form

$$\Phi^{\mathrm{VV10}} = -\frac{3}{2\,g\,g'\,(g+g')}, \qquad g = \omega_0(\mathbf{r})\,R^2 + \kappa(\mathbf{r}), \quad R = |\mathbf{r}-\mathbf{r}'|, \tag{81}$$

where $\omega_0(\mathbf{r})$ is a local excitation frequency built from the density and its gradient, and $\kappa(\mathbf{r})$ controls the short-range behavior. For large R, $g \approx \omega_0 R^2$, and $\Phi \approx -3/(2\omega_0\omega_0'(\omega_0+\omega_0')R^6)$: the R^{-6} law is recovered, with C_6 coefficients that come from the density itself, not from a table. The kernel has two parameters, b and C; b in particular is adjusted for each host functional. The constant β is chosen so that the term vanishes for the uniform electron gas.

rVV10 [95] is a revised form of the kernel that can be evaluated much faster (via a Fourier-space interpolation), with nearly the same accuracy. B97M-rV [77] is B97M-V with rVV10 in place of VV10.

Compared with -D4, nonlocal functionals respond to the actual electron density (they "know" about charge transfer or unusual bonding), at the price of an extra double integral over the grid. This is why B97M-V and ωB97M-V cost more per SCF cycle.

8.2.4.3 Strategy 3: fitted functionals

The third strategy is to include noncovalent complexes in the training set of a highly flexible functional, and let the fit find parameters that reproduce their binding energies. Many Minnesota functionals (M06-2X, M11, MN15) work this way [31], [85], [89]. They describe **medium-range** dispersion (the region around the equilibrium distance of most complexes) surprisingly well. But they are still semilocal (plus HF exchange) in their correlation, so they have no true R^{-6} tail at long range. Dispersion corrections for some Minnesota functionals have been proposed [108], [109], but they are small by design.

Strategy	Example in this course	Long-range R^{-6}	Depends on density?	Extra cost
none	PBE, B3LYP, BLYP	no	n/a	none
pairwise -D4	PBE-D4, B3LYP-D4	yes	no (geometry and model charges)	negligible
nonlocal VV10/ rVV10	B97M-V, ωB97M-V, B97M-rV	yes	yes	moderate
fitted	M06-2X, M11, MN15	no	n/a	none

⚠ Common mistakes about dispersion corrections

- "-D4 changes the orbitals." It does not. It adds an energy computed from the geometry (it does change optimized geometries, through the forces).
- "Adding -D4 always helps." It fixes missing dispersion. It cannot fix other errors of the functional, and if the functional already has spurious binding (or the basis has BSSE), adding -D4 can overbind.
- "The V in ωB97M-V means it is a meta-GGA." V stands for VV10, a correlation term. The rung is set by the exchange and by the ingredients.

8.2.5 Intramolecular dispersion: branched alkanes

Dispersion does not act only *between* molecules. Any two atoms that are a few ångström apart in the same molecule also attract each other. In a long, extended chain most atom pairs are far apart. In a compact, **branched** molecule, more atoms are at medium distance (1,3- and 1,4-contacts between C and H atoms), so the total intramolecular dispersion is larger.

This is part of the reason why branched alkanes are more stable than their linear isomers: for example, neohexane (2,2-dimethylbutane) lies below *n*-hexane. Schleyer and co-workers called the stabilizing 1,3-interactions in alkanes **protobranching**. The effect is small, a few kcal/mol, but it is a classic difficulty for DFT. In 2006 Grimme [110] and Wodrich, Corminboeuf and Schleyer [111] showed that popular functionals such as B3LYP make systematic errors for the relative stability of alkane isomers, and that the errors grow with the size of the molecule. Missing medium-range correlation (dispersion) is one of the causes. Karton, Gruzman and Martin computed benchmark values for all the alkane isomers up to octane [112]; reaction 4b comes from that work, via the AlkIsomer11 subset of MGCDB84.

Notice how different this is from Experiment 3. There, the molecule changed shape but kept the same set of close contacts, so dispersion nearly cancelled. Here, the *number* of medium-range contacts changes.

8.2.6 A reminder about BSSE

When you compute the binding energy of a dimer, E(dimer) − E(monomer A) − E(monomer B), BSSE makes the dimer artificially too stable (Lecture 7). With def2-TZVP the BSSE is much smaller than with double-ζ sets, but it is not zero, and for a dimer bound by 1–2 kcal/mol even a few tenths matter. We do not correct for it in Experiment 4. Lecture 10 defines the counterpoise correction that removes most of it.

8.3 Experiment 4 briefing

8.3.1 The goal

Experiment 4 moves from “easy” to “medium” reactions. Reactants and products are now different molecules, or a complex and its separated parts, so error cancellation can no longer be taken for granted.

Reaction	Description	ΔE_{ref} (kcal/mol)
4a	ethene + butadiene ⟶ cyclohexene (Diels–Alder)	−45.40
4b	*n*-hexane ⟶ neohexane	−2.49
4c	2 ethane ⟶ $(ethane)_2$	−1.34
4d	MeOH(A) + MeOH(B) ⟶ $(MeOH)_2$	−5.81

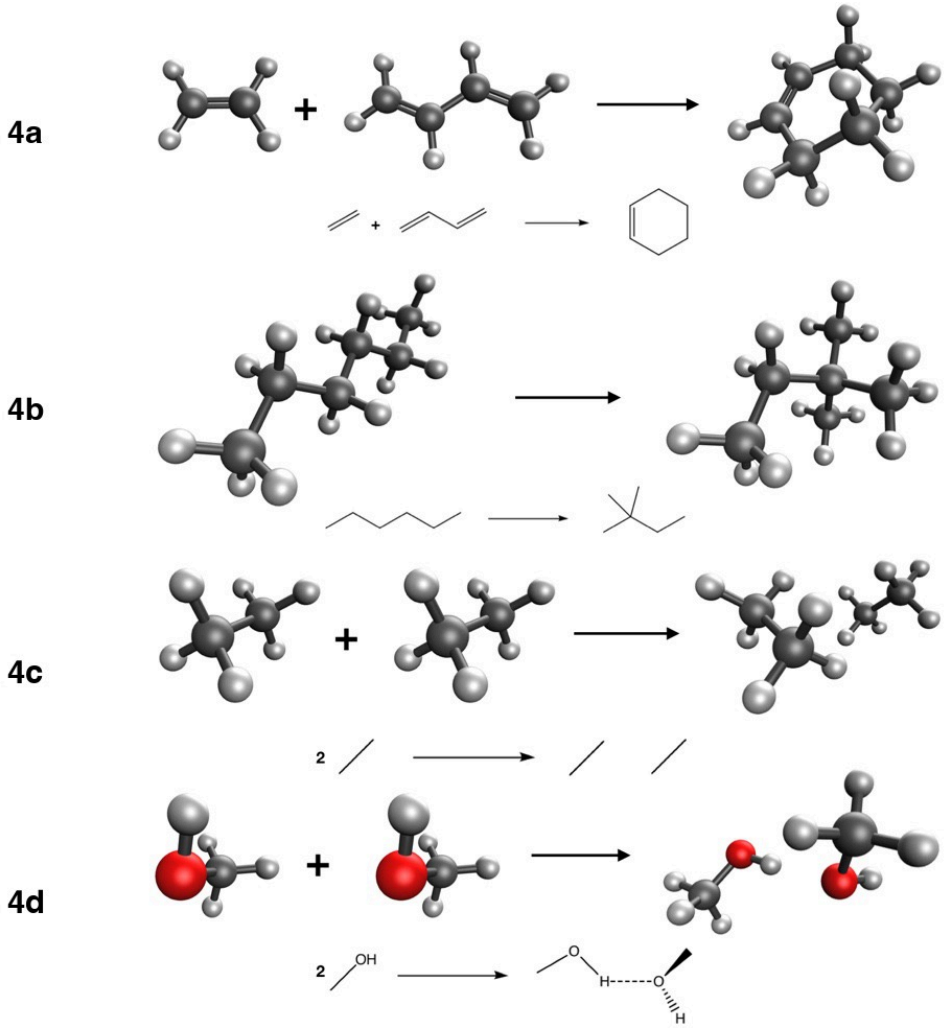


Figure 17: Reactions of Experiment 4: Diels–Alder reaction of ethene and butadiene (4a), isomerization of *n*-hexane to neohexane (4b), and formation of the ethane (4c) and methanol (4d) dimers.

What each reaction tests:

- **4a, Diels–Alder.** Two π bonds become two new C–C σ bonds and one π bond moves. The bonding changes substantially between reactants and product. This is a test of how well a functional describes *bond formation*, not dispersion.
- **4b, alkane isomerization.** Same atoms, same bond types (all C–C and C–H single bonds), but a different number of medium-range contacts. A test of intramolecular dispersion and of subtle electronic effects in alkanes.
- **4c, ethane dimer.** Almost pure dispersion. A direct test of how each method treats dispersion.
- **4d, methanol dimer.** A hydrogen bond. Mostly electrostatics and induction, with a smaller dispersion part.

8.3.2 Methods and jobs

The same 16 methods as Experiment 3, all with def2-TZVP. Ten molecules × 16 methods = **160 jobs**.

8.3.3 What to watch out for

- **Directions and signs.** 4c and 4d are written as *monomers* ⟶ *dimer*, so the reference values are negative: a negative ΔE means the dimer is bound. A positive signed error for a dimer means it is bound **too weakly**.
- **Coefficients.** In 4c the ethane monomer counts twice: $\nu = -2$. In 4a and 4d three species are involved.
- **Two different methanol monomers.** In 4d the monomers are cut from the dimer and have slightly different geometries, so compute both (A and B). Do not use one monomer twice.
- **Pairs of methods.** PBE/PBE-D4 and B3LYP/B3LYP-D4 differ only by the dispersion correction. The *Does -D4 help?* chart compares each pair directly.
- **The dimer chart.** The *binding of the two dimers* chart groups the methods by how they treat dispersion (none, D4, VV10, fitted). Before you look, write down what you expect for each group on 4c and on 4d, and why the two dimers might behave differently.
- Lock in your predictions on the experiment page before you run anything.

8.3.4 Key points

- Conformational and cis–trans reactions are easy because errors cancel between two forms of the same molecule; noncovalent changes (like the H-bond in 3c) cancel less.
- LSDA overbinds weak interactions, HF underbinds them; neither is reliable for noncovalent chemistry.
- Hybrids cost more per SCF cycle than (meta-)GGAs because of the exchange matrix K; range separation and VV10 add more. B3LYP is not special in cost.
- London dispersion is an attractive correlation between fluctuating dipoles. It decays as $-C_6/R^6$, with C_6 set by polarizabilities.
- Semilocal functionals and Hartree–Fock cannot describe dispersion: semilocal ones see one point at a time, HF has no correlation.
- -D4 adds damped pairwise $-C_6/R^6$ and $-C_8/R^8$ terms with coordination- and charge-dependent C_6, plus a three-body term. VV10/rVV10 add a nonlocal correlation functional of the density. Minnesota functionals capture medium-range dispersion by fitting.
- Damping prevents the divergence of $-C_6/R^6$ at short range and avoids double counting with the functional.
- Intramolecular dispersion helps make branched alkanes more stable than linear ones, and is a known difficulty for popular functionals.

8.3.5 Study questions

1. Why does adding -D4 change the reaction energies of Experiment 3 so little?
2. Explain, with the signs, why LSDA and HF tend to err in opposite directions on the hydrogen-bond-breaking reaction 3c.

3. Starting from Equation 72, carry out the expansion and verify Equation 73. Why do the first-order terms cancel?
4. Use Equation 75 to explain why C_6 for two xenon atoms is much larger than for two neon atoms.
5. Two argon atoms are 5.0 Å apart. Using $C_6 = 886$ kcal mol^{-1} Å^6, compute the undamped dispersion energy. Is damping important at this distance? (Look at Figure 16.)
6. Explain in your own words why a GGA gives zero interaction energy for two atoms whose densities do not overlap.
7. What is the role of the damping function in -D4? What would happen in a geometry optimization without it?
8. What does -D4 add over D3? Give one chemical situation where the difference should matter.
9. Compare -D4 and VV10: which one depends on the electron density, which is cheaper, and which one has an R^{-6} tail with coefficients computed from the density?
10. Why is a branched alkane stabilized more by dispersion than its linear isomer?

8.3.6 Further reading

- S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, "Dispersion-corrected mean-field electronic structure methods" [104]. A complete review of dispersion corrections.
- E. Caldeweyher et al., the two -D4 papers [58], [59].
- O. A. Vydrov and T. Van Voorhis, the VV10 paper [60].
- J. Řezáč and P. Hobza, "Benchmark calculations of interaction energies in noncovalent complexes and their applications" [103].
- S. Grimme, "Seemingly simple stereoelectronic effects in alkane isomers and the implications for Kohn–Sham density functional theory" [110].

9 Why some reactions are hard for DFT

> **i** Learning goals
>
> After this lecture you should be able to:
>
> - read the -D4 pair chart and the dimer chart of Experiment 4, and explain which reactions dispersion corrections help and why;
> - explain why error cancellation breaks down when bonds are formed or broken;
> - define the self-interaction error and the delocalization error, and list their chemical consequences;
> - explain what static (nondynamic) correlation is and why a single-determinant method struggles with ozone, stretched bonds and CaO;
> - explain why atomization energies are hard, especially for Hartree–Fock;
> - assign the correct spin multiplicity to O, S, H, Ca and O_2 and set up unrestricted calculations;
> - distinguish absolute and relative errors, and explain why a functional should not be judged on its own training set.

This lecture may be taught over two meetings. The split point is marked in the text.

9.1 Part 1 · Looking back at Experiment 4

> **!** Read this after you submit Experiment 4
>
> This part discusses the trends you probably saw in Experiment 4. If you have not submitted it yet, stop here and come back later.

9.1.1 The big picture: errors got larger

Compare the *signed error* chart of Experiment 4 with the one from Experiment 3. The bars are longer, and they differ much more from method to method. The typical error has grown, sometimes by an order of magnitude. This is what happens when reactants and products are no longer two forms of the same molecule. Each reaction in Experiment 4 failed to cancel errors for a different physical reason, which is why it is worth looking at them one at a time.

9.1.2 4a: the Diels–Alder reaction

The Diels–Alder reaction (ethene + butadiene ⟶ cyclohexene, −45.40 kcal/mol) usually produces the largest absolute errors of the experiment. What you probably saw:

- Several older functionals, BLYP and B3LYP in particular, and also Hartree–Fock, make the reaction **not exothermic enough** (positive signed error), often by many kcal/mol.
- LSDA tends to err in the opposite direction (too exothermic), consistent with its general overbinding.
- The modern meta-GGAs and hybrid meta-GGAs usually do much better.
- -D4 helps B3LYP and PBE, but it does not remove the whole error.

Why is this reaction hard? Two π bonds turn into two new C–C σ bonds. The electron density changes substantially: a delocalized, conjugated π system in butadiene becomes localized single bonds in cyclohexene. Errors in describing a π bond and a σ bond are different, so they no longer cancel. In addition:

- **Delocalization error** (Part 2) makes approximate functionals overstabilize delocalized electrons, like the conjugated π system of butadiene. A reactant that is too stable makes the reaction too little exothermic.

- **Medium-range correlation** is larger in the compact, ring-shaped product than in the two separated reactants. A functional that misses it underestimates the stability of the product. This is the part -D4 can fix, and it explains why -D4 helps but is not enough.

9.1.3 4b: n-hexane → neohexane

This reaction keeps the same atoms and the same types of bonds, but neohexane is more compact. What you probably saw: functionals without any dispersion treatment (BLYP, BP86, B3LYP, B3PW91, PBE, PBE0) do not make neohexane stable enough, so their signed errors are positive and similar in size to the reaction energy itself. Adding -D4 shrinks the error substantially, and the VV10 functionals and the fitted Minnesota functionals are usually good. Hartree–Fock, with no correlation, is among the worst.

You may also have seen LSDA (SPW92) do surprisingly well here. That is the spurious overbinding of LSDA in regions of small overlap, which mimics the missing medium-range attraction. It is "right for the wrong reason", and it will not survive the next experiment.

9.1.4 4c and 4d: the two dimers

The *binding of the two dimers* chart sorts the methods by how they treat dispersion. Remember that positive ε means the dimer is bound too weakly.

For the **ethane dimer (4c)**, which is held together almost only by dispersion:

- methods with **no dispersion** treatment bind too weakly. BLYP, BP86 and B3LYP give almost no binding, or even repulsion, because B88 exchange is repulsive in the overlap region. PBE gives a little binding from its exchange behavior, but not enough. HF does not bind at all;
- **-D4** and **VV10** usually bring the binding very close to the reference;
- the **fitted** Minnesota functionals usually get most of it, because similar complexes were in their training sets;
- **LSDA** tends to overbind.

For the **methanol dimer (4d)**, held together by a hydrogen bond, the picture is different. Most methods, even those without any dispersion term, are within about 1 kcal/mol, because the hydrogen bond is dominated by electrostatics and induction, which all methods describe reasonably. Dispersion is a smaller part of the binding, so -D4 changes it less, and in some cases slightly overbinds. The lesson is that "noncovalent" is not one thing: an interaction dominated by dispersion and one dominated by electrostatics respond differently to dispersion corrections.

9.1.5 The -D4 pair chart

The *Does -D4 help?* chart pairs PBE with PBE-D4 and B3LYP with B3LYP-D4. What you probably saw: -D4 helps most where the *number of close contacts* changes, that is, 4c (two molecules come together) and 4b (a more compact isomer). It helps partially for 4a. It matters little for 4d. Overall, the MAE of each functional drops when -D4 is added. This matches the theory of Lecture 8: -D4 adds only dispersion, so it improves only what was wrong because of missing dispersion.

9.1.6 Lessons of Experiment 4

1. Error cancellation weakens as the chemical change between reactants and products grows.
2. Dispersion corrections fix dispersion. They are cheap and should be used routinely with functionals that lack dispersion.
3. A method can be right for the wrong reason (LSDA on 4b). Check a method on several reactions before trusting it.

4. Bond formation (4a) is a different problem from dispersion, and needs a better functional, not a better dispersion correction.

9.2 Part 2 · Why DFT fails

Experiment 5 contains reactions that are hard for nearly every functional. To understand why, we need three ideas: self-interaction (and delocalization) error, static correlation, and the special difficulty of atomization energies. All are discussed in the excellent review by Cohen, Mori-Sánchez and Yang [57].

9.2.1 Self-interaction error

9.2.1.1 One electron should not repel itself

Consider a system with just one electron, like the hydrogen atom. There is no electron–electron interaction, so the exact energy contains only the kinetic energy and the attraction to the nucleus. In Kohn–Sham DFT the energy is

$$E[\rho] = T_s[\rho] + E_{\mathrm{ne}}[\rho] + J[\rho] + E_{\mathrm{xc}}[\rho], \tag{82}$$

where J[ρ] is the classical Coulomb energy of the density with itself:

$$J[\rho] = \frac{1}{2}\iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\, d\mathbf{r}\, d\mathbf{r}'. \tag{83}$$

For one electron, J is not zero: it is the repulsion of the electron's density with itself, a spurious **self-interaction**. The exact functional cancels it exactly:

$$J[\rho_1] + E_{\mathrm{xc}}[\rho_1] = 0 \quad \text{for any one-electron density} \quad \rho_1. \tag{84}$$

In Hartree–Fock, the condition holds automatically, because the Coulomb and exchange integrals of an orbital with itself are equal ($J_{ii} = K_{ii}$, Lecture 4). Approximate functionals (LSDA, GGAs, meta-GGAs) do not satisfy Equation 84. Perdew and Zunger called the leftover the **self-interaction error** (SIE) and proposed an orbital-by-orbital correction for it [56]. Hybrids reduce the SIE in proportion to their fraction of HF exchange, and range-separated hybrids, with 100% HF exchange at long range, remove it at long range.

9.2.1.2 Many electrons: delocalization error

For many-electron systems the more useful concept is the **delocalization error** [113]. Imagine allowing the number of electrons N to vary continuously, so that a system can hold a fractional number of electrons (think of a statistical mixture, or of one electron shared between two very distant fragments). The exact energy E(N) is a series of **straight lines** between integer values of N [114]. Approximate functionals give curves instead (Figure 18):

- typical density functionals give curves that bow **below** the straight lines (they are convex). Fractional charges, and therefore electrons spread over too large a region, come out **too low in energy**. This is the **delocalization error**.
- Hartree–Fock gives curves that bow **above** the lines (concave). It favors electrons that are too localized: the **localization error**.

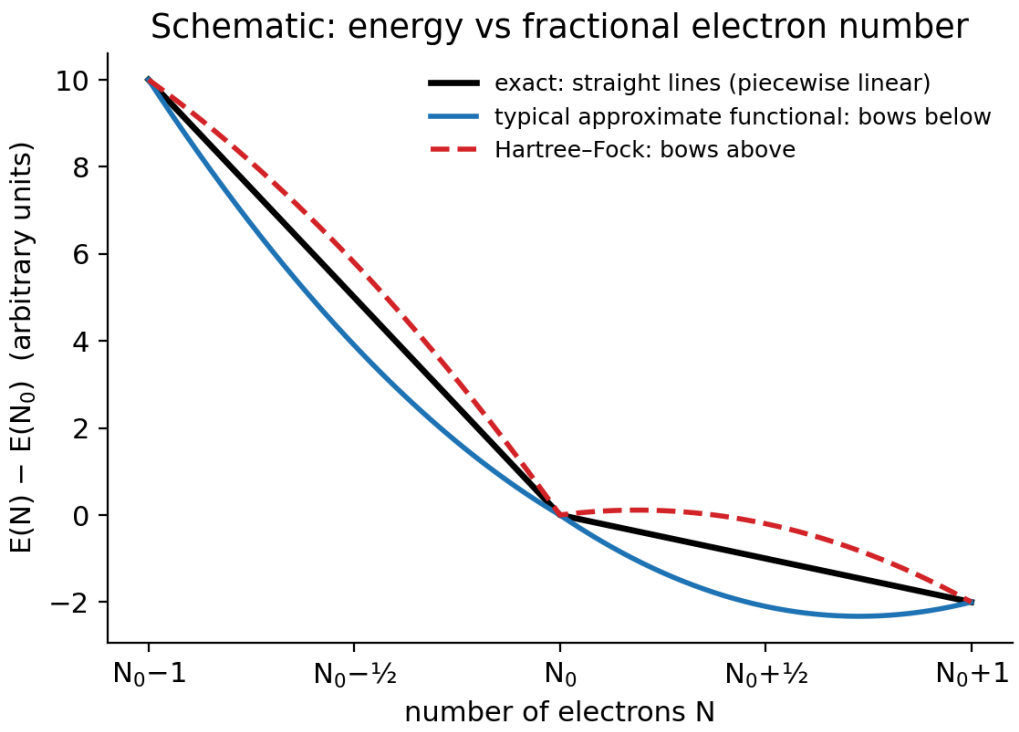


Figure 18: Schematic of the energy of a system as a function of a continuous number of electrons N between $N_0 - 1$ and $N_0 + 1$. The exact energy is piecewise linear. Typical approximate density functionals bow below the lines (delocalization error); Hartree–Fock bows above them (localization error). Arbitrary units; the curves are a schematic, not computed results.

Hybrids mix the two behaviors, and their errors partly cancel. This is one reason why hybrids with 20–50% HF exchange perform well for thermochemistry.

9.2.1.3 Chemical consequences

The delocalization error shows up in many places:

- **Stretched one-electron bonds.** For H_2^+ at long distance, the exact answer is a proton plus a hydrogen atom, or equivalently half an electron on each side with the same energy. Approximate functionals favor the delocalized "half and half" state too much, and give a dissociation curve that is far too low at large distance. The same happens for other radical cations [115].
- **Barrier heights too low.** In a transition state, bonds are partly broken and electrons are more delocalized than in reactants. GGAs overstabilize the transition state and underestimate barriers, often by several kcal/mol. More HF exchange raises barriers, which is why functionals built for kinetics (M06-2X, for example) have a lot of it.
- **Charge-transfer complexes** and donor–acceptor systems are overbound, because partial charge transfer is overstabilized.
- **Conjugated and delocalized systems** are overstabilized relative to localized ones (we saw a hint of this in 4a).
- **Band gaps and orbital energies.** The HOMO energy of a GGA is much too high, and fundamental gaps are too small.
- **Odd-electron systems and radicals** with delocalized spin are overstabilized.

The delocalization error is a property of the *functional*. A larger basis does not help, and neither does -D4.

> **i** Split point
>
> If this lecture is taught over two meetings, the first meeting ends here, after the Experiment 4 debrief and self-interaction error. The second meeting starts with a short recap, then covers static correlation, atomization energies and the Experiment 5 briefing.

9.2.2 Static correlation and multireference character

9.2.2.1 Dynamic and static correlation

In Lecture 4 we split the correlation energy into two kinds:

- **dynamic correlation**: electrons avoid each other at short range. Every system has it. A single determinant is a good starting point, and correlation adds small corrections from many excited determinants.
- **static (nondynamic) correlation**: the system cannot be described by a single determinant at all, because two or more electron configurations have similar energies and mix strongly. Systems like this are said to have **multireference character**.

9.2.2.2 The textbook example: stretching H_2

At its equilibrium distance, H_2 is described well by one determinant with both electrons in the bonding orbital σ_g. As the bond is stretched, the antibonding orbital σ_u comes down in energy, and the configurations σ_g^2 and σ_u^2 become degenerate. The exact wave function at long distance is an equal mixture of the two, which describes one electron on each atom. Restricted Hartree–Fock (RHF), stuck with σ_g^2, puts both electrons on the same atom half of the time ($H^+ + H^-$) and gives a dissociation energy that is far too large. An unrestricted calculation (UHF or UKS) can break the symmetry and dissociate correctly, at the price of a wave function that is no longer a pure spin state (spin contamination, Lecture 12).

Approximate functionals are built for single-determinant Kohn–Sham systems. They capture dynamic correlation reasonably, and they mimic some static correlation through their exchange part, but systems with strong static correlation remain hard. They also make the reference values harder to compute: CCSD(T) itself becomes less reliable when the single-determinant starting point is poor.

9.2.2.3 Where static correlation appears in Experiment 5

- **Ozone.** O_3 has partial **diradical character**: besides the main closed-shell configuration, a second configuration with two electrons promoted into a low-lying π^* orbital has a large weight. This makes O_3 a classic multireference molecule. Reactions 5a (ozonolysis) and 5d ($2\ O_3 \longrightarrow 3\ O_2$) both involve it.
- **CaO.** Calcium oxide is a molecule with low-lying excited states and strongly mixed bonding (ionic $Ca^{2+}O^{2-}$ and less ionic configurations close in energy). Reaction 5b comes from MR-MGM-BE4, a subset of *multireference* main-group bond energies [116].
- **Stretched bonds and bond breaking** in general, and **transition-metal compounds** (Lecture 12).

How do you know whether a molecule has multireference character? Diagnostics exist, for example the T_1 diagnostic from coupled cluster [117], or the fraction of the atomization energy that comes from the (T) triples. Large values are a warning: the reference values themselves may carry larger uncertainties, and DFT results should be treated with caution.

9.2.3 Atomization energies and open-shell atoms

9.2.3.1 Why atomization energies are hard

An **atomization energy** is the energy needed to separate a molecule into its atoms, for example

$$\mathrm{H_2SO_4} \rightarrow \mathrm{S} + 4\,\mathrm{O} + 2\,\mathrm{H}, \qquad \Delta E = E(\mathrm{S}) + 4E(\mathrm{O}) + 2E(\mathrm{H}) - E(\mathrm{H_2SO_4}). \tag{85}$$

Here all the bonds are broken at once, and the products are open-shell atoms while the reactant is a closed-shell molecule. This is the opposite of an isodesmic reaction: there is essentially **no error cancellation**. Every error in the description of each bond adds up. That is why atomization energies are the classic hard test for functionals, and the test on which Becke's first hybrid functionals were judged [79].

9.2.3.2 Why Hartree–Fock fails dramatically

When atoms come together to form bonds, electrons pair up in the same region of space. A molecule therefore has *more* correlation energy than its separated atoms. Hartree–Fock has no correlation at all, so it misses this extra stabilization of the molecule and **severely underestimates atomization energies**. The error grows with the number of bonds, and for a molecule like H_2SO_4 it is enormous. For the same reason, HF is not a useful method for any reaction where the number of electron pairs changes.

9.2.3.3 Open-shell atoms and their multiplicities

The atoms you will compute in Experiment 5 are open-shell, and their ground-state multiplicities follow Hund's rules:

Table 19: Ground-state multiplicities of the open-shell species (and Ca) in Experiment 5.

Species	Configuration	Ground state	Multiplicity 2S + 1
H	$1s^1$	^2S	2 (doublet)
O	[He] $2s^2\,2p^4$	^3P	3 (triplet)
S	[Ne] $3s^2\,3p^4$	^3P	3 (triplet)
Ca	[Ar] $4s^2$	^1S	1 (singlet)
O_2	$(\pi^*)^2$	$^3\Sigma_g^-$	3 (triplet)

Two practical rules follow.

1. **Set the multiplicity correctly.** Line 2 of every .xyz file in this lab is "charge multiplicity". A triplet O atom run as a singlet gives a perfectly converged, perfectly wrong energy, too high by tens of kcal/mol, and every reaction that uses it is wrong.
2. **Use an unrestricted calculation** (`UNRESTRICTED = TRUE` in Q-Chem) for open shells, so that α and β electrons can have different orbitals. The generated inputs do this for you.

Open-shell atoms are hard for approximate functionals for additional reasons. The functional must describe **spin polarization** (different α and β densities) correctly. The p-shell of O and S is partly filled, and an approximate functional may give different energies for different ways of occupying the p orbitals, or break the spherical symmetry of the atom. The molecule, by contrast, is closed-shell. The errors in the two kinds of systems are different and do not cancel.

Worked example: an atomization energy from total energies

Suppose a calculation gives these total energies (illustrative numbers, not from any experiment in this course): E(H_2O) = −76.40000 E_h, E(O, triplet) = −75.05000 E_h, E(H, doublet) = −0.50000 E_h. For $H_2O \longrightarrow O + 2\,H$, with ν($H_2O$) = −1, ν(O) = +1 and ν(H) = +2:

$$\Delta E = (+1)(-75.05000) + (+2)(-0.50000) + (-1)(-76.40000) = -76.05000 + 76.40000 = 0.35000\ E_h.$$

In kcal/mol: 0.35000 × 627.5095 = 219.63 kcal/mol. Note the sign: atomization is endothermic, so ΔE > 0. If the O atom had been run as a singlet, its energy would be too high, and the atomization energy would come out too large.

9.2.4 Absolute and relative errors

Reaction 5c, the atomization of sulfuric acid, has a reference value of 602.18 kcal/mol, by far the largest number in the course. A method that is "only" 1% off is wrong by about 6 kcal/mol. Is that good or bad?

- In **relative** terms (the relative error defined in Lecture 7), 1% looks excellent, much better than a 0.3 kcal/mol error on a 1 kcal/mol rotational barrier (30%).
- In **absolute** terms, 6 kcal/mol is six times chemical accuracy. It would change an equilibrium constant by a factor of about $e^{6/0.5925} \approx 2.5 \times 10^4$ at room temperature.

Chemistry is decided by absolute energies. Relative errors are useful to compare the difficulty of very different quantities, but they must never be used to excuse a large absolute error. Experiment 5 has a chart that shows the same errors in both ways.

9.2.5 Training sets and fitted functionals

Many functionals contain parameters fitted to data (Lecture 6). B3LYP has 3; the Minnesota functionals have several dozen. The data used for the fit are the **training set**. A fitted functional reproduces its training set well *by construction*, so a good result on a training-set molecule is expected, not impressive. What matters for a user is how the functional performs on systems it was **not** fitted to.

Reaction 5b (CaO ⟶ Ca + O) belongs to the MR-MGM-BE4 subset of Minnesota Database 2015B [116], and MN15 was parametrized on that database [31]. Experiment 5 has a *spotlight* chart for this reaction. When you look at it, ask: is this result a test of MN15, or a check that the fit worked? The same caution applies to any benchmark: before you use a dataset to judge a functional, check whether it was part of the functional's training set.

Common mistakes in hard cases

- Running open-shell atoms as singlets, or with restricted orbitals.
- Using the wrong coefficient for the atoms: in 5c there are 4 O atoms and 2 H atoms (ν = +4 and +2).
- Expecting -D4 to fix everything. Dispersion corrections fix dispersion, not self-interaction or static correlation.
- Judging a functional by a single good result, especially on its own training data.

9.3 Experiment 5 briefing

9.3.1 The goal

Experiment 5 is the last experiment on functionals. It contains four reactions that look simple chemically (the largest molecules have only five atoms heavier than hydrogen) but are known to be hard for DFT.

Reaction	Description	ΔE_{ref} (kcal/mol)
5a	$O_3 + C_2H_4 \longrightarrow$ primary ozonide	−58.7
5b	$CaO \longrightarrow Ca + O$	96.15
5c	$H_2SO_4 \longrightarrow S + 4\,O + 2\,H$	602.18
5d	$2\,O_3 \longrightarrow 3\,O_2$	−67.60

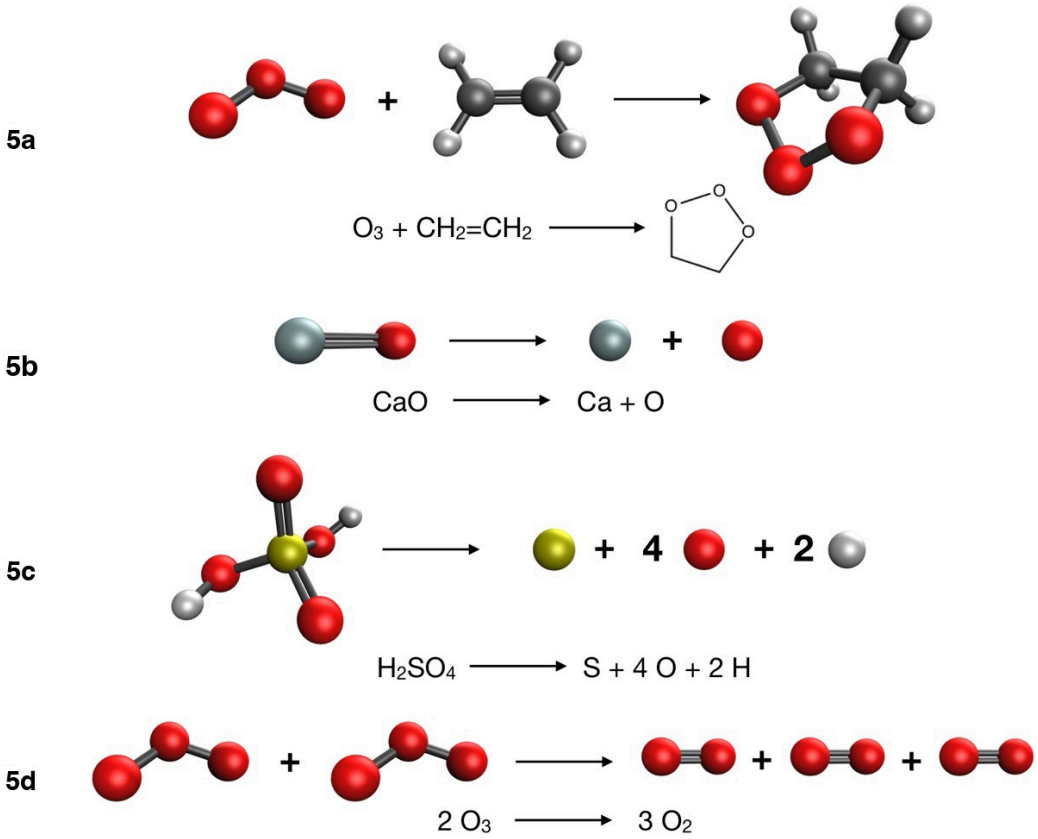


Figure 19: Reactions of Experiment 5: formation of the primary ozonide from ozone and ethene (5a), dissociation of CaO (5b), atomization of sulfuric acid (5c), and conversion of ozone into oxygen (5d).

For each reaction, think about which of the difficulties of Part 2 it contains: delocalization error, static correlation, open-shell atoms, no error cancellation. Some reactions contain more than one.

9.3.2 Methods and jobs

The same 16 methods as Experiments 3 and 4, all with def2-TZVP. Twelve species × 16 methods = **192 jobs**. Note that 5a and 5d use two different ozone geometries (from two different databases), and that the O atom of 5b and the O atom of 5c also come from different databases: use the files listed for each reaction.

9.3.3 What to watch out for

- **Multiplicities.** O, S and O_2 are triplets, H is a doublet, Ca and all the closed-shell molecules are singlets (Table 19). The generated inputs already set the multiplicity and `UNRESTRICTED = TRUE`. If you build an input by hand, check line 2 of the .xyz file.
- **Coefficients.** In 5c, $\nu = -1$ for H_2SO_4, +1 for S, +4 for O and +2 for H. In 5d, $\nu = -2$ for O_3 and +3 for O_2.
- **SCF convergence.** Open-shell atoms and ozone can take more SCF cycles. Check that every job converged before using its energy.
- **The charts.** Besides the signed errors, you will see the *relative-error* chart, the *CaO spotlight* and the *easy → hard* chart, which combines your Experiments 3, 4 and 5 (from this browser) and shows how your functional-draft pick did.
- Lock in your predictions on the experiment page first: how much larger will the errors be, which method will win, and will -D4 help?

9.3.4 Key points

- In Experiment 4 errors grew because reactants and products differ: bonds change in 4a, the number of close contacts in 4b and 4c.
- -D4 and VV10 fix missing dispersion (4b, 4c) but not errors in bond formation (4a); hydrogen bonds (4d) depend mostly on electrostatics.
- The self-interaction error is the spurious repulsion of an electron with itself; HF has none, approximate functionals do.
- The delocalization error makes approximate functionals overstabilize spread-out electrons: stretched radical cations, transition states, charge transfer, conjugated systems. HF has the opposite (localization) error.
- Static correlation arises when several configurations are nearly degenerate (stretched bonds, ozone, CaO); single-determinant methods struggle.
- Atomization energies have no error cancellation. HF underestimates them badly because molecules have more correlation energy than their atoms.
- Open-shell atoms need the right multiplicity (O, S triplet; H doublet; Ca singlet) and unrestricted orbitals.
- Absolute errors decide chemistry; relative errors can hide large absolute errors.
- A functional's result on its own training data is not an independent test.

9.3.5 Study questions

1. Why did -D4 help much more for 4c than for 4d? Answer in terms of the physical components of each interaction.
2. Give two reasons why the Diels–Alder reaction 4a is harder for DFT than the ethane rotation 3a.
3. Write Equation 84 in words. Why does Hartree–Fock satisfy it automatically?
4. Sketch E(N) between $N_0 - 1$ and $N_0 + 1$ for the exact functional, a GGA and HF. Use your sketch to explain why GGAs underestimate reaction barriers.
5. Why is a GGA's dissociation curve for H_2^+ too low at large distance?
6. Explain the difference between dynamic and static correlation, using H_2 at equilibrium and at a long bond length.
7. Why does Hartree–Fock underestimate atomization energies? Would you expect the error to grow or shrink for larger molecules?
8. Determine the ground-state multiplicity of N, C, F and Mg atoms. Which need an unrestricted calculation?
9. An atomization energy of 400 kcal/mol is computed with a 1.5% error. A conformational energy of 0.8 kcal/mol is computed with a 50% error. Which is the larger absolute error, and by how much?

10. A new functional reproduces a molecule's bond energy within 0.1 kcal/mol, and that molecule was in its training set. What can you conclude? What would you check next?

9.3.6 Further reading

- A. J. Cohen, P. Mori-Sánchez and W. Yang, "Challenges for density functional theory" [57]. The best review of delocalization error and static correlation.
- P. Mori-Sánchez, A. J. Cohen and W. Yang, "Localization and delocalization errors in density functional theory" [113].
- J. P. Perdew and A. Zunger, the original self-interaction-correction paper [56].
- E. R. Johnson, P. Mori-Sánchez, A. J. Cohen and W. Yang, "Delocalization errors in density functionals and implications for main-group thermochemistry" [118].
- K. H. Hopmann's editorial in *Organometallics* on what makes a computational study useful [119]: "Do not use DFT for mechanisms it cannot handle."

10 Basis-set error vs functional error

ℹ Learning goals

After this lecture you should be able to:

- summarize what Experiment 5 showed about hard reactions, -D4, relative errors and training sets;
- read the easy ⟶ hard chart across Experiments 3–5 and say what makes a method good overall;
- write the total error of a calculation as a method error plus a basis-set error, and explain how the two can cancel;
- define the basis-set incompleteness error (BSIE) and the basis-set superposition error (BSSE), and write the counterpoise correction;
- explain why B3LYP/6-31G* became the most used level of theory, and why it can be right for the wrong reasons;
- count the basis functions of a molecule in 6-31G*, def2-SVP, cc-pVDZ and pc-1;
- start Experiment 6 knowing what to run and what to look for.

10.1 Part 1 · Looking back at Experiment 5

! Read this after you submit Experiment 5

This part discusses the trends you probably saw in Experiment 5. If you have not submitted it yet, stop here and come back later.

10.1.1 The errors exploded

The first thing you probably noticed in Experiment 5 is the scale of the *signed error* chart. In Experiment 3 the bars were measured in tenths of a kcal/mol; in Experiment 4 in kcal/mol; here many are measured in tens of kcal/mol, and a few in hundreds. No method was accurate for all four reactions. Most methods had at least one error well above chemical accuracy.

What you probably saw, reaction by reaction, and why:

- **Hartree–Fock** is catastrophic for the two dissociations, 5b (CaO) and especially 5c (H_2SO_4 atomization). It badly *underestimates* both, because the molecule has much more correlation energy than the separated atoms, and HF has none (Lecture 9). Its error on 5c is so large that it dominates the chart.
- **LSDA (SPW92)** errs the other way: it **overbinds** molecules, so its atomization energies are far too large. This is the classic failure of LSDA that motivated GGAs in the first place.
- **GGAs** (BP86, BLYP, PBE) still make large errors, typically overbinding CaO by many kcal/mol. Their delocalization error and the difficulty of open-shell atoms add up, with no cancellation.
- **Hybrids and hybrid meta-GGAs** usually reduce the errors, but none of them is uniformly good. The sign of the error changes from reaction to reaction and from functional to functional.
- The two **ozone** reactions (5a and 5d) are hard for almost everyone. Ozone's multireference character is not described well by any single-determinant functional, and different functionals miss it in different directions.

10.1.2 -D4 is not a universal fix

Compare B3LYP with B3LYP-D4 and PBE with PBE-D4. What you probably saw is that -D4 changes these reaction energies only a little, and not always in the right direction. That is exactly what the theory predicts. -D4 adds the dispersion energy between atom pairs. In an atomization or a dissociation, the dispersion energy

of the molecule is a tiny fraction of the bond energies, and the errors in Experiment 5 come from somewhere else: self-interaction, static correlation, the description of open-shell atoms. A correction designed for one physical effect cannot repair another.

10.1.3 Absolute and relative errors

The *relative-error* chart shows the same errors as a percentage of each reference value. Reaction 5c has the largest reference (602.18 kcal/mol), so its percentage errors look modest even when the absolute errors are the largest of the experiment. For example, a 12 kcal/mol error on 5c is only 2%, but it is twelve times chemical accuracy. A rotational barrier of 1 kcal/mol computed with a 0.3 kcal/mol error looks bad at 30%, but it is chemically harmless. The lesson of Lecture 9 stands: chemistry is decided by absolute energies; relative errors only help compare quantities of very different size.

10.1.4 The CaO spotlight

The *spotlight* chart for 5b highlights MN15. You probably saw that MN15 has one of the smallest errors on CaO, while most other functionals miss it by many kcal/mol. This is not proof that MN15 understands CaO better. CaO is part of the database MN15 was fitted to [31], [116], so reproducing it well is what the fit was designed to do. On the other reactions of Experiment 5, MN15 is not exceptional. A functional must be judged on data it was **not** trained on. This is true of every fitted functional, Minnesota or not.

10.1.5 From easy to hard

The *easy* → *hard* chart puts your Experiments 3, 4 and 5 side by side, with the MAE of each method on a log scale. Three trends usually appear:

1. **Everything gets worse from left to right.** The MAE of a typical method grows by one to two orders of magnitude from Experiment 3 to Experiment 5. Error cancellation does the work in the easy reactions; in the hard ones, the functional is on its own.
2. **The ranking changes.** Methods that look equivalent in Experiment 3 separate in Experiment 4 (dispersion) and again in Experiment 5 (self-interaction and static correlation). LSDA and HF, which hid among the others in Experiment 3, become the worst.
3. **Modern functionals win on average.** The best overall performers are usually among the modern meta-GGAs and hybrid meta-GGAs with good dispersion treatment, such as ωB97M-V, B97M-V or the Minnesota hybrids. But no method wins every reaction, and the margin between the best few is smaller than the spread of errors within each experiment.

Look at the thick line on the chart: your Experiment 3 "draft pick". Was it still a good choice by Experiment 5? If not, what would you have needed to know in advance? This is the question every computational chemist faces when choosing a method, and it is why we read benchmarks on the *subsets* closest to our problem (Lecture 14).

> ⚠ Common mistake: "Functional X is the best"
>
> There is no best functional, only functionals that are better on average or better for a class of problems. Experiment 5 shows that a functional can be excellent on one reaction and poor on the next, even among reactions with five heavy atoms. As Hopmann put it, "Do not use DFT for mechanisms it cannot handle" [119].

10.2 Part 2 · Basis-set error vs functional error

10.2.1 Two approximations, two errors

Every calculation makes the two approximations of Lecture 1: a finite basis set (approximating Ψ, or the Kohn–Sham orbitals) and an approximate method (approximating $\hat{H}$, here the exchange–correlation functional). Each introduces an error. To separate them, imagine computing the same reaction with the same functional in a **complete basis set** (CBS). Then

$$\underbrace{\Delta E^{\text{method/basis}} - \Delta E^{\text{ref}}}_{\varepsilon_{\text{total}}} = \underbrace{\Delta E^{\text{method/CBS}} - \Delta E^{\text{ref}}}_{\varepsilon_{\text{method}}} + \underbrace{\Delta E^{\text{method/basis}} - \Delta E^{\text{method/CBS}}}_{\varepsilon_{\text{basis}}}. \tag{86}$$

This is an identity: we have only added and subtracted $\Delta E^{\text{method/CBS}}$. But it is a useful one.

- $\boldsymbol{\varepsilon_{\text{method}}}$ is the error of the functional itself. It does not depend on the basis. It is what benchmark papers report, using large basis sets.
- $\boldsymbol{\varepsilon_{\text{basis}}}$ is the error caused by the finite basis. It goes to zero as the basis grows, as you saw in Experiment 1.

In practice there are other terms too: the integration grid (Lecture 11), SCF convergence and stability (Lecture 12), the geometry, and the uncertainty of the reference itself. With the settings used in this lab they are usually smaller than the first two, but they are not always negligible.

The two main errors have **different origins**, and so they have **independent signs**. When they have the same sign, they add up. When they have opposite signs, they partly **cancel**, and the total error can be smaller than the method error alone (Figure 20). A small basis can then make a method look better than it really is. This is luck, not accuracy: change the reaction, and the two errors may no longer have opposite signs.

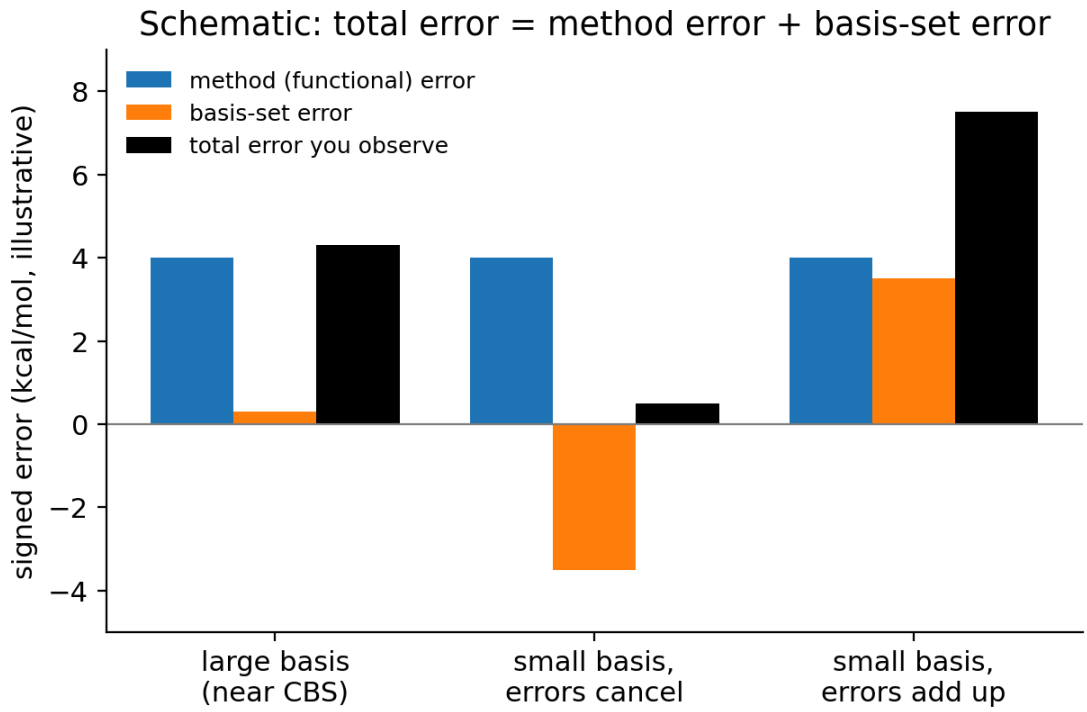


Figure 20: Schematic error budget. In each case the method error is the same (+4 kcal/mol). Near the CBS limit the basis error is small, and the total error is essentially the method error. With a small basis, the basis error can have the opposite sign (errors cancel, the result looks good) or the same sign (errors add up). The numbers are illustrative, not results.

Worked example: splitting the error

A functional gives these results for one reaction (illustrative numbers, not from any experiment in this course): $\Delta E = -26.00$ kcal/mol with a large, nearly complete basis, and $\Delta E = -29.50$ kcal/mol with a small basis. The reference is −30.00 kcal/mol. Treating the large-basis value as the CBS limit:

- $\varepsilon_{\text{method}} = -26.00 - (-30.00) = +4.00$ kcal/mol;
- $\varepsilon_{\text{basis}} = -29.50 - (-26.00) = -3.50$ kcal/mol;
- $\varepsilon_{\text{total}} = -29.50 - (-30.00) = +0.50$ kcal/mol $= 4.00 + (-3.50)$.

The small-basis result is eight times closer to the reference than the large-basis one. But the functional did not get better: two errors of opposite sign cancelled. This is the situation in the middle of Figure 20.

10.2.2 Basis-set incompleteness error

The **basis-set incompleteness error** (BSIE) is the error in a total energy caused by the finite basis. For HF and DFT, the variational principle guarantees that total energies decrease toward the CBS limit as the basis grows (Lecture 2). The BSIE of a total energy is therefore always positive. But a reaction energy is a *difference*, and its BSIE can have either sign. It depends on which side of the reaction is described worse by the basis.

A basis is **balanced** for a reaction if it describes reactants and products about equally well, so that their BSIEs cancel. Basis sets are usually less balanced when:

- the bonding changes a lot (new bonds formed, lone pairs created or destroyed);
- one side has atoms and the other a molecule, as in an atomization: an atom needs fewer functions than the same atom in a molecule, where the density must polarize into bonds;
- one side has anions or diffuse electrons and the other does not;
- the basis lacks **polarization functions** on some atoms. Without p functions on hydrogen, for example, the density around H cannot polarize into a bond.

10.2.3 Basis-set superposition error

The **basis-set superposition error** (BSSE) appears whenever two fragments come close: two molecules in a dimer, or two distant parts of one molecule. Each fragment is described by its own basis functions, but in the complex it can also use the functions centered on its partner. The isolated fragment has no such help. So the complex is described with an effectively larger basis than the separated fragments, and it comes out artificially **too stable**. For a dimer, BSSE always makes the binding too strong.

BSSE is not limited to dimers. In a large, compact molecule, an atom can borrow functions from other atoms that are close in space but far along the chain. This **intramolecular BSSE** favors compact structures over extended ones. It matters, for example, when comparing a branched alkane with a linear one, or a folded conformer with an unfolded one [120].

BSSE shrinks quickly as the basis grows, because a fragment that already has a nearly complete basis gains little from its partner. It is a significant problem for double-ζ sets and much smaller for triple-ζ and quadruple-ζ sets. It is a different thing from BSIE, although both are consequences of an incomplete basis.

10.2.3.1 The counterpoise correction

Boys and Bernardi proposed a simple way to remove most of the BSSE of a dimer AB [26]: compute the monomers with the *same* basis they enjoy in the dimer. In practice, the monomer A is computed at its geometry in the dimer, together with the basis functions of B placed at B's atomic positions, but without B's nuclei and

electrons. These are called **ghost atoms**. Writing E_X^Y for the energy of fragment X computed in the basis of Y, all at the dimer geometry, the **counterpoise-corrected** interaction energy is

$$\Delta E_{\mathrm{int}}^{\mathrm{CP}} = E_{AB}^{AB} - E_{A}^{AB} - E_{B}^{AB}. \tag{87}$$

The uncorrected interaction energy uses E_A^A and E_B^B instead. The difference between the two is the counterpoise estimate of the BSSE:

$$\delta_{\mathrm{BSSE}} = \left(E_A^{AB} - E_A^A \right) + \left(E_B^{AB} - E_B^B \right) \leq 0. \tag{88}$$

Each term is negative (or zero), because a larger basis can only lower the energy variationally. If the monomers relax their geometry in the dimer, a small deformation term is added, but the idea is the same. Q-Chem and most other programs can do counterpoise calculations with ghost atoms.

The counterpoise correction is the standard tool for dimers, but it has limits. It is defined for well-separated fragments, not for intramolecular BSSE. And it removes the BSSE but not the BSIE, so a counterpoise-corrected double-ζ result is still far from the CBS limit. For large systems, Kruse and Grimme proposed a geometric counterpoise correction (gCP) that estimates the inter- and intramolecular BSSE from the geometry alone [120].

10.2.4 Why B3LYP/6-31G* became so popular

10.2.4.1 A brief history

In the 1990s two things happened at once. Becke introduced hybrid functionals [79], and in 1994 Stephens, Frisch and co-workers defined B3LYP [80], which quickly became available in the most widely used quantum chemistry program of the time (Gaussian). The 6-31G* basis set of Pople and co-workers [15], [16] was already one of the most used basis sets, in that program and in the textbooks. Computers were slow: a double-ζ basis was what you could afford for a molecule with ten or twenty heavy atoms.

B3LYP/6-31G* was a huge improvement over what organic chemists used before (Hartree–Fock, semi-empirical methods). Geometries were good, reaction energies were often within a few kcal/mol, and calculations finished overnight. Papers that used it worked, so the next papers used it too. B3LYP's original papers became some of the most cited in all of chemistry, and the combination B3LYP/6-31G* became a de facto standard, often chosen because "everybody uses it" rather than because it was tested for the problem at hand.

10.2.4.2 Right for the wrong reasons

B3LYP has no dispersion (Lecture 8). With a complete basis, it underbinds dimers and gives too little stability to compact molecules. 6-31G* is a small, unbalanced basis with a large BSSE, which artificially stabilizes compact structures and complexes. The two errors have **opposite signs**. In many organic systems, the BSSE of 6-31G* partly compensates the missing dispersion of B3LYP, and the combination looks better than B3LYP with a large basis.

Kruse, Goerigk and Grimme analyzed this in detail in 2012 [121]. They showed that much of the apparent success of B3LYP/6-31G* comes from this fortuitous cancellation, that it fails badly when the cancellation does not happen, and that adding dispersion and BSSE corrections (or simply using a better basis and a modern functional) removes the problem. Their title says it plainly: the standard B3LYP/6-31G* model chemistry should not be used for molecular thermochemistry. A broader analysis of small basis sets and their error sources was given by Sure, Brandenburg and Grimme [122].

The worrying part is not that B3LYP/6-31G* is sometimes accurate. It is that we cannot predict **when**. Lucky cancellations depend on the reaction. Experiment 6 lets you see this for yourself.

⚠ Common mistakes with small basis sets

- Concluding that a small basis is "more accurate" because it gave a smaller error once.
- Adding -D4 to a functional used with a small basis without thinking: if BSSE is already mimicking dispersion, adding -D4 counts the attraction twice and can overbind.
- Comparing two functionals with a small basis and drawing conclusions about the functionals. Compare methods at a (nearly) converged basis first.
- Assuming all double-ζ sets are equivalent because they share the label "DZ".

10.2.5 Not all double-ζ sets are equal

10.2.5.1 Four polarized double-ζ sets

All four basis sets of Experiment 6 are "polarized double-ζ": two functions per valence atomic orbital, plus polarization functions. But they were built differently:

- **6-31G*** (Pople) [15], [16]: a split-valence set from the 1970s. The star adds one set of d functions on non-hydrogen atoms only. Hydrogen has no polarization functions (that would be 6-31G**). The s and p functions of each valence shell share exponents ("sp shells"), which was a trick to speed up integral evaluation on old computers.
- **def2-SVP** (Karlsruhe) [17]: split valence plus polarization on all atoms, including p functions on hydrogen. Optimized mostly at the HF level and designed to be balanced across the periodic table.
- **cc-pVDZ** (Dunning) [19]: correlation consistent. The polarization functions were chosen to recover correlation energy in wave-function methods. Same composition as def2-SVP for first-row atoms, but different exponents and contractions.
- **pc-1** (Jensen) [22], [123]: polarization consistent, optimized specifically for DFT. If it is not available, the smallest Karlsruhe set, def2-SV(P), which drops the p functions on hydrogen, is a fallback.

10.2.5.2 How to count basis functions

The number of basis functions K (or N_{bf}) sets the cost of a calculation. You can find it in every Q-Chem output, but you should also be able to count it by hand. You need the **contracted** composition of the basis for each element, written in square brackets as the number of contracted s, p, d, … shells. Each shell contributes:

- s: 1 function;
- p: 3 functions;
- d: 5 functions if spherical ("pure", 5d), or 6 if Cartesian (6d);
- f: 7 (spherical) or 10 (Cartesian).

The compositions you need for first-row atoms and hydrogen are:

Table 20: Contracted compositions of the four polarized double-ζ sets for H and first-row atoms. For 6-31G*, the core 1s is one contracted s function, and the inner and outer valence shells each contribute one s and one p shell.

Basis	H	C, N, O
6-31G*	[2s]	[3s2p1d]
def2-SVP	[2s1p]	[3s2p1d]
cc-pVDZ	[2s1p]	[3s2p1d]
pc-1	[2s1p]	[3s2p1d]

Two details matter:

1. **Spherical or Cartesian d.** The Pople sets were defined with six Cartesian d functions; the def2, cc and pc sets use five spherical d functions. Many programs follow these conventions by default, but they can be changed. Always check the number your output reports.
2. **Same count, different quality.** Two sets with the same number of contracted functions can differ in the number of primitive Gaussians in each contraction and in the exponents. That affects both the accuracy and the cost of computing the integrals.

Worked example: water and ethene

Water, H_2O. One O and two H.

- 6-31G*: O = 3 s + 2 × 3 p + 1 × 6 d (Cartesian) = 3 + 6 + 6 = 15; each H = 2. Total = 15 + 2 × 2 = **19** (or 18 with spherical d).
- def2-SVP, cc-pVDZ, pc-1: O = 3 + 6 + 5 = 14; each H = 2 + 3 = 5. Total = 14 + 2 × 5 = **24**.

Ethene, C_2H_4. Two C and four H.

- 6-31G*: 2 × 15 + 4 × 2 = **38** (or 36 with spherical d).
- def2-SVP, cc-pVDZ, pc-1: 2 × 14 + 4 × 5 = 28 + 20 = **48**.

The difference comes almost entirely from hydrogen. In 6-31G* each H has 2 functions; in the other three it has 5, because of the p polarization functions. The more hydrogen atoms a molecule has, the bigger the gap. In Experiment 6 you will do the same count for cyclohexene and compare it with your outputs.

Since cost grows as a power of K (Experiment 1), a basis with 25% more functions can cost roughly twice as much. That was a real advantage of 6-31G* in the 1990s. Today it is rarely a good reason.

10.3 Experiment 6 briefing

10.3.1 The goal

Experiment 6 asks a simple question: is B3LYP/6-31G* special? You will run B3LYP and B3LYP-D4 with four polarized double-ζ basis sets and compare them with the def2-TZVP results you already have from Experiments 4 and 5.

Reaction	Description	ΔE_{ref} (kcal/mol)
4a	ethene + butadiene ⟶ cyclohexene (Diels–Alder)	−45.40

Reaction	Description	ΔE_{ref} (kcal/mol)
4b	*n*-hexane ⟶ neohexane	−2.49
5b	CaO ⟶ Ca + O	96.15

The three reactions were chosen because they respond to small basis sets in very different ways. Your job is to find out how, and to explain why.

10.3.2 Methods and jobs

- Functionals: **B3LYP** and **B3LYP-D4**.
- Basis sets: **6-31G***, **def2-SVP**, **cc-pVDZ** and **pc-1**, plus **def2-TZVP** as the comparison.
- Eight molecules × 2 functionals × 4 small basis sets = **64 new jobs**. They are all small; running one basis set at a time works well.
- The def2-TZVP energies are not recomputed: the **Import** button on the results form copies your B3LYP and B3LYP-D4/def2-TZVP energies from Experiments 4 and 5 (from this browser). If you worked on another computer, load your saved files on those pages first.
- For every job, record the total energy and the **number of basis functions**. The *basis-size* table uses them to compare the four sets for cyclohexene.

10.3.3 What to watch out for

- **Think in terms of Equation 86.** For each reaction, the def2-TZVP result is your best estimate of the B3LYP method error. The difference between a small-basis result and the def2-TZVP result estimates the basis-set error. Ask whether the two have the same or opposite signs.
- **The anomaly spotter.** The *spot the anomaly* chart asks you to flag cells where a small basis does better than def2-TZVP. Flag them before you explain them, then ask, for each one: is this a better description, or a cancellation of errors?
- **-D4 with small basis sets.** Ask whether BSSE and -D4 describe the same attraction or different ones, and what that implies when both are present.
- **CaO.** Ca is a period-4 element with a large, polarizable electron cloud. Think about how well each double-ζ set might describe it, and remember that the products are atoms.
- **pc-1.** If `BASIS = pc-1` is not available on the server, use def2-SV(P) and say so in your report.
- **Optional.** Run the three reactions with ωB97M-V/6-31G* and compare with your ωB97M-V/def2-TZVP results. Does a modern functional benefit from the small basis the same way?
- Lock in your predictions on the experiment page before you run anything.

10.3.4 Key points

- In Experiment 5, errors grew to tens of kcal/mol for hard reactions; HF and LSDA failed in opposite directions on dissociations, and ozone was hard for nearly everyone.
- -D4 corrects dispersion only; it does not fix self-interaction, static correlation or open-shell atoms.
- MN15′s good result on CaO comes from its training set. Judge a functional on data it was not fitted to.
- Total error = method error + basis-set error (+ smaller terms). The two have independent signs and can cancel.
- BSIE is the error from a finite basis; BSSE is the artificial stabilization of a complex (or a compact molecule) that borrows its partner's basis functions. Both shrink as the basis grows.
- The counterpoise correction computes the monomers in the full dimer basis using ghost atoms: $\Delta E_{\text{int}}^{\text{CP}} = E_{AB}^{AB} - E_{A}^{AB} - E_{B}^{AB}$.

- B3LYP/6-31G* became popular for historical and practical reasons. Its successes often come from BSSE compensating missing dispersion, a cancellation that is unpredictable.
- To count basis functions, add up the contracted shells of each atom: s = 1, p = 3, d = 5 (spherical) or 6 (Cartesian). 6-31G* has no polarization on H, so it is smaller than def2-SVP, cc-pVDZ and pc-1.

10.3.5 Study questions

1. Why does -D4 barely change the error of B3LYP on the atomization of H_2SO_4?
2. A functional has errors of +0.3, +2.5 and +25 kcal/mol in Experiments 3, 4 and 5. Another has +0.5, +1.5 and +8. Which would you choose for a new organic reaction, and what else would you want to know?
3. Using Equation 86, a method gives ΔE = 10.0 kcal/mol at the CBS limit and 12.5 kcal/mol with a small basis; the reference is 12.0 kcal/mol. Compute $\varepsilon_{\text{method}}$, $\varepsilon_{\text{basis}}$ and $\varepsilon_{\text{total}}$. Is the small-basis result trustworthy?
4. Explain the difference between BSIE and BSSE. Which one can be corrected with the counterpoise method?
5. Write the counterpoise-corrected interaction energy for a water dimer. Which calculations do you need, and which use ghost atoms?
6. Why is the counterpoise estimate of the BSSE (Equation 88) always negative or zero?
7. Explain how the BSSE of 6-31G* and the missing dispersion of B3LYP can cancel. Give an example of a reaction where you would expect them not to cancel.
8. Count the basis functions of methanol (CH_3OH) in 6-31G* (Cartesian d) and in def2-SVP.
9. Why does the difference in size between 6-31G* and def2-SVP grow with the number of hydrogen atoms?
10. A colleague reports a reaction energy with B3LYP-D4/6-31G* that agrees with experiment within 0.5 kcal/mol. What would you ask before believing that the method is reliable for this reaction?

10.3.6 Further reading

- H. Kruse, L. Goerigk and S. Grimme, “Why the standard B3LYP/6-31G* model chemistry should not be used in DFT calculations of molecular thermochemistry: understanding and correcting the problem” [121].
- S. F. Boys and F. Bernardi, the original counterpoise paper [26].
- R. Sure, J. G. Brandenburg and S. Grimme, “Small atomic orbital basis set first-principles quantum chemical methods for large molecular and periodic systems: a critical analysis of error sources” [122].
- F. Jensen, “Atomic orbital basis sets” [27]. A review of basis-set families and how they are built.
- P. Morgante and R. Peverati, “The devil in the details” [32].

11 Numerical integration grids

> **i** Learning goals
>
> After this lecture you should be able to:
>
> - explain why B3LYP/6-31G* can beat B3LYP/def2-TZVP for one reaction and fail badly for another, and why that should worry you;
> - explain why the exchange–correlation energy has to be computed numerically, while the Hartree–Fock energy does not;
> - describe Becke's partition of a molecule into fuzzy atomic cells, and the radial × angular grid built on each atom;
> - read grid names such as (75,302), SG-1, SG-3 and `XC_GRID 000099000590`, and estimate how many points they contain;
> - explain why meta-GGAs and heavily parametrized functionals need finer grids;
> - plan and run Experiment 7, and collect its 480 results with the extraction script.

11.1 Part 1 · Looking back at Experiment 6

> **!** Read this after you submit Experiment 6
>
> This part discusses the trends you most likely saw in Experiment 6. Your own numbers may differ a little. What matters is the reason behind each trend.

In Experiment 6 you ran B3LYP and B3LYP-D4 with four polarized double-ζ basis sets (6-31G*, def2-SVP, cc-pVDZ and pc-1). You compared them with the def2-TZVP results from Experiments 4 and 5 on three reactions: the Diels–Alder reaction **4a** (reference −45.4 kcal/mol), the alkane isomerization **4b** (−2.49 kcal/mol) and the CaO dissociation **5b** (96.15 kcal/mol). The reactions were chosen because they react to a small basis in three different ways.

11.1.1 How to read the charts

The **MAE by basis set** chart shows one bar per basis set for each functional. If the variational principle were the whole story, the bars would get shorter from double-ζ to triple-ζ for every reaction. The **Spot the anomaly** panel asked you to flag, before you saw the numbers, every cell where you expected a small basis to beat def2-TZVP. Each confirmed anomaly is a small basis "winning". The **How big is each basis?** table counts the basis functions for one molecule. It reminds you that "double-ζ" is a family name, not a size.

11.1.2 Three reactions, three behaviors

4a, the Diels–Alder reaction. Most of you probably found that B3LYP/6-31G* is *closer* to the reference than B3LYP/def2-TZVP, and by a wide margin. This is not because 6-31G* is a better basis. At a nearly converged basis, B3LYP makes this reaction too little exothermic: it has no dispersion (Lecture 8), and it has a known tendency to underbind when π bonds turn into σ bonds. A small basis makes an error of its own, and here that error has the opposite sign. The product, cyclohexene, is one compact molecule with two new C–C σ bonds. With a small basis it can "borrow" functions from neighboring atoms more than the two separate reactants can (the intramolecular cousin of basis-set superposition error, Lecture 10). A small basis therefore favors the product and makes the reaction more exothermic. Two errors with opposite signs cancel, and the result looks good.

4b, n-hexane → neohexane. Here the small bases probably gave nearly the same answer as def2-TZVP. Both sides of the reaction have the same number and kind of bonds, so the basis-set errors almost cancel between reactant and product. B3LYP's error for this reaction is mostly the missing intramolecular dispersion, which no basis set can supply. -D4 fixes most of it.

5b, CaO → Ca + O. This is where 6-31G* probably failed badly, by tens of kcal/mol, while the other double-ζ sets stayed in a more reasonable range. CaO is a strongly ionic molecule (close to $Ca^{2+}O^{2-}$), and the products are two neutral atoms. Nothing on one side of the reaction resembles the other side, so there is no cancellation to rely on. The 6-31G* set for calcium is a much later addition to the Pople family [124] than the sets for C, H and O, and it describes the ionic molecule and the separated atoms very unevenly. The other three double-ζ families were built to be more balanced across the periodic table [17], [22].

> ⚠ Common mistake: "the small basis wins, so it's better"
>
> A small basis that beats a large one for a single reaction tells you that two errors cancelled. It tells you nothing about the next reaction. Experiment 6 showed both outcomes with the *same* functional and basis: help for 4a, a disaster for 5b. A method you can trust gets better, or at least not worse, as the basis grows.

11.1.3 -D4 with a small basis

With def2-TZVP, adding -D4 to B3LYP improves 4a and 4b, as you saw in Experiment 4. With the double-ζ sets the picture is muddier, and for some reactions -D4 probably made things worse. The reason is **double counting**. Basis-set superposition error acts like an artificial attraction between fragments that are close together. Dispersion is a real attraction between the same fragments. A small basis already adds a fake version of the effect, and -D4 then adds the real one on top. The sum overbinds. Dispersion corrections are fitted with large basis sets, where BSSE is small, so they are meant to be used with them.

11.1.4 Not all double-ζ sets are equal

The basis-function table showed that 6-31G* is the smallest of the four sets for cyclohexene. Its "*" adds *d* functions on carbon only, with no polarization functions on hydrogen. def2-SVP, cc-pVDZ and pc-1 all put *p* functions on hydrogen, so they are larger even though they share the "polarized double-ζ" label. Size alone did not decide which basis gave the smallest error, which is one more sign that the errors you saw were not systematic.

11.1.5 Lessons from Experiment 6

1. **Converge the basis first, then judge the functional.** A functional's error is only meaningful at (or near) the complete-basis-set limit. In this course, def2-TZVP is our practical stand-in for it.
2. **A small basis that wins is a warning, not a result.** Kruse, Goerigk and Grimme reached the same conclusion for B3LYP/6-31G* on a much larger data set [121].
3. **Corrections are designed for a context.** -D4 assumes a reasonable basis. Using it with 6-31G* mixes a physical correction with an unphysical error.

The same logic returns today. The integration grid is a third numerical approximation, after the basis set and the SCF convergence threshold, and its errors can also hide or mimic chemistry.

11.2 Part 2 · Numerical integration grids

11.2.1 Why the exchange–correlation energy needs quadrature

Recall the Kohn–Sham energy (see Lecture 5):

$$E[\rho] = T_s[\rho] + E_{\mathrm{ne}}[\rho] + J[\rho] + E_{\mathrm{xc}}[\rho].$$

All terms except E_{xc} reduce to integrals over Gaussian basis functions. The kinetic energy, the nuclear attraction and the Coulomb energy J need one- and two-electron integrals such as $(\mu\nu|\lambda\sigma)$. The Gaussian product theorem (Lecture 2) turns every one of them into a closed formula. Hartree–Fock exchange is built from the same two-electron integrals. That is why a Hartree–Fock calculation is exact for a given basis set, up to the SCF convergence threshold.

The exchange–correlation energy is different. In its simplest form, LSDA exchange,

$$E_{\mathrm{x}}^{\mathrm{LDA}} = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} \int \rho(\mathbf{r})^{4/3}\, d\mathbf{r},$$

the integrand is the density to the power 4/3. The density is a sum of products of Gaussians, and a non-integer power of such a sum has no analytic integral. GGAs, meta-GGAs and every other rung of Jacob's ladder make the integrand even more complicated:

$$E_{\mathrm{xc}} = \int f\,\left(\,\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta, \tau_\alpha, \tau_\beta\,\right) d\mathbf{r}.$$

The only general way to evaluate it is **numerical quadrature**: replace the integral by a weighted sum over a finite set of points $\mathbf{r}_i$,

$$E_{\mathrm{xc}} \approx \sum_{i=1}^{N_{\mathrm{pts}}} w_i\, f(\mathbf{r}_i). \tag{89}$$

The same points are used for the exchange–correlation potential in the Kohn–Sham matrix,

$$V_{\mu\nu}^{\mathrm{xc}} = \int \chi_\mu(\mathbf{r})\, v_{\mathrm{xc}}(\mathbf{r})\, \chi_\nu(\mathbf{r})\, d\mathbf{r} \approx \sum_i w_i\, \chi_\mu(\mathbf{r}_i)\, v_{\mathrm{xc}}(\mathbf{r}_i)\, \chi_\nu(\mathbf{r}_i),$$

at every SCF cycle. The set of points and weights is the **integration grid**. A finer grid gives a more accurate sum and costs more time. Unlike the basis set, the grid is **not variational**: a coarse grid can make the energy too high or too low, and the error does not shrink smoothly.

11.2.2 The problem with molecules

A single, uniform 3D grid would be a terrible choice. The density has a sharp cusp at every nucleus and decays exponentially far away. Near a nucleus the integrand changes on a scale of hundredths of a bohr; in the bonding region, on a scale of a bohr. A grid fine enough for the core would waste millions of points in the valence region and in empty space.

Spherical grids centered on an atom work well for that atom. In spherical polar coordinates,

$$\int F(\mathbf{r})\, d\mathbf{r} = \int_0^\infty r^2\, dr \int_\Omega F(r,\theta,\phi)\, d\Omega,$$

so we need a set of **radial shells** and, on each shell, a set of **angular points**. A molecule has several nuclei, though, and each needs its own spherical grid. The question is how to combine them without counting any region twice.

11.2.3 Becke's fuzzy cells

In 1988 Becke solved this problem in a way that almost every program still uses [125]. The idea is to split the integrand into atomic pieces,

$$F(\mathbf{r}) = \sum_A w_A(\mathbf{r})\, F(\mathbf{r}), \qquad \sum_A w_A(\mathbf{r}) = 1 \;\; \text{at every point},$$

and to integrate each piece $w_A F$ on the spherical grid of atom A. The weight $w_A(\mathbf{r})$ is close to 1 near nucleus A and close to 0 near every other nucleus. The atoms are "fuzzy cells": like Voronoi cells, but with smooth boundaries.

For a pair of atoms A and B at distance R_{AB}, Becke defines the elliptical coordinate

$$\mu_{AB}(\mathbf{r}) = \frac{|\mathbf{r}-\mathbf{R}_A| - |\mathbf{r}-\mathbf{R}_B|}{R_{AB}},$$

which is -1 at nucleus A, $+1$ at nucleus B and 0 on the plane halfway between them. He then applies the polynomial

$$p(\mu) = \frac{3}{2}\mu - \frac{1}{2}\mu^3$$

three times, $p_3(\mu) = p(p(p(\mu)))$, and defines the cell function

$$s(\mu_{AB}) = \frac{1}{2}\left[\,1 - p_3(\mu_{AB})\right].$$

Each application of p makes the step sharper while keeping it smooth (Figure 21 b). The cell weight of atom A is the product over all other atoms, normalized:

$$P_A(\mathbf{r}) = \prod_{B \neq A} s\left(\,\mu_{AB}(\mathbf{r})\right), \qquad w_A(\mathbf{r}) = \frac{P_A(\mathbf{r})}{\sum_C P_C(\mathbf{r})}.$$

For atoms of different sizes, Becke shifts the boundary toward the smaller atom with an adjustment based on atomic radii. The full molecular integral becomes a double sum over atoms and their grid points:

$$\int F(\mathbf{r})\, d\mathbf{r} \approx \sum_A \sum_{i \in A} w_A(\mathbf{r}_{Ai})\, w_i\, F(\mathbf{r}_{Ai}). \tag{90}$$

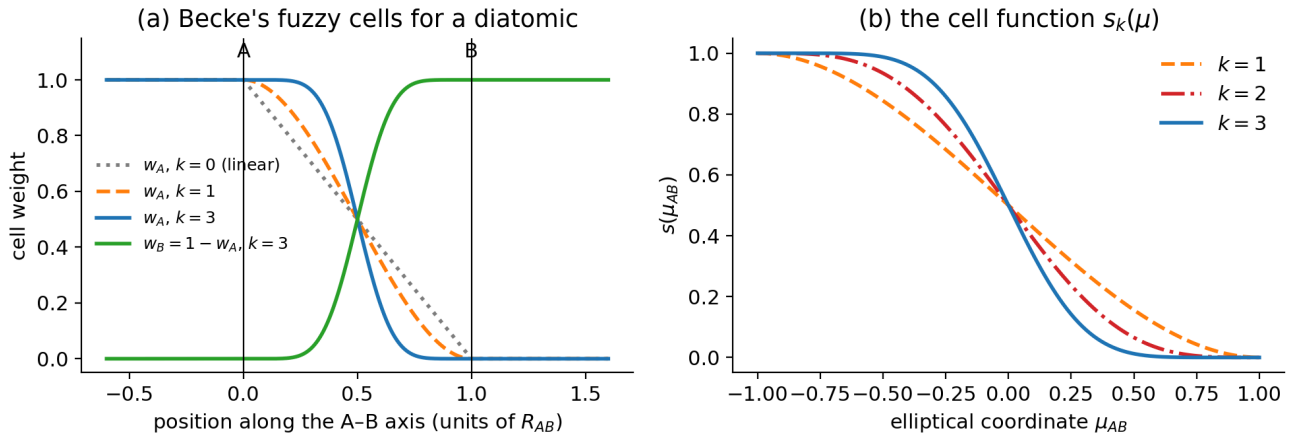


Figure 21: Becke's partitioning for a homonuclear diatomic. (a) Cell weights along the bond axis. With the plain linear function ($k = 0$) the cells overlap too much; three iterations of Becke's polynomial ($k = 3$) give a smooth but fairly sharp boundary halfway between the atoms. (b) The cell function $s_k(\mu)$ for $k = 1, 2, 3$. Exact functions from Becke's 1988 paper; no fitted data.

Worked example: weights at the midpoint

At the midpoint of a homonuclear diatomic, $\mu_{AB} = 0$. Then $p(0) = 0$, so $p_3(0) = 0$ and $s(0) = \frac{1}{2}$. Both atoms have $P = \frac{1}{2}$, and $w_A = w_B = \frac{1}{2}/(\frac{1}{2} + \frac{1}{2}) = \frac{1}{2}$. At a quarter of the way from A to B on the axis, $|\mathbf{r} - \mathbf{R}_A| = R/4$ and $|\mathbf{r} - \mathbf{R}_B| = 3R/4$, so $\mu = -\frac{1}{2}$. Then $p(-0.5) = -0.6875$, $p(-0.6875) = -0.8688$, $p(-0.8688) = -0.9753$, and $s = \frac{1}{2}(1 + 0.9753) = 0.988$. With $s(+\frac{1}{2}) = 1 - 0.988 = 0.012$ for atom B, $w_A = 0.988$. A point a quarter of the way along the bond belongs almost entirely to atom A.

The key consequence for us: **the grid moves with the atoms.** Each atom carries its own spherical grid, and the cell weights change whenever the distance between two atoms changes. Keep this in mind for Experiment 7.

11.2.4 The atomic grid: radial shells × angular points

Radial shells. The radial integral runs from 0 to ∞, so programs map it onto a finite interval and use a standard one-dimensional rule. Becke used Gauss–Chebyshev points x_i in (−1, 1) mapped with $r = r_m(1 + x)/(1 - x)$, where r_m is about half the atom's Bragg–Slater radius [125]. Murray, Handy and Laming proposed an Euler–Maclaurin scheme that many programs, Q-Chem among them, still use [126]. Every good radial scheme places many shells close to the nucleus, where the density changes fastest, and only a few far away (Figure 22).

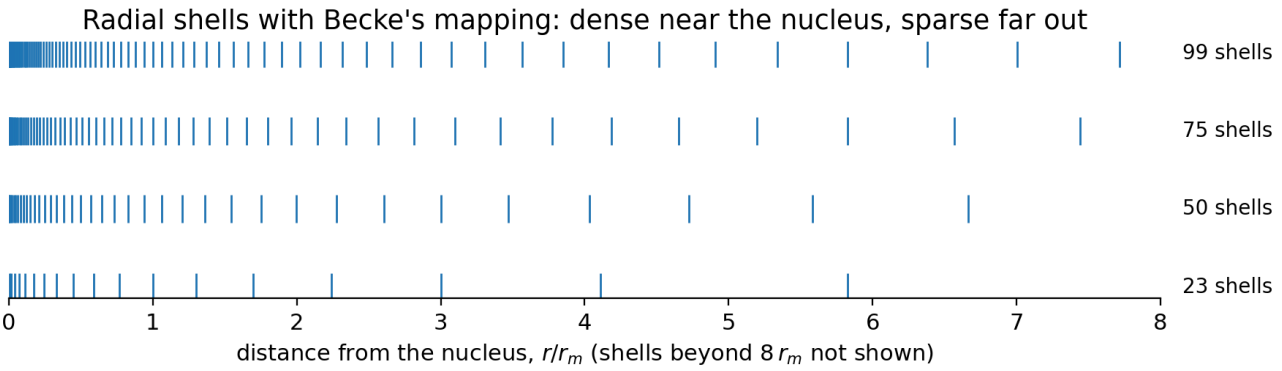


Figure 22: Radial shells generated by Becke's mapping for 23, 50, 75 and 99 shells. Each tick is one shell. Shells crowd near the nucleus and spread out at long range; shells beyond $8\,r_m$ are not shown. This is one common radial scheme, drawn from its exact formula; Q-Chem's SG-n grids use related schemes.

Angular points. On each shell the angular integral is done with a **Lebedev** grid [127]. Lebedev grids place points on the sphere with octahedral symmetry, and their weights are chosen so the grid integrates every spherical harmonic Y_{lm} up to a certain degree L exactly. They are the most efficient rules known for this job. Only certain sizes exist:

Table 21: Common Lebedev grids. A grid of degree L integrates all functions on the sphere that are combinations of Y_{lm} with $l \leq L$ exactly.

Lebedev points	110	194	302	590	974
exact up to degree L	17	23	29	41	53

The notation (n_r, n_Ω). A grid with n_r radial shells and n_Ω Lebedev points on each shell is written (n_r, n_Ω). The grid (75,302) has 75 shells × 302 points = 22,650 points per atom. The grid (99,590) has 58,410 points per atom.

11.2.5 Pruning

Close to a nucleus the density is almost spherical. Far from all nuclei it is small and smooth. Only in the valence region does the angular structure of bonds and lone pairs matter. A **pruned** grid uses this: it keeps the full angular grid only in the valence region and uses smaller Lebedev grids on the inner and outer shells. SG-1, for example, is built from a (50,194) grid by dividing the 50 shells into five regions with 6, 38, 86, 194 and 86 angular points [128]. Pruning removes about two thirds of the points with almost no loss of accuracy.

11.2.6 Standard grids in Q-Chem

Q-Chem offers four **standard grids**, SG-0 to SG-3, selected with the `$rem` keyword `XC_GRID` [129]:

Table 22: The standard grids in Q-Chem. Point counts are for one carbon atom, as listed in the Q-Chem manual; they differ a little between elements.

Grid	`XC_GRID`	Pruned from	Points for a carbon atom	Introduced by
SG-0	0	(23,170)	≈ 1,400	S.-H. Chien and P. M. W. Gill [130]
SG-1	1	(50,194)	≈ 3,800	P. M. W. Gill, B. G. Johnson, and J. A. Pople [128]
SG-2	2	(75,302)	≈ 7,800	S. Dasgupta and J. M. Herbert [131]
SG-3	3	(99,590)	≈ 17,700	S. Dasgupta and J. M. Herbert [131]

You can also ask for an unpruned grid directly. The value is written as two six-digit numbers glued together: the number of radial shells, then the number of Lebedev points. `XC_GRID 000099000590` means an unpruned (99,590) grid, and `XC_GRID 000075000302` an unpruned (75,302) grid. Experiment 9 uses this format.

The **default grid** depends on the program, its version and the functional. Recent Q-Chem versions use SG-1 for LSDA, most GGAs and hybrids, SG-2 for meta-GGAs and B97-based functionals, and SG-3 for Minnesota functionals [129]. Older versions used SG-1 for everything. Other programs have their own defaults. Because you can't be sure what a default means on a given server, **always set the grid explicitly and report it.**

Worked example: how many grid points for Ar_2?

With an unpruned (99,590) grid, each argon atom gets 99 × 590 = 58,410 points, so Ar_2 has 2 × 58,410 = 116,820. With SG-1 (pruned from 50 × 194 = 9,700 points per atom, roughly 40% kept), each atom has about 3,800–4,000 points. A calculation on a 50-atom organic molecule with SG-3 handles close to a million points at every SCF cycle. That is why programs screen out points where all basis functions are negligible.

11.2.7 How big is a grid error?

Grid errors in total energies are tiny on the scale of hartrees. They matter because reaction energies are small differences of large numbers, and the grid errors of the species in a reaction do not cancel reliably. Three effects make this worse:

1. **The grid moves with the atoms.** When a bond length changes, the points around each atom shift relative to the other atoms, and the cell weights change. The integration error changes erratically from one geometry to the next. A potential energy curve picks up small oscillations, or "wiggles", that have nothing to do with chemistry.
2. **Rotational variance.** A Lebedev grid has a fixed orientation in space. If you rotate a molecule while keeping the grid fixed, the energy changes slightly. Programs reduce this by putting every molecule in a standard orientation first. Two conformers or two molecules in a reaction have no common orientation, though, so the effect never vanishes completely.
3. **Derivatives are more sensitive than energies.** Forces and, especially, second derivatives amplify grid noise. Small vibrational frequencies (methyl rotations, soft torsions) can change by tens of cm^{-1} or even become imaginary on a coarse grid. That matters when you compute thermal corrections (Lecture 13).

Worked example: noise vs signal in Ar_2

The reference Ar_2 well depth in Experiment 7 is 0.28 kcal/mol. In hartree that is

$$\frac{0.28\ \text{kcal/mol}}{627.5095\ \text{kcal/mol per}\ \ E_h} = 4.5 \times 10^{-4}\ E_h.$$

The interaction energy is $\Delta E(r) = E(\mathrm{Ar_2}) - 2E(\mathrm{Ar})$, and each argon atom has a total energy of about $-527\ E_h$. So the binding energy is less than one part in a million of the total energy. A grid error of only $10^{-4}\ E_h$ in $E(\mathrm{Ar_2})$, about 0.06 kcal/mol, is already more than 20% of the well depth. For a reaction energy of 50 kcal/mol, the same error would be invisible.

11.2.8 Which functionals are grid-sensitive?

The quadrature error in Equation 89 depends on how smooth the integrand is. A smooth integrand is integrated well by few points; one with sharp features needs many. The integrand of a functional depends on its ingredients and on its functional form.

- **LSDA and GGAs** depend on ρ and $\nabla\rho$. Their integrands are fairly smooth, and SG-1 is usually adequate for energies.
- **Meta-GGAs** also depend on the kinetic-energy density $\tau(\mathbf{r}) = \frac{1}{2}\sum_i |\nabla\phi_i(\mathbf{r})|^2$. Many use τ through a ratio such as $\alpha = (\tau - \tau_\mathrm{W})/\tau_\mathrm{unif}$, where $\tau_\mathrm{W} = |\nabla\rho|^2/8\rho$ is the von Weizsäcker kinetic-energy density. In the tails of the density and between atoms, α is a ratio of two small numbers. It varies rapidly and is numerically delicate.
- **Heavily parametrized functionals** expand their enhancement factors in power series with many fitted coefficients, often large and of alternating sign. The fit makes the functional accurate on its training set but can make the enhancement factor oscillate as a function of α or of the reduced gradient. Oscillations in the functional become structure in the integrand, which a coarse grid samples badly.

Both effects are well documented. Wheeler and Houk showed that the M06 family gives unreliable torsional profiles and reaction energies with the grids that were defaults at the time [85], [132]. Mardirossian and Head-Gordon examined two generations of Minnesota functionals and found that meta-GGAs, in particular the heavily parametrized ones, need much finer grids than GGAs [133], [134]. More recently, Lehtola and Marques tested hundreds of functionals and found that many modern ones, including SCAN, converge impractically slowly with standard grids [135]. The design of a functional, not only its accuracy, decides how much grid it needs. Functionals designed with smoothness in mind, such as ωB97M-V, tend to be better behaved [90].

Aside: functionals with nonlocal correlation

Functionals with VV10 nonlocal correlation (B97M-V, ωB97M-V, the "-V" family) need a double integral over two grid points. Programs evaluate this part on a second, usually coarser grid that you can set separately (in Q-Chem, `NL_GRID`). The `XC_GRID` keyword controls only the semilocal part.

11.2.9 Hartree–Fock needs no exchange–correlation grid

Hartree–Fock has no E_xc. Its exchange energy is built from analytic two-electron integrals, so the HF energy does not depend on any grid. For a hybrid functional, the exact-exchange part is analytic and only the DFT

part is on the grid. Hybrids with a large fraction of exact exchange are therefore often (not always) less grid-sensitive than their semilocal parents.

> **i** Aside: grids in HF too?
>
> Some programs speed up large Hartree–Fock and hybrid calculations by evaluating exchange *seminumerically*, on a grid (for example the COSX or "chain-of-spheres" method). Those calculations do have a grid dependence. In this course, Q-Chem's HF calculations are fully analytic.

11.2.10 What a finer grid costs

The time spent on the exchange–correlation part of each SCF cycle grows roughly in proportion to the number of grid points. With screening, it also grows only about linearly with the size of the molecule. The other parts of the cycle (Coulomb and exact-exchange integrals, diagonalization) don't depend on the grid. How much the *total* time grows from SG-1 to SG-3 therefore depends on the functional and the basis set. You will measure it in Experiment 7, so it's worth making a prediction first.

> **⚠** Common mistakes with grids
>
> - **Mixing grids within one reaction.** Every species in a reaction energy must be computed with the same grid. Different grids, and even different radial schemes, give different absolute energies.
> - **Trusting the default.** The default grid changes between programs and versions. If your methods section says "default grid", nobody can reproduce your numbers.
> - **Blaming the functional for grid noise.** Wiggles in a potential curve, a spurious minimum, or a small imaginary frequency may be numerical. Check with a finer grid before you interpret them.
> - **Thinking HF has grid noise.** If an HF curve wiggles, look for another cause (SCF convergence, a symmetry change, a wrong geometry).

11.3 Experiment 7 briefing

11.3.1 Goal

Experiment 7 asks a simple question: how fine a grid does each functional need before its answer stops changing? You will find out on the argon dimer, where the signal is very small and grid noise has nowhere to hide.

11.3.2 The system and the reference

Ar_2 is held together only by dispersion. The geometries come from the RG10 subset [136] of MGCDB84 [63]: 31 Ar–Ar distances from 3.0 to 6.0 Å in steps of 0.1 Å, plus the argon atom. The interaction energy at each distance is

$$\Delta E(r) = E\left(\mathrm{Ar}_2(r)\right) - 2\,E(\mathrm{Ar}),$$

with signed stoichiometric coefficients $\nu = +1$ for the dimer and $\nu = -2$ for the atom. The reference curve is repulsive at short range (+2.24 kcal/mol at 3.0 Å), has a single minimum of −0.28 kcal/mol at 3.7–3.8 Å, and approaches zero slowly (−0.02 kcal/mol at 6.0 Å). The full table is on the experiment page.

11.3.3 Methods and jobs

Setting	Value
Functionals	B3LYP, M06-L, M11-L, M11, ωB97M-V
Grids	SG-1, SG-2, SG-3 (`XC_GRID` 1, 2, 3)
Basis set	def2-TZVP (with def2-QZVP, the SG-3 jobs would not fit in the 10-minute limit)
Structures	31 dimers + 1 atom = 32
Jobs	5 × 3 × 32 = **480**: seconds each with SG-1, a few minutes each with SG-2 and SG-3

Before you run anything, place the five functionals on Jacob's ladder (Lecture 6) and think about which ingredients each one uses. That is the basis for your predictions.

11.3.4 Running 480 jobs without losing your mind

- **Work one functional and one grid at a time.** Run the 32 jobs of one curve, check that the curve looks reasonable, then go on.
- **Run the atom with every functional and every grid.** The atom's energy must come from the same functional and grid as the dimer. That's why there are 32 jobs per curve, not 31.
- **Check `XC_GRID` in the inputs.** The website's input generator writes it for you. If you build inputs by hand and forget it, you get the default grid, which is not SG-1, SG-2 or SG-3 in general.
- Your instructor may split the functionals across the class.

11.3.5 Collecting the results with the script

Typing 480 total energies by hand is slow and error-prone. Experiment 7 is the only experiment where a script does it for you. Download extract_ex07.py and follow the instructions in the section Collect your results with a script of the experiment page. In short:

1. Put all your Experiment 7 output files in one folder (subfolders are fine).
2. Run `python3 extract_ex07.py <folder>` (macOS/Linux) or `py extract_ex07.py <folder>` (Windows). Google Colab works if you can't install Python.
3. The script reads every Q-Chem output, recognizes each job from the functional, `XC_GRID`, basis set and Ar–Ar distance that Q-Chem echoed, and writes `ex07_results.csv`. It never changes your outputs.
4. **Read the report it prints.** You need 32 usable jobs for every functional and grid, 480 in all. It lists missing jobs, failed or unconverged jobs (from which it takes no energy), duplicates and files it didn't recognize. Fix the problems and run it again.
5. On the results form, use **Import CSV** to load the file. The page computes every $\Delta E(r)$ and draws the curves.

⚠ Pitfalls with the script

- **Check its work** on two or three outputs, say one per grid: compare the total energy, SCF cycles and wall time with what you read in the file (Guide 3). The numbers you submit are your responsibility.
- **Don't open the CSV in Excel and save it.** Excel may round the energies or change the separators. Import the file exactly as the script wrote it.
- **Keep your output files.** The CSV is a summary, not a replacement.

11.3.6 What to look at

The results page has three views:

- **Curves**: your $\Delta E(r)$ for each functional, one panel per grid, with the reference as a dashed black line. Use the zoom switch to look inside the well.
- **Wiggle report**: for each functional and grid, the number of minima ("smooth" means exactly one, like the reference), the RMS deviation from the reference, and the largest difference from the same functional on SG-3. The last number measures grid convergence without using the reference at all.
- **Timing**: the total wall time of each curve by grid.

Figure 23, from the experiment page, shows what grid noise can look like for a functional that is *not* part of your set (VSXC). Your job is to find out which of your five functionals behave like this, and on which grids.

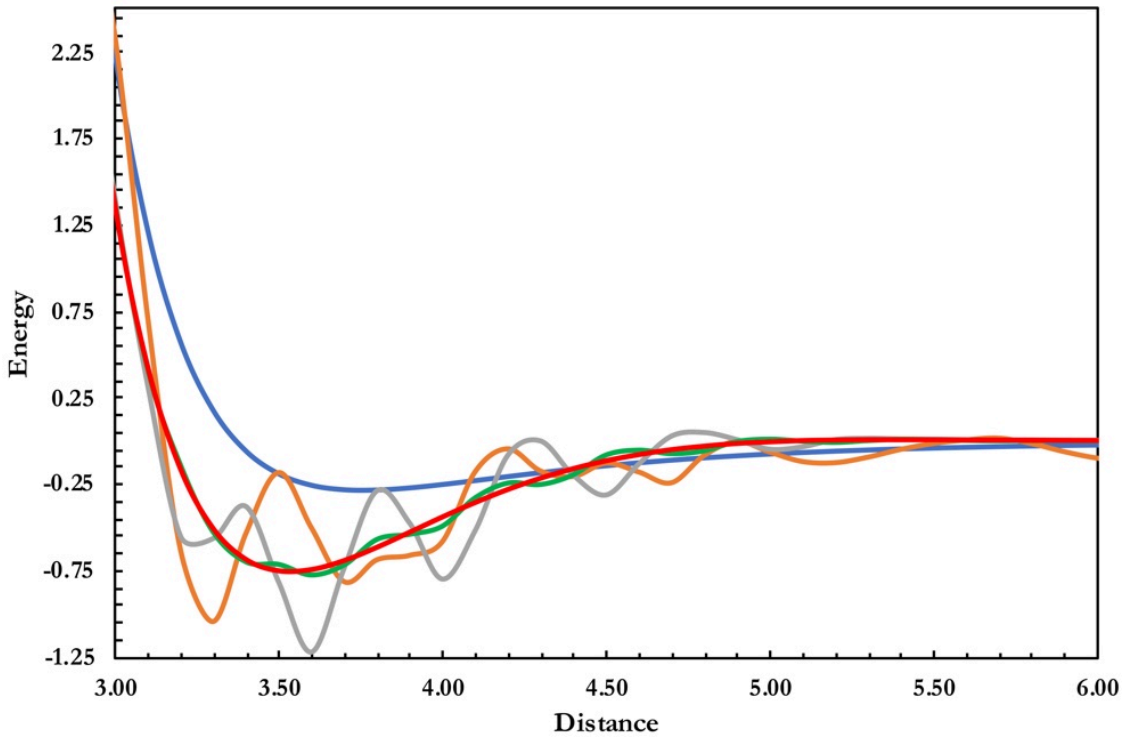


Figure 23: The VSXC functional on the Ar_2 curve with four integration grids: SG-1 (orange), (75,302) (grey), (99,590) (green) and (175,974) (red); reference in blue. Energy in kcal/mol, distance in Å. Reproduced from the Experiment 7 page.

Two questions to keep in mind while you work. First, a curve can be perfectly smooth and still wrong: grid convergence is not accuracy. Second, think about what the grid explorer on the experiment page shows. Why would a wiggle appear at one distance and not at the next?

11.3.7 Key points

- B3LYP/6-31G* can beat B3LYP/def2-TZVP when the basis-set error and the functional error cancel. The cancellation is luck: Experiment 6 showed it helping for 4a and failing badly for 5b.
- Dispersion corrections are designed for converged basis sets. With small basis sets, BSSE and -D4 can double count.
- The exchange–correlation energy has no analytic integral over Gaussians. It is computed by numerical quadrature on a grid of points around each atom, at every SCF cycle.
- Becke's scheme splits a molecule into fuzzy atomic cells with smooth weights. Each atom carries a radial × angular (Lebedev) grid, written (n_r, n_Ω).
- Pruned standard grids SG-0 to SG-3 are selected in Q-Chem with `XC_GRID` 0–3. `XC_GRID 000099000590` is an unpruned (99,590) grid.
- Grid errors are not variational. They change erratically with geometry because the grid moves with the atoms, and they are largest for small energy differences and for derivatives.
- Meta-GGAs (through τ) and heavily parametrized functionals need finer grids than GGAs. Hartree–Fock needs no exchange–correlation grid.
- Always set the grid explicitly, use the same grid for every species in a reaction, and report it.

11.3.8 Study questions

1. For reaction 4a, B3LYP at the basis-set limit and the B3LYP/6-31G* basis-set error have opposite signs. Explain, in terms of the reactants and the product, why a small basis tends to favor the product.
2. Why is the cancellation of basis-set errors nearly complete for 4b but absent for 5b?
3. Explain "double counting" between BSSE and -D4. Would you expect the same problem with def2-QZVP? Why or why not?
4. Show that Becke's polynomial $p(\mu) = \frac{3}{2}\mu - \frac{1}{2}\mu^3$ satisfies $p(\pm 1) = \pm 1$ and $p'(\pm 1) = 0$. Why are these properties useful for a cell function?
5. How many grid points per atom are in the unpruned grids (50,194), (75,302) and (99,590)? By what factor does the point count grow from the first to the last?
6. A colleague reports a reaction energy computed with "B3LYP/def2-TZVP, default settings". List what you would need to know about the grid to reproduce the number.
7. Explain, without equations, why the integration error changes from one Ar–Ar distance to the next.
8. Why are meta-GGAs more grid-sensitive than GGAs? Why can a heavily parametrized functional be more sensitive than a simple one on the same rung?
9. Would a Hartree–Fock Ar_2 curve show grid wiggles in Q-Chem? Would a B3LYP curve? Explain the difference.
10. Your geometry optimization with a meta-GGA ends with one small imaginary frequency of 25i cm^{-1}. What would you check before concluding that the structure is a transition state?

11.3.9 Further reading

- A. D. Becke, "A multicenter numerical integration scheme for polyatomic molecules" [125]. The original fuzzy-cell paper; short and readable.
- P. M. W. Gill, B. G. Johnson and J. A. Pople, the SG-1 paper [128], and S. Dasgupta and J. M. Herbert on SG-2 and SG-3 [131]: how a standard grid is designed and tested.

- S. E. Wheeler and K. N. Houk on grid sensitivity of the M06 family [132], and N. Mardirossian and M. Head-Gordon on grid sensitivity of meta-GGAs [133].
- S. Lehtola and M. A. L. Marques, “Many recent density functionals are numerically ill-behaved” [135].
- P. Morgante and R. Peverati, “The devil in the details” [32]: the review behind several experiments of this course, including the grid experiment.
- H. Kruse, L. Goerigk and S. Grimme on why B3LYP/6-31G* is “right for the wrong reasons” [121].
- The Q-Chem manual, section on DFT numerical quadrature [129].

12 SCF stability, open shells and transition metals

i Learning goals

After this lecture you should be able to:

- tell grid noise from physics in a potential energy curve, and choose a grid for a given functional;
- explain why a converged SCF solution is a stationary point that may not be a minimum;
- define the electronic Hessian and say what a negative eigenvalue means;
- distinguish internal from external (restricted → unrestricted, real → complex) instabilities;
- compute the exact $\langle S^2 \rangle$ for a spin state and judge spin contamination in an unrestricted calculation;
- explain why transition-metal atoms have many low-lying states, and what an effective core potential does;
- plan Experiment 8.

i Note

This lecture may take two meetings. If it does, the natural split point is just before Section 12.3.2: the first meeting covers the Experiment 7 debrief, SCF stability and spin contamination, the second covers transition metals, effective core potentials and the Experiment 8 briefing.

12.1 Part 1 · Looking back at Experiment 7

! Read this after you submit Experiment 7

This part discusses the trends you most likely saw in Experiment 7. Your curves may differ in the details, and a classmate's may too. What matters is how to read them.

12.1.1 Wiggles are grid noise, not physics

The reference Ar_2 curve has one minimum. Most of you probably found curves with two, three or more minima, especially on SG-1. Those extra minima are not chemistry. Argon atoms do not have several preferred distances. As Lecture 11 explained, the grid moves with the atoms, so the integration error changes erratically from one distance to the next. The binding energy is only 0.28 kcal/mol, or $4.5 \times 10^{-4}\ E_h$, so an integration error of a few $10^{-5}\ E_h$ is enough to draw a new minimum.

Two simple tests separate noise from physics:

1. **Does the feature survive a finer grid?** A physical minimum stays where it is when you go from SG-2 to SG-3. A grid artifact moves, shrinks or disappears.
2. **Does it make physical sense?** A real potential between two closed-shell atoms has one well. Several minima a few tenths of an ångström apart are a red flag.

The **max Δ vs SG-3** column of the wiggle report measures the first test. It compares each curve with the same functional on the finest grid, without using the reference, so it measures grid convergence alone.

12.1.2 Which functionals needed the finest grid?

The typical pattern, which you probably reproduced:

- **SG-1 is too coarse for this problem for almost everyone.** Even B3LYP may show small ripples at SG-1 on a scale of 0.1 kcal/mol, invisible in ordinary reaction energies but obvious here.
- **The meta-GGAs M06-L and M11-L and the hybrid meta-GGA M11** were usually the most sensitive. Their curves often still had spurious minima at SG-2, and sometimes small irregularities even at SG-3. This matches what Wheeler and Houk [132] and Mardirossian and Head-Gordon [133] reported for Minnesota functionals.
- **ωB97M-V** is also a meta-GGA, but it was designed with smoothness in mind [90]. It usually converges by SG-2.
- **B3LYP** converges quickly with the grid, because a GGA hybrid has a smooth integrand.

12.1.3 Grid convergence is not accuracy

Now look at the *shape* of the converged curves. B3LYP has no dispersion, and argon is held together only by dispersion. On a converged grid, B3LYP's curve is probably repulsive at every distance, or nearly so. Any "well" it showed on SG-1 was made by the grid. Some Minnesota functionals bind Ar_2 only weakly on a converged grid. ωB97M-V includes VV10 nonlocal correlation and usually has a well of about the right depth. A smooth curve tells you the grid is good enough. It does not tell you the functional is right.

12.1.4 What the grid costs

Table 23: Grid sizes from Lecture 11 (Q-Chem manual values for carbon).

Grid	SG-1	SG-2	SG-3
Points for a carbon atom	≈ 3,800	≈ 7,800	≈ 17,700

The timing chart probably surprised some of you: going from SG-1 to SG-3 multiplies the number of grid points by more than four (Table 23), but the total wall time grows much less. Only the exchange–correlation part of each SCF cycle scales with the number of points. Compared with the effect of a larger basis set (Experiment 1), a finer grid is cheap insurance.

12.1.5 Recommendations

- Use **at least SG-2** for meta-GGAs, and **SG-3 or (99,590)** for Minnesota functionals and other heavily parametrized ones. SG-1 is acceptable for GGAs and GGA hybrids in ordinary reaction energies, but not for weak interactions or frequencies.
- Use **the same grid for every species** in a reaction.
- **Report the grid** in every methods section, by name or as (n_r, n_Ω).
- **Hartree–Fock has no grid problem.** If you answered question 7 with "no, because HF exchange is computed analytically", you were right.

12.2 Part 2 · Is the SCF solution the right one?

12.2.1 The SCF finds stationary points

In Lecture 4 you saw the SCF loop: guess orbitals, build the Fock (or Kohn–Sham) matrix, diagonalize, repeat until nothing changes (see Lecture 4). When it stops, the energy no longer changes to first order when you mix occupied and virtual orbitals. That is a statement about the **gradient**. It says nothing about the curvature.

To make this precise, write any small change of the occupied orbitals as a rotation that mixes occupied orbital i with virtual orbital a by an amount κ_{ai}:

$$\phi_i \to \phi_i + \sum_a \kappa_{ai}\,\phi_a + \dots$$

The energy is a function of all the rotation parameters $\boldsymbol{\kappa}$. Expanded around the converged solution,

$$E(\boldsymbol{\kappa}) = E_0 + \mathbf{g}^T\boldsymbol{\kappa} + \frac{1}{2}\boldsymbol{\kappa}^T\mathbf{H}\,\boldsymbol{\kappa} + \ldots$$

The SCF convergence criterion is $\mathbf{g} = 0$ (the off-diagonal occupied–virtual block of the Fock matrix vanishes; this is Brillouin's theorem). A point with zero gradient is a **stationary point**. It can be

- a **minimum**: all eigenvalues of $\mathbf{H}$ positive;
- a **saddle point**: some eigenvalues negative;
- a **maximum**: all eigenvalues negative (rare in practice).

The matrix $\mathbf{H}$ of second derivatives of the energy with respect to orbital rotations is the **electronic Hessian** (or orbital Hessian). Note what is being differentiated: the *energy*, with respect to *orbital rotations*. This is not the nuclear Hessian you use for vibrational frequencies.

12.2.2 A picture of an unstable solution

Figure 24 shows the toy model used by the stability-landscape widget of Experiment 8. One angle θ mixes two orbitals, and the energy is

$$E(\theta) = \Delta \sin^2\theta - \sin^2 2\theta.$$

At $\theta = 0$ the gradient is zero for every value of Δ, so an SCF can converge there. The curvature is $E''(0) = 2\Delta - 8$. For a large gap ($\Delta > 4$) the curvature is positive and $\theta = 0$ is a minimum. For a small gap ($\Delta < 4$) it is negative: $\theta = 0$ is a maximum along this direction, and lower solutions exist at $\cos 2\theta^* = \Delta/4$.

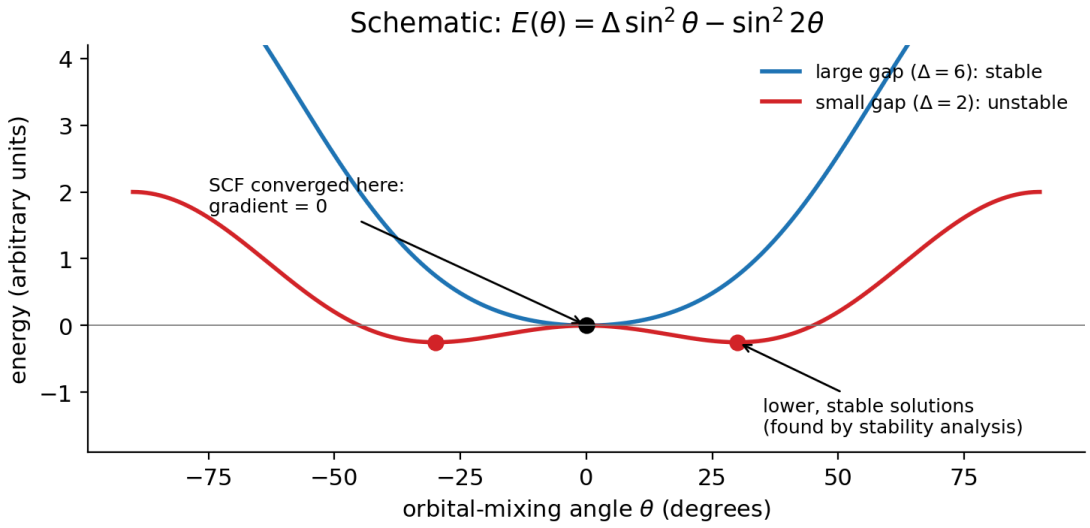


Figure 24: Toy model of SCF (in)stability (schematic; the same function as the Experiment 8 widget). With a large gap Δ between two low-lying states (blue), the converged point θ = 0 is a minimum. With a small gap (red), the same point still has zero gradient but negative curvature, and two lower solutions exist.

Worked example: when does the toy model become unstable?

For $\Delta = 2$, $E''(0) = 2(2) - 8 = -4 < 0$, so the converged point is unstable. The lower solutions satisfy $\cos 2\theta^* = 2/4 = 0.5$, so $2\theta^* = 60^\circ$ and $\theta^* = 30^\circ$. Their energy is

$$E(30^\circ) = 2\sin^2 30^\circ - \sin^2 60^\circ = 2(0.25) - 0.75 = -0.25,$$

below the converged value $E(0) = 0$. For $\Delta = 6$, $E''(0) = +4$ and there is nothing to fix. The crossover is at $\Delta = 4$, where the curvature changes sign. Small gaps between states make instabilities likely.

12.2.3 The electronic Hessian

For a real, closed-shell Hartree–Fock determinant, the Hessian has a simple block structure built from two matrices [137], [138]:

$$A_{ia,jb} = (\varepsilon_a - \varepsilon_i)\,\delta_{ij}\,\delta_{ab} + \langle aj \parallel ib\rangle, \qquad B_{ia,jb} = \langle ab \parallel ij\rangle,$$

where i, j label occupied orbitals, a, b virtual orbitals, ε orbital energies and $\langle pq \parallel rs\rangle$ antisymmetrized two-electron integrals. The solution is stable with respect to real rotations if $\mathbf{A} + \mathbf{B}$ has no negative eigenvalues, and with respect to complex rotations if $\mathbf{A} - \mathbf{B}$ has none. In Kohn–Sham DFT the same structure holds, with the exchange–correlation kernel (the second derivative of E_{xc}) added to the two-electron terms.

Two remarks help the intuition:

- The first term is the orbital-energy gap $\varepsilon_a - \varepsilon_i$. A large HOMO–LUMO gap makes the diagonal large and positive, and the solution is almost certainly stable. Small gaps are where trouble lives.
- $\mathbf{A}$ and $\mathbf{B}$ are the same matrices that appear in time-dependent HF and DFT for excitation energies. A negative eigenvalue of the Hessian corresponds to a "negative excitation energy": there is a state of the same kind *below* the one you converged to.

12.2.4 Internal and external instabilities

Every SCF calculation works within constraints: restricted or unrestricted spin orbitals, real or complex coefficients. These form a hierarchy [138], [139]:

$$\mathrm{RHF} \subset \mathrm{UHF} \subset \mathrm{GHF}, \qquad \text{real} \subset \text{complex}\ .$$

- An **internal instability** means a lower solution exists *within the same constraints*: for example, a UHF (or UKS) solution that can lower its energy by rotating to a different UHF solution. This is the typical case for open-shell atoms that converged to the wrong occupation of their d orbitals.
- An **external instability** means a lower solution exists only if you *relax a constraint*. The classic example is a restricted solution with a lower unrestricted solution (an **RHF → UHF**, or triplet, instability). Stretch H_2 and the RHF solution, stable near equilibrium, becomes externally unstable past the Coulson–Fischer point (roughly 1.2 Å for HF). From there on, UHF lets the two electrons localize on separate atoms and gives a lower energy.

A stability analysis computes the lowest eigenvalues of the Hessian. If one is negative, the program rotates the orbitals along the corresponding eigenvector, which is downhill by construction, converges the SCF again from there, and checks the new solution. It repeats until all eigenvalues are non-negative [140].

12.2.5 Stability analysis in Q-Chem

In Q-Chem, add one line to the `$rem` section:

```
INTERNAL_STABILITY  TRUE
```

After the SCF converges, Q-Chem checks for internal instabilities. If it finds one, it follows it, reconverges and checks again, and the output reports what it found and the new energy. The energy after a successful stability analysis can only be equal to or lower than the energy without it. Q-Chem [1] has separate options for external (restricted ⟶ unrestricted) checks; the manual describes them. In this course the open-shell atoms already use unrestricted orbitals, so internal stability is what matters.

A stability analysis costs extra time, usually a few times the cost of the SCF itself, because finding the lowest Hessian eigenvalues requires work similar to a few excited-state calculations. For a small molecule that is nothing. For a large one, run the analysis on the final geometry at least.

⚠ Common mistakes with SCF solutions

- **"It converged, so it's right."** Convergence means $\mathbf{g} = 0$, nothing more.
- **Comparing energies from different solutions.** If one state of a reaction was stabilized and the other wasn't, the reaction energy mixes two different kinds of solution.
- **Ignoring an energy that goes *up*.** A stable solution can never be higher than the unstable one it came from. If the "with stability" energy is higher, the two jobs did not use the same settings.

12.3 Part 3 · Open shells, spin contamination and transition metals

12.3.1 Spin and ⟨S²⟩

A state with total spin quantum number S is an eigenfunction of $\hat{S}^2$ with eigenvalue $S(S+1)$ (in units of $\hbar^2$), and has multiplicity $2S+1$. For N_α α electrons and N_β β electrons in the high-spin component, $S = M_S = (N_\alpha - N_\beta)/2$:

Table 24: Spin states and their exact ⟨S²⟩ values.

Multiplicity	S	exact $\langle S^2 \rangle$	Example in this course
doublet	1/2	0.75	H, OH, *tert*-butyl, Cl
triplet	1	2.00	O, C, excited Fe
quartet	3/2	3.75	N, ground-state Ru^+
quintet	2	6.00	ground-state Fe
sextet	5/2	8.75	excited Ru^+, FeCl

A **restricted open-shell** (ROHF/ROKS) determinant uses the same spatial orbitals for α and β electrons and is a proper spin eigenfunction. An **unrestricted** (UHF/UKS) determinant lets α and β orbitals differ. That gives more variational freedom and a lower energy, but the determinant is no longer an eigenfunction of $\hat{S}^2$. For UHF,

$$\langle S^2 \rangle_{\mathrm{UHF}} = S(S+1) + N_\beta - \sum_{i}^{N_\alpha} \sum_{j}^{N_\beta} \left| \langle \phi_i^\alpha | \phi_j^\beta \rangle \right|^2,$$

which is always at least $S(S+1)$. The excess is **spin contamination**: the determinant contains some admixture of higher spin states. Q-Chem prints ⟨S²⟩ after every unrestricted SCF.

Worked example: is this triplet contaminated?

A UKS calculation on a triplet ($S = 1$) prints $\langle S^2 \rangle$ = 2.03. The exact value is $S(S+1) = 2$, so the excess is 0.03, or 1.5%. That is small; the state is essentially a triplet. If it printed 2.60, the excess would be 0.60 (30%), a sign that the determinant is far from a pure triplet, for example a broken-symmetry mixture with a quintet ($\langle S^2 \rangle = 6$).

In Kohn–Sham DFT, $\langle S^2 \rangle$ of the determinant is not strictly the $\langle S^2 \rangle$ of the real system: the KS determinant describes a fictitious non-interacting system. It is still a useful diagnostic. A value far from $S(S+1)$ means the calculation has a problem worth looking at. Note that following an instability can *increase* spin contamination: the lower solution may break spin symmetry to gain energy.

12.3.2 Why transition metals are hard

In main-group chemistry, a closed-shell molecule usually has one dominant electron configuration with a large HOMO–LUMO gap, and the SCF converges to it easily. Transition metals break both assumptions.

Near-degenerate configurations. The 3d and 4s orbitals (4d and 5s in the second row) have similar energies. The iron atom has the ground configuration [Ar]$3d^6 4s^2$ (a quintet, ^{5}D), but states of the $3d^7 4s^1$ configuration lie only about 1–2 eV higher. Each configuration also has many ways to place the d electrons in five d orbitals, so there are many low-lying states and many SCF solutions close in energy. In the language of Figure 24, the gap Δ is small, and instabilities are common.

Symmetry breaking. An open-shell atom with a partially filled d shell has a spherically symmetric density only on average over all the degenerate components. A single KS determinant picks one component and has a non-spherical density. Different starting guesses can land on different components, which in an approximate functional need not have exactly the same energy.

Spin-state energetics. The energy difference between spin states of the same metal depends strongly on the functional. Exact (HF) exchange stabilizes parallel spins, so functionals with more HF exchange tend to favor high-spin states, and semilocal functionals tend to favor low-spin states. No single functional is reliable for all transition-metal spin-state gaps, which is why they are a demanding benchmark [141].

SCF convergence. With many nearly degenerate orbitals, the SCF can oscillate between occupations from one cycle to the next and never converge. Remedies exist (better guesses, damping, level shifting, direct-minimization algorithms), but some functionals fail more often than others. Experiment 8 keeps only the functional/atom pairs that converged reliably on the IQmol server; the ones that were removed are a lesson in themselves.

12.3.3 Effective core potentials and relativity

Ruthenium has 44 electrons and palladium 46. Their core electrons are chemically inert, but they cost basis functions and they move fast. Near a nucleus with charge Z, an inner electron's speed is a sizable fraction of the speed of light (roughly $Z/137$ in atomic units). **Scalar relativistic effects** contract and stabilize s and p orbitals, and through better screening expand and destabilize d and f orbitals. For 4d metals they shift orbital energies and excitation energies by amounts that matter chemically.

An **effective core potential** (ECP) solves both problems at once. It replaces the core electrons with a potential fitted to accurate relativistic atomic calculations, so the scalar relativistic effects of the core are built in [142]. Only the valence electrons are treated explicitly, with a basis set designed for that ECP. The def2 basis sets use

small-core ECPs for elements beyond krypton [17]: for the 4d metals, 28 core electrons (the shells 1s through 3d) are replaced.

Worked example: how many electrons does Q-Chem see for Ru^+?

Ruthenium has $Z = 44$. Ru^+ has 43 electrons. The def2 ECP replaces 28 core electrons, so the calculation treats $43 - 28 = 15$ electrons explicitly, moving in the field of an effective core of charge $44 - 28 = +16$. For the sextet, $N_\alpha - N_\beta = 5$ and $N_\alpha + N_\beta = 15$, so $N_\alpha = 10$ and $N_\beta = 5$. For the quartet, $N_\alpha = 9$ and $N_\beta = 6$.

In Q-Chem, ECPs are requested with `ECP = def2-ECP`; the website's input generator adds it for Ru and Pd. Iron (a 3d metal) is treated with all its electrons in def2 basis sets. Its relativistic effects are smaller, and in this course we ignore them. Spin–orbit coupling, the other relativistic effect, is not included in standard ECPs or in standard DFT. It returns in Lecture 13.

12.4 Experiment 8 briefing

12.4.1 Goal

Experiment 8 asks whether stability analysis changes anything for two transition-metal "reactions": excitations between spin states of an atom.

12.4.2 The two excitations

The data come from the 3dAEE8 [143], [144], [145] and 4dAEE5 [146] subsets of the Minnesota 2015B database [31]. Each excitation energy is a reaction energy with $\nu = -1$ for the ground state and $\nu = +1$ for the excited state:

Table 25: The two excitation energies of Experiment 8.

Reaction	Ground state	Excited state	Reference (kcal/mol)
Fe: quintet ⟶ triplet	Fe, charge 0, multiplicity 5	Fe, charge 0, multiplicity 3	34.32
Ru^+: quartet ⟶ sextet	Ru^+, charge +1, multiplicity 4	Ru^+, charge +1, multiplicity 6	21.96

$$\Delta E = E(\text{excited}) - E(\text{ground}).$$

Note that for Ru^+ the *higher* multiplicity is the excited state. Ruthenium's ground-state configuration has more d electrons than iron's, and the balance between d and s occupation is different in the two rows.

12.4.3 Methods and jobs

Table 26: Functional/atom pairs in Experiment 8.

Functional	Rung	Atoms
PBE	GGA	Fe and Ru^+
BP86	GGA	Ru^+
PBE0	hybrid GGA	Ru^+
B97M-V	meta-GGA	Fe
M06-2X	hybrid meta-GGA	Fe

All jobs use def2-TZVP (with `ECP = def2-ECP` for Ru). Each functional/state pair is run twice, without and with `INTERNAL_STABILITY TRUE`. That is 12 calculations per run, **24 jobs** in all. They are small atoms, so each job is quick, although the stability runs take longer.

12.4.4 Pitfalls

- **Same settings in both runs.** The only difference between the two inputs must be `INTERNAL_STABILITY`. Use the website's inputs.
- **Read the stability section of each output.** Look for whether an instability was found and whether the wave function was reoptimized. Then compare the final energy with the one from the first run.
- **Check $\langle S^2 \rangle$** in every output against Table 24. Note any large deviation in your answers.
- **Charge and multiplicity** are on line 2 of every geometry file. Ru^+ has charge +1.
- **SCF failures.** If a job doesn't converge, record it and tell your instructor. Don't silently replace the functional.
- **Units.** Total energies in E_h with all their digits; excitation energies in kcal/mol (× 627.5095).

12.4.5 What to look at

The **stability landscape** widget on the experiment page lets you play with the toy model of Figure 24 before you run anything. After you enter your results, the **energy-lowering table** shows E(with) − E(without) for every functional and state, in kcal/mol: zero means nothing changed, a negative number means a lower solution was found. The **error chart** shows each excitation-energy error without (grey) and with (red) stability analysis.

Two questions to think about while you work, without guessing the answers here: if stability analysis lowers only one of the two states, in which direction does the excitation energy move? And once every solution is stable, is the remaining error small?

12.4.6 Key points

- Grid wiggles are numerical noise. Test any feature of a curve with a finer grid. Meta-GGAs, especially heavily parametrized ones, need SG-2 to SG-3; report the grid.
- Grid convergence is not accuracy: B3LYP gives a smooth but essentially unbound Ar_2 curve because it has no dispersion.
- A converged SCF solution has zero gradient with respect to orbital rotations. It can be a minimum or a saddle point.
- Stability analysis computes the lowest eigenvalues of the electronic Hessian (second derivatives of the energy with respect to orbital rotations). A negative eigenvalue means a lower solution exists; the program follows it and reconverges.

- Internal instabilities lead to lower solutions of the same type; external ones require relaxing a constraint (RHF ⟶ UHF, real ⟶ complex).
- Unrestricted determinants are spin-contaminated: $\langle S^2 \rangle$ > S(S+1). Use it as a diagnostic.
- Transition metals have many near-degenerate configurations and spin states, so instabilities and convergence failures are common.
- Elements beyond Kr use effective core potentials with def2 basis sets (`ECP = def2-ECP`); the ECP includes scalar relativistic effects of the core.

12.4.7 Study questions

1. A classmate's M06-L curve on SG-2 has three minima, at 3.6, 4.1 and 4.6 Å. List two checks that tell you whether any of them is physical.
2. Why can B3LYP give a smooth curve with no minimum for Ar_2, while a less accurate functional gives a well of about the right depth?
3. Explain the difference between the gradient condition $\mathbf{g} = 0$ and the stability condition on the electronic Hessian. Which one does an ordinary SCF check?
4. In the toy model $E(\theta) = \Delta \sin^2\theta - \sin^2 2\theta$, find the lower solutions and their energy for $\Delta = 3$.
5. Why does a small HOMO–LUMO gap make an instability more likely? Point to the term in $A_{ia,jb}$ responsible.
6. Give one example each of an internal and an external instability.
7. A UKS calculation on a doublet radical gives $\langle S^2 \rangle$ = 0.78. Another gives 1.10. Comment on each.
8. How many electrons are treated explicitly for neutral Pd with def2-TZVP? What is the charge of the effective core?
9. If stability analysis lowers the ground state of an atom by 5 kcal/mol and leaves the excited state unchanged, how does the computed excitation energy change?
10. Why do functionals with a large fraction of HF exchange tend to favor high-spin states?

12.4.8 Further reading

- R. Seeger and J. A. Pople, "Self-consistent molecular orbital methods. XVIII. Constraints and stability in Hartree–Fock theory" [138]: the classic paper on the hierarchy of instabilities.
- J. Čížek and J. Paldus [137]: the original derivation of the stability conditions.
- R. Bauernschmitt and R. Ahlrichs, stability analysis for HF and DFT wave functions [140].
- H. Fukutome, a group-theoretical classification of unrestricted solutions [139].
- D. Andrae et al., the Stuttgart effective core potentials for the 4d and 5d metals used with def2 basis sets [142].
- N. Mardirossian and M. Head-Gordon on grid sensitivity [133].
- P. Morgante and R. Peverati, "The devil in the details" [32], sections on grids and stability.

13 From electronic energies to experiment

i Learning goals

After this lecture you should be able to:

- explain what stability analysis changed in Experiment 8, and why a stable solution can still be inaccurate;
- distinguish ΔE_{el}, ΔE_0, ΔH_{298} and ΔG_{298}, and say which one a Q-Chem single-point energy gives you;
- compute a zero-point energy and the thermal corrections of an ideal gas from harmonic frequencies;
- explain the difference between D_e and D_0;
- compute a spin–orbit correction for an atom from its energy levels;
- explain why an atomization energy can't be compared directly with an enthalpy of formation, and connect the two with NIST data;
- plan Experiment 9.

i Note

This lecture may take two meetings. If it does, the natural split point is after Section 13.2.3: the first meeting covers the Experiment 8 debrief and the thermal corrections, the second covers spin–orbit coupling, atomization energies, experimental data and the Experiment 9 briefing.

13.1 Part 1 · Looking back at Experiment 8

! Read this after you submit Experiment 8

This part discusses the trends you most likely saw in Experiment 8. Your own numbers may differ. Focus on the reasons.

13.1.1 How much did stability analysis change?

Start with the **energy-lowering table**: E(with stability analysis) − E(without), for each functional and state. The typical picture is lopsided. For most functional/state pairs the number is 0.00: the first SCF solution was already stable. For one or two pairs it is large and negative, often tens of kcal/mol. In those cases the first SCF had landed on an unstable solution, typically a wrong occupation of the d orbitals, and the stability analysis found a much lower one. Hybrid functionals are often more prone to this: exact exchange makes the energy more sensitive to how the d electrons are arranged, which creates more competing solutions.

What happens to the excitation energy depends on *which* state was unstable. If only the ground state is lowered, the gap grows. If only the excited state is lowered, the gap shrinks. If both are lowered by different amounts, it can go either way. That is why you must check every species of a reaction, not just the one that "looks" difficult.

⚠ Common mistake: an energy that went up

If any number in your table was positive by more than a few hundredths of a kcal/mol, the two runs did not use the same settings (a different basis, a missing ECP, a different multiplicity). A stability analysis can't raise the energy.

13.1.2 Stable is not the same as accurate

Now look at the error chart. After stability analysis, every solution is a true minimum, but the errors that remain are probably still large, particularly for Fe, where errors of 10 kcal/mol or more are common. Stability analysis fixes a *numerical* problem: it makes sure you found the lowest solution of the equations you set up. It can't fix the *functional's* error.

For Fe, the excitation changes the configuration from $3d^6 4s^2$ to $3d^7 4s^1$: one electron moves from 4s to 3d. Approximate functionals are known to be biased in how they treat a localized d electron compared with a diffuse s electron, mostly through self-interaction error (Lecture 9). Many semilocal functionals make the $3d^7 4s^1$ states too low, so they underestimate this excitation. Ru^+ is a different case, and you probably saw smaller errors there for some functionals. One atom and one functional rarely tell the whole story.

13.1.3 Why several functionals were removed

SPW92, B3LYP, SCAN, ωB97M-V and Hartree–Fock were taken out of Experiment 8 because their SCF calculations on these atoms often did not converge on the IQmol server. Near-degenerate d orbitals make the SCF oscillate between occupations. Better initial guesses, damping, level shifting or direct-minimization algorithms (Q-Chem's `SCF_ALGORITHM` options) often help, but they take experience. The lesson: for transition metals, expect convergence problems, always run a stability analysis, and get help from someone who has done it before [141].

13.1.4 Spin-state energetics are a hard benchmark

The reference values you used came from atomic spectroscopy. They are among the most accurate numbers in this course, and still most functionals miss them by many kcal/mol. When a paper reports a spin-state gap or a metal–ligand bond energy computed with one functional, ask which functional, why, and whether anyone checked it against data for similar systems. That question returns in Lecture 14.

The reference values also raise a question we have avoided so far. An experiment measures an atom's energy levels, a molecule's heat of reaction, or a bond enthalpy at room temperature. Our calculations give an electronic energy at fixed nuclei. How are the two related? That is the subject of the rest of this lecture.

13.2 Part 2 · From electronic energies to experiment

13.2.1 What a single-point energy is

The "total energy" in a Q-Chem output is the electronic energy at fixed nuclei plus the nuclear repulsion (Lecture 1):

$$E_{\mathrm{el}}(\mathbf{R}) = \langle \Psi_{\mathrm{el}} | \hat{H}_{\mathrm{el}} | \Psi_{\mathrm{el}} \rangle + V_{\mathrm{nn}}(\mathbf{R}).$$

It is the potential energy surface at one point $\mathbf{R}$: no nuclear motion, no temperature. Every reaction energy you have computed in this course is

$$\Delta E_{\mathrm{el}} = \sum_i \nu_i \, E_{\mathrm{el},i},$$

with signed stoichiometric coefficients. The benchmark references in Experiments 1–8 are values of this same quantity, derived from high-level calculations or from experiment after removing all the effects below. That is why you could compare directly.

An experiment never measures ΔE_{el}. Real molecules vibrate even at 0 K, and at room temperature they also translate, rotate and populate excited vibrational levels. Experiments report one of these:

Table 27: The energies a "reaction energy" can mean.

Quantity	What it includes	Typical source
ΔE_{el}	electronic energy only, nuclei fixed at the minima	calculations; benchmark databases
ΔE_0 (or D_0)	ΔE_{el} + zero-point vibrational energy	spectroscopy (dissociation limits)
ΔH_T	ΔE_0 + thermal energy at temperature T + $\Delta(pV)$	calorimetry, thermochemical tables
ΔG_T	$\Delta H_T - T\Delta S_T$	equilibrium constants, electrochemistry

13.2.2 Zero-point energy and harmonic frequencies

Near a minimum, the potential energy surface is approximately quadratic in the nuclear displacements. Diagonalizing the mass-weighted matrix of second derivatives of E_{el} with respect to the nuclear coordinates (the *nuclear* Hessian, not the electronic one of Lecture 12) gives $3N_{\mathrm{atoms}} - 6$ normal modes for a nonlinear molecule ($3N_{\mathrm{atoms}} - 5$ for a linear one), each with a harmonic wavenumber $\tilde{\nu}_k$. In Q-Chem this is `JOBTYPE FREQ`. It is only meaningful at a stationary point optimized with the same method and basis set. At a minimum all frequencies are real. An imaginary frequency marks a saddle point (or, if it is small, possibly grid noise; see Lecture 11).

A quantum harmonic oscillator can't be at rest: its lowest level has energy $\frac{1}{2}hc\tilde{\nu}$. The **zero-point vibrational energy** of a molecule is the sum over modes,

$$\mathrm{ZPE} = \frac{1}{2}\sum_k hc\,\tilde{\nu}_k.$$

Atoms have no vibrations and no ZPE. A molecule always has some, so breaking bonds always *releases* zero-point energy, and

$$\Delta E_0 = \Delta E_{\mathrm{el}} + \sum_i \nu_i\,\mathrm{ZPE}_i.$$

For a diatomic, the **equilibrium dissociation energy** D_e is measured from the bottom of the well, and the **spectroscopic dissociation energy** D_0 from the lowest vibrational level:

$$D_0 = D_e - \mathrm{ZPE}.$$

Harmonic frequencies from DFT are usually a few percent too high, partly because of anharmonicity and partly because of the functional, so they are often multiplied by an empirical scale factor. Anharmonicity lowers the real ZPE a little below the harmonic value.

Worked example: D_e and D_0 of H_2

Spectroscopy gives, for H_2, the harmonic wavenumber $\omega_e = 4401.2$ cm^{-1} and the anharmonicity constant $\omega_e x_e = 121.3$ cm^{-1}. The zero-point level of an anharmonic diatomic is $\frac{1}{2}\omega_e - \frac{1}{4}\omega_e x_e$:

$$\mathrm{ZPE} = 2200.6 - 30.3 = 2170.3\ \mathrm{cm}^{-1}.$$

With 1 cm^{-1} = 0.0028591 kcal/mol, ZPE = 6.21 kcal/mol. The measured D_0 is 36118 cm^{-1} = 103.27 kcal/mol, so

$$D_e = D_0 + \mathrm{ZPE} = 103.27 + 6.21 = 109.48\ \mathrm{kcal/mol},$$

which matches the W4-11 reference you will use in Experiment 9 (109.5 kcal/mol). As a check from first principles: the exact nonrelativistic energy of H_2 at its equilibrium bond length is −1.17448 E_h, and each H atom has exactly −0.5 E_h. Then $D_e = 2(-0.5) - (-1.17448) = 0.17448\ E_h$, and $0.17448 \times 627.5095 = 109.49$ kcal/mol.

13.2.3 Thermal corrections from statistical mechanics

At temperature T molecules also carry thermal energy. For an ideal gas, statistical mechanics splits the molecular partition function into independent factors,

$$q = q_{\mathrm{trans}}\, q_{\mathrm{rot}}\, q_{\mathrm{vib}}\, q_{\mathrm{elec}},$$

and each factor gives a contribution to the internal energy U and the entropy S. With the rigid-rotor, harmonic-oscillator model, the molar thermal energies above the zero-point level are:

$$\begin{aligned}
E_{\mathrm{trans}} &= \frac{3}{2}RT, \\
E_{\mathrm{rot}} &= RT\ (\text{linear}), \qquad \frac{3}{2}RT\ (\text{nonlinear}), \\
E_{\mathrm{vib}} &= R\sum_k \frac{\theta_k}{e^{\theta_k/T}-1}, \qquad \theta_k = \frac{hc\tilde{\nu}_k}{k_{\mathrm{B}}}, \\
E_{\mathrm{elec}} &\approx 0\ (\text{unless low-lying excited states exist})\ .
\end{aligned}$$

For one mole of ideal gas $H = U + pV = U + RT$, so the enthalpy at temperature T is

$$H_T = E_{\mathrm{el}} + \mathrm{ZPE} + E_{\mathrm{trans}} + E_{\mathrm{rot}} + E_{\mathrm{vib}} + RT.$$

At 298.15 K, $RT = 0.5925$ kcal/mol. An atom has only translation: $H_{298} - H_0 = \frac{5}{2}RT = 1.48$ kcal/mol. A diatomic adds rotation, $\frac{7}{2}RT = 2.07$ kcal/mol, plus a vibrational term that is negligible when $\tilde{\nu}$ is large. Figure 25 shows why: a stiff mode is "frozen" at room temperature, and only modes below a few hundred cm^{-1} contribute much.

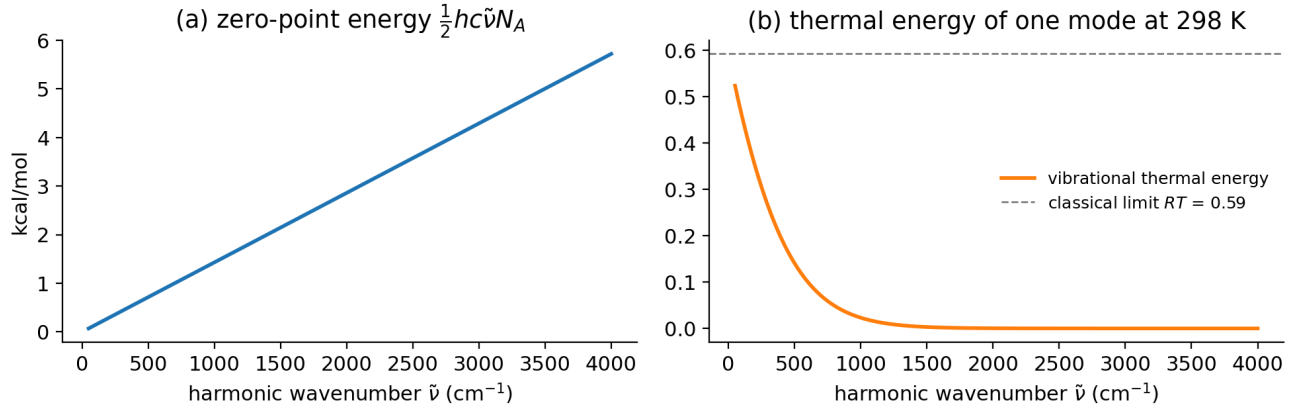


Figure 25: (a) Zero-point energy and (b) thermal vibrational energy at 298 K of a single harmonic mode, as functions of its wavenumber. The ZPE grows linearly; the thermal energy is close to the classical value RT only for very soft modes and nearly zero above about 1500 cm^{-1}. Exact harmonic-oscillator formulas.

Worked example: ΔH_{298} for $H_2 \longrightarrow 2\,H$

Products: two H atoms, each $\frac{5}{2}RT$. Reactant: H_2, $\frac{7}{2}RT$ (its vibration, 4401 cm^{-1}, contributes nothing at 298 K). The thermal part of the reaction enthalpy is

$$2 \times \frac{5}{2}RT - \frac{7}{2}RT = \frac{3}{2}RT = 1.5 \times 0.5925 = 0.89\ \text{kcal/mol}\ .$$

So $\Delta H_{298} = D_0 + 0.89 = 103.27 + 0.89 = 104.16$ kcal/mol. The NIST WebBook gives $\Delta_f H^\circ_{298}(\text{H}, g) = 217.998$ kJ/mol = 52.10 kcal/mol, and $\Delta_f H^\circ(\text{H}_2) = 0$, so the experimental $\Delta H_{298} = 2 \times 52.10 = 104.20$ kcal/mol. The small difference comes from treating H_2's rotation classically.

Figure 26 puts the three numbers side by side. For H_2 the ZPE is the biggest correction, about 6 kcal/mol, and the thermal correction is under 1 kcal/mol. For molecules with many hydrogen atoms the ZPE becomes large: tens of kcal/mol for the atomization of methane.

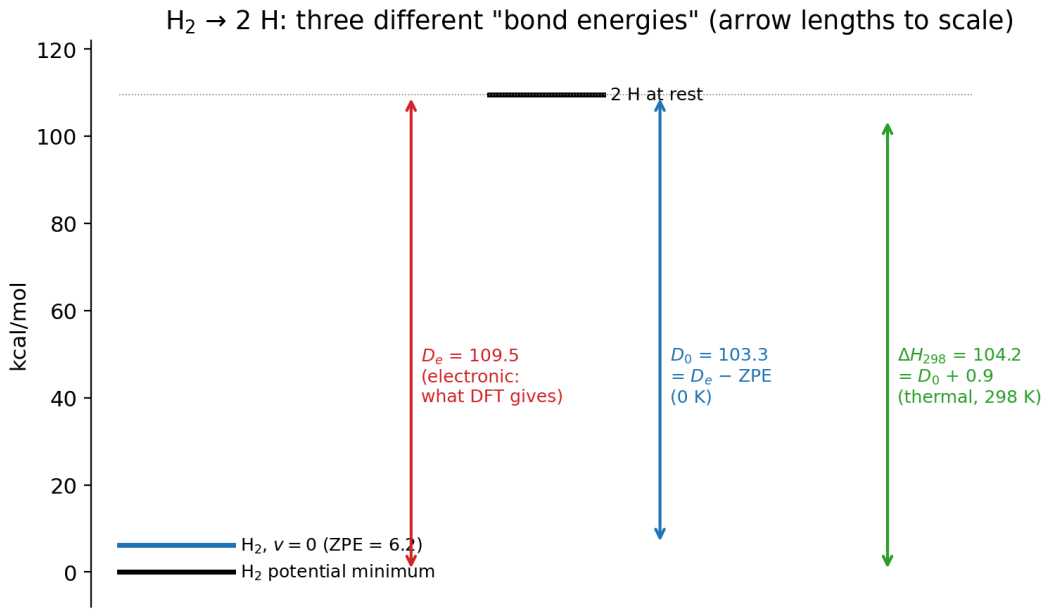


Figure 26: Three "bond energies" of H_2, to scale: the electronic D_e (what a single-point calculation gives), the spectroscopic D_0 (after removing the zero-point energy) and the reaction enthalpy at 298 K. Literature values; schematic layout.

13.2.4 Entropy and Gibbs energy

The same partition functions give the entropy. Translation dominates, and it depends on the number of particles: breaking one molecule into two fragments gains a lot of translational entropy, about 30 cal mol^{-1} K^{-1} or more. At 298 K, $T\Delta S$ for a dissociation is typically 8–12 kcal/mol, so ΔG is much smaller than ΔH. Vibrational entropy comes mostly from soft modes, and the harmonic model treats very soft modes (internal rotations, floppy ligands) poorly. That makes computed entropies and Gibbs energies more uncertain than enthalpies. In this course we compare energies and enthalpies only.

> ⚠ Common mistakes with thermochemistry
>
> - **Frequencies at a non-optimized geometry.** The harmonic analysis assumes a stationary point of the same method. The geometries in our experiments come from benchmark databases, not from your functional, so we don't compute frequencies on them.
> - **Mixing levels.** ZPE and thermal corrections are often computed with a cheaper method than the electronic energy. That's fine, but say so.
> - **Comparing ΔE_{el} with ΔH_{298}.** For bond breaking the difference is several kcal/mol, larger than the chemical accuracy you are aiming for.

13.2.5 Spin–orbit coupling in atoms

Standard DFT, as used in this course, is nonrelativistic for the light atoms and scalar-relativistic for the heavy ones (Lecture 12). It ignores **spin–orbit coupling**, the interaction of the electron spin with its orbital motion. In an atom with orbital angular momentum $L > 0$ and spin $S > 0$, spin–orbit coupling splits the term into levels with total angular momentum $J = |L - S|, \ldots, L + S$. A nonrelativistic calculation gives, approximately, the energy of the term *averaged* over all its levels. The real ground level lies below that average.

The **spin–orbit correction** is the difference between the lowest level and the degeneracy-weighted average:

$$\Delta E_{\mathrm{SO}} = E_{J,\mathrm{lowest}} - \frac{\sum_J (2J+1)\,E_J}{\sum_J (2J+1)}.$$

It is always negative (or zero). Atoms in S states (H, N) have $L = 0$ and no first-order correction. Closed-shell molecules have none either. Open-shell atoms such as C, O, Cl and Fe do.

> Worked example: the spin–orbit correction of the oxygen atom
>
> The ground term of O is ^{3}P, with levels $J = 2$ at 0 cm^{-1}, $J = 1$ at 158.3 cm^{-1} and $J = 0$ at 227.0 cm^{-1}. The weighted average is
>
> $$\bar{E} = \frac{5(0) + 3(158.3) + 1(227.0)}{5+3+1} = \frac{701.9}{9} = 78.0\ \mathrm{cm}^{-1}.$$
>
> So $\Delta E_{\mathrm{SO}} = 0 - 78.0 = -78.0$ cm^{-1} = $-78.0 \times 0.0028591 = -0.22$ kcal/mol. The same calculation for Cl (^{2}P, $J = 3/2$ at 0 and $J = 1/2$ at 882.4 cm^{-1}) gives $-2(882.4)/6 = -294.1$ cm^{-1} = −0.84 kcal/mol. For C it is −0.08 kcal/mol, and for Fe (^{5}D, five levels up to 978 cm^{-1}) about −1.2 kcal/mol.

Because the computed atoms are too high by $|\Delta E_{\mathrm{SO}}|$, a computed atomization energy without spin–orbit coupling is too *large* by the sum of the atomic corrections. For $O_3 \longrightarrow 3$ O, that is 3 × 0.22 = 0.66 kcal/mol. Benchmark references such as W4-11 report values with spin–orbit effects removed [147], so you can compare with them directly. An experimental number needs this correction first. For heavier atoms the correction grows quickly: for iodine it is several kcal/mol.

13.2.6 Atomization energies and enthalpies of formation

The **total atomization energy** (TAE) of a molecule is the energy to separate it into ground-state atoms in the gas phase. For methane, $CH_4 \longrightarrow$ C + 4 H. The **standard enthalpy of formation** $\Delta_f H^\circ$ is the enthalpy of forming the molecule from its *elements in their standard states* at 298.15 K and 1 bar: for methane, C(graphite) + 2 H_2(g) $\longrightarrow$ CH_4(g).

The two are not the same reaction. Graphite and H_2 gas are not atoms. To connect them you need the enthalpies of formation of the gas-phase atoms, which are themselves experimental numbers:

$$\Delta H_{298}(\text{atomization of M}) = \sum_{\text{atoms}} \Delta_f H^\circ(\text{atom}, g) - \Delta_f H^\circ(\mathrm{M}, g).$$

Then, to reach the electronic TAE that a calculation gives, you must remove the thermal enthalpies, add back the ZPE, and remove the spin–orbit corrections.

Worked example: from NIST to the atomization energy of methane

From the NIST WebBook [148], in kJ/mol: $\Delta_f H^\circ(\mathrm{C}, g) = 716.68$, $\Delta_f H^\circ(\mathrm{H}, g) = 217.998$ and $\Delta_f H^\circ(\mathrm{CH_4}, g) = -74.87$. In kcal/mol (÷ 4.184): 171.29, 52.10 and −17.89. Then

$$\Delta H_{298} = 171.29 + 4(52.10) - (-17.89) = 397.6\ \text{kcal/mol}\ .$$

The W4-11 electronic reference used in Experiment 9 is 420.4 kcal/mol. The 22.8 kcal/mol between them is mostly the zero-point energy of methane (about 27.7 kcal/mol), partly offset by the thermal enthalpies (about +5 kcal/mol in favor of the five atoms) and carbon's small spin–orbit correction (0.08 kcal/mol). Comparing a computed ΔE_{el} with 397.6 would make any functional look about 23 kcal/mol wrong.

13.2.7 Experimental uncertainty

Experimental numbers come with error bars, and they matter. The enthalpy of formation of the H atom is known to ±0.006 kJ/mol. Values for small organic molecules are typically good to a few tenths of a kJ/mol. Gas-phase data for transition-metal compounds can be uncertain by several kcal/mol, comparable to the errors of a good functional. When you combine several values, uncertainties add in quadrature for independent errors:

$$\sigma_{\Delta H} = \sqrt{\sum_i \nu_i^2\, \sigma_i^2}.$$

If the experimental uncertainty is ±3 kcal/mol, a functional that agrees to within 1 kcal/mol is not "better" than one that is off by 2. Both are within the error bar. Before declaring a winner, check the experiment's uncertainty.

13.2.8 The NIST Chemistry WebBook

The NIST Chemistry WebBook [148] is the standard free source of gas-phase thermochemistry. To use it:

1. Search by **formula** (e.g. `FeCl`) or name, and choose the right species (check the structure and the phase).
2. Open **Gas phase thermochemistry data**. Read $\Delta_f H^\circ_{\mathrm{gas}}$, its **units** (usually kJ/mol) and its **uncertainty**.
3. Note the **reference** for the value. Several measurements may be listed; prefer a recent evaluated compilation and say which you used.
4. For atoms, search the element and take the gas-phase value (for example H, C, O, Cl, Fe).
5. Convert with 1 kcal/mol = 4.184 kJ/mol, and use signed stoichiometric coefficients.

13.3 Experiment 9 briefing

13.3.1 Goal

Experiment 9 tests five methods on two bond dissociations and five atomization energies, all reactions that break molecules into radicals or atoms, and asks you to compare one of them with experimental data from NIST.

13.3.2 The reactions

Table 28: Reactions and reference values of Experiment 9. The bond dissociations come from the Minnesota 2015B database [31]; the atomization energies from W4-11 [147] in GMTKN55 [43].

Label	Reaction	Reference ΔE_{el} (kcal/mol)	Source
9a	$(CH_3)_3C{-}OH \longrightarrow (CH_3)_3C\cdot + \cdot OH$	115.02	SR-MGN-BE107
9b	$FeCl \longrightarrow Fe + Cl$	78.5	SR-TM-BE17
	$H_2 \longrightarrow 2\,H$	109.5	W4-11
	$CH_4 \longrightarrow C + 4\,H$	420.4	W4-11
	$NH_3 \longrightarrow N + 3\,H$	298.0	W4-11
	$H_2O \longrightarrow O + 2\,H$	233.0	W4-11
	$O_3 \longrightarrow 3\,O$	147.4	W4-11

All reference values are electronic energies (D_e-like), so you compare your ΔE_{el} with them directly. The exception is the NIST exercise for FeCl, where you do the opposite: start from experiment and see how far it is from the electronic reference.

Multiplicities to get right: H, OH, *tert*-butyl and Cl are doublets; C and O are triplets; N is a quartet; Fe is a quintet; FeCl is a sextet; all closed-shell molecules are singlets. The geometry files have them on line 2.

13.3.3 Methods and settings

Setting	Value
Methods	MN15-L, MN15, B3LYP, PBE, HF
Basis set	def2-TZVP (def2-QZVP would not fit the 10-minute limit for *tert*-butanol)
Grid	(99,590): `XC_GRID 000099000590`
Stability	`INTERNAL_STABILITY TRUE` on every job
Extra	B3LYP/6-31G* for the three species of reaction 9a only
Jobs	15 species × 5 methods + 3 = **78**

The website's inputs already contain the grid and the stability keyword. Everything you learned in Lectures 11 and 12 is switched on.

Worked example: signs in an atomization reaction

For $CH_4 \longrightarrow C + 4\,H$, the coefficients are $\nu(CH_4) = -1$, $\nu(C) = +1$, $\nu(H) = +4$:

$$\Delta E = E(\mathrm{C}) + 4\,E(\mathrm{H}) - E(\mathrm{CH_4}).$$

With total energies in E_h, multiply the result by 627.5095. Forgetting the 4 in front of $E(\mathrm{H})$ is the most common error in this experiment. It gives a "reaction energy" off by three hydrogen atoms, about 3×0.5 $E_h \approx 940$ kcal/mol. A number that far from the reference should stop you immediately.

13.3.4 Pitfalls

- **Big numbers, small relative errors.** An error of 4 kcal/mol is 1% of methane's atomization energy but 3.6% of H_2's. The **relative errors** chart puts them on the same footing; look at both.
- **The H atom is used many times.** An error in $E(\mathrm{H})$ is multiplied by 4 in $CH_4 \longrightarrow C + 4\,H$. Check that every H-atom job used the right method.
- **Radicals and atoms.** Every product in this experiment is open-shell. Check $\langle S^2 \rangle$ in each output (Lecture 12).
- **Time.** *tert*-Butanol with def2-TZVP, (99,590) and stability analysis is the largest job. If MN15 or MN15-L is slow, start those first.

13.3.5 The NIST exercise

The **Theory meets experiment** panel asks for $\Delta_f H^\circ$ of FeCl(g), Fe(g) and Cl(g) from the NIST WebBook. It computes the experimental bond enthalpy at 298 K and compares it with the electronic reference of 78.5 kcal/mol. Your job is to explain the difference with the corrections of this lecture: ZPE (small for a heavy diatomic with a low frequency), thermal enthalpy, spin–orbit coupling in the Fe and Cl atoms and in FeCl itself, and the experimental uncertainty. Estimate each one if you can.

Before you run anything, think back to Lecture 9 and make your predictions: which methods include electron correlation, which reactions have no error cancellation between reactants and products, and which basis-set behavior from Experiment 6 might return for *tert*-butanol?

13.3.6 Key points

- Stability analysis changes nothing for most species and a lot for a few. Check every species; a stable solution can still have a large functional error.
- Transition-metal SCF calculations often fail to converge. Expect it, and use stability analysis.
- A single-point calculation gives ΔE_{el}. Experiments give D_0, ΔH_T or ΔG_T.
- $\Delta E_0 = \Delta E_{\mathrm{el}} + \Delta\mathrm{ZPE}$, with ZPE = ½ Σ $hc\tilde{\nu}_k$ from harmonic frequencies at an optimized geometry. $D_0 = D_e - \mathrm{ZPE}$.
- Thermal enthalpy at 298 K: $\frac{3}{2}RT$ translation, RT or $\frac{3}{2}RT$ rotation, a vibrational term from soft modes, and RT for pV.
- Open-shell atoms with $L > 0$ need a spin–orbit correction (O: −0.22, Cl: −0.84 kcal/mol). Without it, computed atomization energies are too large.
- An atomization energy is not an enthalpy of formation. Connect them through the enthalpies of formation of the gas-phase atoms.
- Always check the experimental uncertainty before ranking methods.

13.3.7 Study questions

1. In Experiment 8, the stability analysis lowered the excited state of one atom by 40 kcal/mol and its ground state by 25 kcal/mol. By how much and in which direction did the excitation energy change?
2. Give two reasons why a stable SCF solution can still give a large error for an atomic excitation energy.
3. Order ΔE_{el}, ΔE_0 and ΔH_{298} for the dissociation of $\mathrm{H_2}$, from largest to smallest, and explain the order.
4. The harmonic wavenumber of HCl is about 2990 $\mathrm{cm^{-1}}$. Estimate its ZPE in kcal/mol, neglecting anharmonicity.
5. Show that the thermal part of ΔH_{298} for a diatomic AB $\longrightarrow$ A + B is about $\tfrac{3}{2}RT$ when the vibrational frequency is high. What changes if the frequency is 300 $\mathrm{cm^{-1}}$?
6. Why do atoms in S states (H, N) need no spin–orbit correction? Why does the correction make computed atomization energies too large?
7. The F atom has a ^{2}P term with levels at 0 ($J = 3/2$) and 404.1 $\mathrm{cm^{-1}}$ ($J = 1/2$). Compute its spin–orbit correction in kcal/mol.
8. Explain why you can't compare a computed atomization energy of water with $\Delta_f H^\circ(\mathrm{H_2O}, g)$. List the data you need to connect them.
9. A functional gives a bond energy within 1.5 kcal/mol of an experimental value with an uncertainty of ±3 kcal/mol. Another gives 2.5 kcal/mol. Can you say one is better?
10. Why is ΔG for a dissociation reaction much smaller than ΔH at room temperature?

13.3.8 Further reading

- C. J. Cramer and D. G. Truhlar, "Density functional theory for transition metals and transition metal chemistry" [149]: spin states, multireference character and practical advice.
- W. Zhang, D. G. Truhlar and M. Tang, and X. Xu et al., on comparing computed and experimental bond energies of transition-metal compounds [150], [151].
- A. Karton, S. Daon and J. M. L. Martin, the W4-11 dataset [147]: how electronic reference values are built from experiment and high-level theory.
- L. A. Curtiss et al. on the G2 test set [152]: a standard source for spin–orbit and thermal corrections in thermochemistry.
- W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory* [8], the chapters on thermochemistry and harmonic frequencies.
- The NIST Chemistry WebBook [148].
- For a review of statistical thermodynamics, see the P-Chem 2 textbook.

14 Choosing, running and reporting a method

> **i** Learning goals
>
> After this lecture you should be able to:
>
> - explain why reactions that produce atoms and radicals are harder than reactions between closed-shell molecules, and why Hartree–Fock fails for them;
> - explain the gap between an experimental bond enthalpy and an electronic reference value;
> - choose a functional, basis set, dispersion treatment, grid and SCF settings for a new problem, and justify each choice with evidence;
> - check whether a calculation fits a computer-time budget;
> - write a methods paragraph that lets someone else reproduce your numbers;
> - plan Experiment 10.

14.1 Part 1 · Looking back at Experiment 9

> **!** Read this after you submit Experiment 9
>
> This part discusses the trends you most likely saw in Experiment 9. Your own numbers may differ. The reasons are what you should take away.

14.1.1 Hartree–Fock and the atomization energies

The first thing on the signed-error chart was probably Hartree–Fock. Its atomization energies are far too small: tens of kcal/mol too small for the hydrides, and for ozone the error can be so large that HF barely binds the molecule at all relative to three oxygen atoms. You saw the same failure for H_2SO_4 in Experiment 5.

The reason is the one from Lecture 9. Correlation energy is larger in the molecule than in the separated atoms, because bonding packs more electron pairs into the same region of space. Hartree–Fock has no correlation at all, so it misses this extra stabilization of the molecule, and every bond comes out too weak. In an atomization nothing on the product side looks like the reactant, so there is no cancellation to rescue it. Even H_2, a two-electron problem, is underbound by HF: the missing energy is exactly the correlation energy of the H_2 bond, since the hydrogen atom has only one electron and no correlation.

14.1.2 The density functionals

The density functionals all include correlation in some approximate way, and most of them got the hydrides within a few kcal/mol. Look at the errors in two ways:

- **Absolute errors** grow with the size of the number. Methane's atomization energy is 420 kcal/mol. An error of 4 kcal/mol there is less than 1%.
- **Relative errors** put H_2 and CH_4 on the same scale. The relative-errors chart probably showed that some functionals that look fine on the absolute scale are not as good per bond.

Ozone was probably the hardest atomization for everyone. It has strong multireference character (Lecture 9), and a single determinant can't describe it well. GGAs such as PBE tend to overbind ozone by a lot, inheriting LSDA's overbinding. Hybrids and the Minnesota functionals usually do better, but not uniformly.

14.1.3 Radicals and the C–O bond of *tert*-butanol

Reaction 9a breaks one C–O bond and produces two radicals, *tert*-butyl and OH. The reactant is a closed-shell molecule; the products are two open-shell doublets. As with atomization, the two sides are different kinds of species, so error cancellation is weak. Self-interaction error tends to overstabilize the delocalized, partly broken bond relative to the separated radicals, and the open-shell products are where unrestricted calculations, spin contamination and stability all matter.

If you compared B3LYP/6-31G* with B3LYP/def2-TZVP for this reaction, you may have seen the small basis do slightly *better*. You know the lesson from Experiment 6: before you call it an improvement, ask whether two errors cancelled.

14.1.4 FeCl

The FeCl dissociation combines everything difficult: a transition metal, a sextet molecule, a quintet Fe atom and a doublet Cl atom. The spread between functionals is wide, and Hartree–Fock is far off. The functional that does best on FeCl is not necessarily the one that does best on the atomizations. Experiment 9 contains seven reactions, and no single method is best at all of them.

14.1.5 What NIST told you

The **Theory meets experiment** panel probably gave an experimental bond enthalpy a few kcal/mol away from the electronic reference of 78.5 kcal/mol. Several corrections stand between the two numbers (Lecture 13):

- **ZPE.** FeCl is a heavy diatomic with a low vibrational frequency, a few hundred cm^{-1}, so its ZPE is well under 1 kcal/mol.
- **Thermal enthalpy.** About $\frac{3}{2}RT \approx 0.9$ kcal/mol in favor of the atoms, a little less once FeCl's soft vibration is counted.
- **Spin–orbit coupling.** The Fe atom (−1.2 kcal/mol) and the Cl atom (−0.84 kcal/mol) both have it, and so does the FeCl molecule. Together these can amount to a couple of kcal/mol.
- **Experimental uncertainty.** Check the error bar NIST quotes for FeCl. Gas-phase data for transition-metal diatomics are often uncertain by several kcal/mol, comparable to all the corrections combined.

The take-home message: for a transition-metal diatomic, the experimental uncertainty alone may be larger than the difference between a good functional and a mediocre one. Comparing a computed electronic energy directly with an experimental enthalpy is not meaningful without the corrections, and sometimes not meaningful even with them.

14.2 Part 2 · Choosing, running and reporting a method

14.2.1 The question you will face after this course

So far, every experiment told you which methods to run. In research, nobody does. You will have a molecule, a property and a budget, and you will have to decide. There is no single correct answer, but there are bad answers, and there is a process that avoids most of them. The process is the same whether the system is a drug molecule, a catalyst or a battery material.

14.2.2 A decision process

1. Define the question. What property do you need: a reaction energy, a barrier, a geometry, a spin-state gap, a noncovalent binding energy? How accurate must it be? A trend across a series of similar molecules tolerates larger errors than an absolute number to be compared with experiment.

2. Identify the chemistry. Ask the questions this course has trained you to ask:

- Is everything closed-shell, or are there radicals, atoms or open-shell metals? (Experiments 5, 8, 9)
- Do bonds form or break, or is it an isomerization or conformational change with good error cancellation? (Experiments 3–5)
- Are noncovalent interactions or large, polarizable groups involved? (Experiment 4)
- Are there transition metals, near-degenerate states or elements beyond krypton? (Experiments 8–10)
- Is there any sign of multireference character, such as a small HOMO–LUMO gap, ozone-like bonding or stretched bonds? (Experiment 5)

3. Look for benchmark evidence for that kind of reaction. Large benchmark databases rank functionals by reaction type: GMTKN55 for main-group thermochemistry, kinetics and noncovalent interactions [43], MGCDB84 [63], and the Minnesota databases [31], which include transition-metal bond energies. Specialized sets exist for transition-metal chemistry [153], [154]. Read the summary tables, but also the category closest to your problem: a functional that is excellent on average can be poor for your reaction type. Be wary of evidence from a functional's own training set (the CaO lesson of Experiment 5).

4. Choose the basis set. For DFT, a triple-ζ basis such as def2-TZVP is the practical standard for energies (Experiment 1). Double-ζ sets are for screening and geometries, and their results can be right for the wrong reasons (Experiment 6). Use a quadruple-ζ set to check convergence if you can afford it. Elements beyond krypton need an effective core potential: with def2 sets, `ECP = def2-ECP` (Lecture 12). Anions and very weak interactions need diffuse functions.

5. Decide on dispersion. Unless the functional already includes it, add a dispersion correction such as -D4 [58], [59]. Never add -D4 on top of a functional that has VV10 (the "-V" functionals) or that was parametrized to include dispersion. That double counts (Experiment 4). Check that -D4 parameters exist for your functional.

6. Choose the grid. SG-2 at least for meta-GGAs, SG-3 or (99,590) for Minnesota and other heavily parametrized functionals (Experiment 7). Always state it.

7. Check the SCF. Run a stability analysis for anything open-shell or containing a transition metal, check $\langle S^2 \rangle$, and make sure you are in the intended spin state (Experiment 8). For a closed-shell molecule a stability check is cheap insurance.

8. Check the cost. Count basis functions, remember how cost grows with them (you measured the exponent in Experiment 1), and run the smallest species first. Hybrids cost more per SCF cycle than GGAs and meta-GGAs; range-separated hybrids and VV10 add a little more (Experiment 3). On the IQmol server every job must finish within 10 minutes.

9. Sanity-check the answer. If you can, compare two functionals from different families, or two basis sets. If they disagree by more than your accuracy target, you have learned something important about your problem.

> **i** Aside: one functional, or several?
>
> In a research paper, a single "best" functional is rarely enough to convince a referee. Reporting a second functional from a different family, or a basis-set check, shows how robust the conclusion is. In Experiment 10 you report one main result, but a sentence on how it might change with another choice makes the report stronger.

14.2.3 What the course has shown you

Table 29: The evidence from Experiments 1–9.

Experiment	What you saw	What it means for your choice
1	Errors fall from double- to quadruple-ζ; cost rises steeply	Use triple-ζ for energies; check convergence when possible
2	Functionals differ in their ingredients	Know the rung and the dispersion treatment of what you use
3	Easy reactions: most methods do well	Error cancellation makes some problems forgiving; don't generalize from them
4	Dispersion matters even within molecules	Include dispersion; avoid double counting
5	Hard reactions: large errors, HF fails, training sets mislead	Find benchmarks for *your* reaction type
6	A small basis can win by luck	Don't trust a result that gets worse with a better basis
7	Meta-GGAs need fine grids	Set and report the grid
8	SCF solutions can be unstable; metals are hard	Run stability analysis; check spin states
9	Atoms and radicals are hard; experiment needs corrections	Compare like with like

14.2.4 A cost budget

Budgeting computer time is part of choosing a method. A rough rule from Experiment 1: the time per SCF cycle grows roughly as N_{bf}^{p}, with an exponent p that is smaller in practice than the formal value of 4 (you measured it). The number of SCF cycles adds a factor of about 10–30, more for difficult open-shell systems, and a stability analysis adds a few more.

Worked example: will the big job fit?

Suppose a small fragment with 200 basis functions took 40 s with your method, and the full complex has 500 basis functions. With $p = 3$, the estimate for the complex is

$$40\ \mathrm{s} \times \left(\frac{500}{200}\right)^{3} = 40 \times 15.6 = 625\ \mathrm{s} \approx 10.4\ \mathrm{min}\ .$$

That is right at the 10-minute limit, before counting a stability analysis or extra SCF cycles. You could move to a cheaper functional (a meta-GGA instead of a hybrid), a smaller grid where it is safe, or a smaller basis, and say in your report what the compromise costs in accuracy. With $p = 2.5$, the estimate is $40 \times 2.5^{2.5} = 40 \times 9.9 = 395$ s, about 6.6 minutes, comfortably inside. The exponent matters; measure it rather than guess.

The method card in Experiment 10 makes a similar estimate for you from the number of basis functions. It is a rough guide, not a guarantee.

14.2.5 Reproducible reporting

A computed number is only useful if someone else can reproduce it. That requires a methods section that lists every choice that changes the number. A reader should be able to rebuild your input file from your text.

A complete computational methods paragraph states:

1. **Program and version**, with its citation (Q-Chem [1], run through IQmol [29]).
2. **Method**: the functional with its original reference(s); for a hybrid or range-separated functional, the name is enough if it is standard.
3. **Dispersion**: none, -D4 (with reference), or built in (VV10, fitted).
4. **Basis set** with its reference, and the **ECP** for heavy elements (which atoms, which ECP).
5. **Integration grid**: by name (SG-3) or as (n_r, n_Ω).
6. **SCF settings**: convergence threshold (`SCF_CONVERGENCE 8` = 10^{-8}), restricted or unrestricted, stability analysis.
7. **Spin states**: the charge and multiplicity of every species.
8. **Geometries**: where they come from (optimized at which level, or taken from a database, with its reference).
9. **What is reported**: electronic energies, or with ZPE or thermal corrections (and how those were obtained); units and conversion factor.
10. **Reference values** for any comparison, with their source.

Example: a methods paragraph for reaction 4a

"All calculations were performed with Q-Chem [1] through the IQmol interface [29]. We used the ωB97M-V functional [90], which includes VV10 nonlocal correlation [60], so no additional dispersion correction was applied. The def2-TZVP basis set [17] was used for all atoms. The exchange–correlation energy was integrated on the SG-3 grid [131]. SCF calculations used a convergence criterion of 10^{-8} and were followed by an internal stability analysis; all species are closed-shell singlets and were computed with restricted orbitals. Geometries were taken from the DARC subset of GMTKN55 [43] without further optimization. Reported values are electronic reaction energies, with no zero-point or thermal corrections, converted with 1 E_h = 627.5095 kcal/mol, and compared with the GMTKN55 reference value of −45.4 kcal/mol."

About 120 words, and someone else could rebuild every input from it.

⚠ Common mistakes in methods sections

- **"Default settings."** Defaults change between programs and versions. Name the grid, the convergence threshold and the ECP.
- **Citing only the program.** Each functional, basis set and dispersion correction has its own reference.
- **Mixing up the quantity.** Say whether you report ΔE_{el}, ΔE_0 or ΔH, and compare with a reference of the same kind.
- **Hiding the compromise.** If the 10-minute limit forced a smaller basis, say so and estimate the consequence. An honest limitation is a strength of a report, not a weakness.
- **Over-precise numbers.** Report reaction energies to 0.1 or 0.01 kcal/mol, not to eight decimals. Keep all digits in the total energies you used to compute them.

14.3 Experiment 10 briefing

14.3.1 You are the consultant

In Experiment 10, a colleague asks you for the energy needed to remove a hexatriene ligand from a palladium(0) complex:

$$\mathrm{Pd(PH_3)_2(C_6H_8)} \rightarrow \mathrm{Pd(PH_3)_2} + \mathrm{C_6H_8}.$$

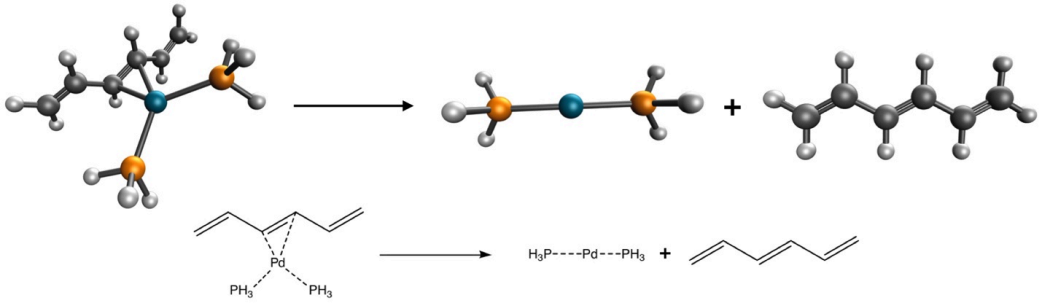


Figure 27: Dissociation of the diphosphinohexatrienepalladium(0) complex. Carbon is black, phosphorus orange, hydrogen white and palladium light blue; dotted lines are dative bonds. From the Experiment 10 page.

The reaction energy, with signed coefficients, is

$$\Delta E = E\left(\ \mathrm{Pd(PH_3)_2}\ \right) + E\left(\ \mathrm{C_6H_8}\ \right) - E\left(\ \mathrm{Pd(PH_3)_2(C_6H_8)}\ \right).$$

The structures come from the PdBE2 subset [155] of the Minnesota 2015B database [31], and the reference value is **16.2 kcal/mol** (an electronic bond dissociation energy, like every reference in this course). All three species are closed-shell singlets: palladium(0) is d^{10}.

14.3.2 What you have to decide

Every setting is yours: functional, dispersion, basis set, grid and stability analysis. The **method card** on the experiment page collects your choices and your reasons, warns you about common pitfalls (double-counting dispersion, grid-sensitive functionals without an explicit grid, no stability check), estimates whether each job fits in 10 minutes, and writes the three Q-Chem inputs. Only three jobs are required. You are welcome to run extra ones to check your choice, for example a second basis set or a second functional.

Use the decision process of this lecture. Some questions to answer before you pick:

- What kind of bond breaks here? Is it covalent, dative, noncovalent, or a mix? What does that say about dispersion (Experiment 4)?
- Which benchmarks include metal–ligand bond energies, and which functionals do well there? Start from the references on the experiment page [30], [31], [43], [63], [153], [154].
- Pd is beyond krypton. Which basis sets come with a suitable ECP?
- Does your functional need a fine grid? Does it already include dispersion?
- Which of your three jobs is the largest, and will it fit?

14.3.3 The report

Write a short report of **about 300 words** that:

1. has a title;
2. states the method you chose and **why**, citing the literature and your own evidence from Experiments 1–9;
3. describes the computational setup completely, as in the checklist above;
4. presents your result and compares it with the reference value of 16.20 kcal/mol, with the signed error;
5. ends with a reference list.

A good report is not the one closest to 16.2. It is the one with a well-justified choice, a reproducible setup, and an honest discussion of the result. A method that lands 2 kcal/mol away for well-stated reasons is better consulting than one that lands on 16.2 by luck and can't explain why. Your instructor may ask you to review two classmates' reports with a short rubric: is the method justified, is the setup reproducible, is the comparison honest?

⚠ Pitfalls in Experiment 10

- **Forgetting the ECP.** Without `ECP = def2-ECP`, a def2 basis on Pd fails or gives meaningless energies. The method card adds it automatically; check your inputs.
- **Double-counting dispersion.** No -D4 on top of a "-V" functional or one that already contains dispersion.
- **Choosing by the reference value.** Choose and justify the method *before* you compute. Changing the functional until the number matches 16.2 is not a method choice, and a reviewer will see it.
- **Different settings for different species.** All three jobs must use exactly the same method, basis, grid and ECP.
- **Running out of time.** Run the smallest species (hexatriene) first to check the settings, then the complex.

14.3.4 Key points

- Hartree–Fock badly underbinds atomization energies because it has no correlation; the missing correlation is larger in the molecule than in the atoms.
- Reactions that produce atoms or radicals have little error cancellation. Report relative as well as absolute errors.
- An experimental bond enthalpy differs from an electronic reference by ZPE, thermal enthalpy and spin–orbit corrections, and for metal compounds the experimental uncertainty can be as large as all of them together.
- Choosing a method means defining the question, identifying the chemistry, finding benchmark evidence for that reaction type, and then choosing basis set, ECP, dispersion, grid, SCF checks and cost to match.

- A methods section must let a reader rebuild your input: program, method, dispersion, basis, ECP, grid, SCF settings, spin states, geometry source, the quantity reported, and references.
- In Experiment 10 you are judged on the justification and the reproducibility of your choice, not only on how close you land to 16.2 kcal/mol.

14.3.5 Study questions

1. Why does Hartree–Fock underbind H_2, even though the H atom has no correlation energy at all?
2. A functional has a mean absolute error of 3 kcal/mol on the atomizations of Experiment 9 and 8 kcal/mol on FeCl. What would you conclude about using it for a new iron compound?
3. List the corrections needed to compare the FeCl bond enthalpy from NIST with the electronic reference, and estimate the size of each.
4. Your reaction involves a Cu(I) complex with a large aromatic ligand. Go through steps 2–7 of the decision process and write one sentence for each.
5. Why is it a mistake to add -D4 to ωB97M-V? Why might it be a mistake to add it to a Minnesota functional?
6. A job with 300 basis functions took 2 minutes. Estimate the time for 600 basis functions with exponents of 2.5 and 3.5. Would either fit in 10 minutes?
7. Rewrite this methods sentence so it is reproducible: "Calculations were done with DFT using Q-Chem with default settings."
8. Why should you choose the method for Experiment 10 before you know how close it gets to the reference?
9. A classmate reports 16.3 kcal/mol for Experiment 10 with B3LYP/def2-SVP. What questions would you ask in a peer review?

14.3.6 Further reading

- L. Goerigk et al., "A look at the density functional theory zoo with the advanced GMTKN55 database" [43]: the most widely used main-group benchmark, with explicit recommendations per rung.
- N. Mardirossian and M. Head-Gordon, "Thirty years of density functional theory in computational chemistry" [63]: a review with practical recommendations.
- H. S. Yu, X. He, S. L. Li and D. G. Truhlar on MN15 and the Minnesota 2015 database [31].
- B. Chan, P. M. W. Gill and M. Kimura on transition-metal benchmarks [153], and S. Dohm et al. on metal–organic reactions [154].
- R. Peverati and D. G. Truhlar, "Quest for a universal density functional" [30].
- P. Morgante and R. Peverati, "The devil in the details" [32]: this course's experiments, and why each detail matters.

15 What we learned

> **i** Learning goals
>
> After this lecture you should be able to:
>
> - look at the spread of methods and results in Experiment 10 and explain what made the good choices good;
> - place every source of error you met in this course in a hierarchy, with typical sizes;
> - explain why a quantum chemistry program can't be used as a black box;
> - name three directions in which electronic-structure methods are developing, and what problem each addresses.

15.1 Part 1 · Looking back at Experiment 10

> **!** Read this after you submit Experiment 10
>
> This part discusses what the class typically chose and found in Experiment 10. If your instructor has released the class results, open the class results page alongside it.

15.1.1 The spread of choices

For the first time in the course, everyone ran something different. The class results usually show a wide spread: several functionals from different rungs, with and without dispersion, two or three basis sets, different grids. The computed bond energies typically spread over more than 10 kcal/mol around the reference of 16.2 kcal/mol. That spread is the most important result of the experiment. With the same three geometries and the same program, the choices alone moved the answer by more than any chemist would accept.

15.1.2 What tended to work

A few patterns are typical.

Dispersion made the biggest difference. The hexatriene ligand is large and polarizable, and it sits close to the two phosphines. Part of the binding is dispersion between the ligand and the rest of the complex, the same physics as the alkane and dimer reactions of Experiment 4. Functionals with no dispersion at all, B3LYP being the usual example, typically *underbind* the complex by many kcal/mol, far outside chemical accuracy. Adding -D4, or choosing a functional with VV10 or with dispersion built into its parametrization, usually closed most of that gap.

Modern functionals with good transition-metal benchmarks did well. Students who looked up metal–ligand bond energies in the benchmark literature (Lecture 14) and chose a dispersion-inclusive hybrid or meta-GGA that performs well there usually landed within a few kcal/mol of the reference, often closer.

Triple-ζ basis sets with the def2 ECP were the sensible choice. def2-TZVP fits the 10-minute limit for this complex with most functionals. def2-SVP is faster, but its basis-set superposition error makes the complex look more stable than it is. Combined with -D4, it can overbind: the double counting of Experiment 6 again. A def2-SVP result that lands close to 16.2 should make you suspicious, not happy.

The details mattered less here, but they still mattered. A palladium(0) d^{10} complex is closed-shell with a reasonable HOMO–LUMO gap, so the stability analysis usually found nothing to fix. That was not a wasted

check: you didn't know it in advance. Grid effects on a 16 kcal/mol reaction energy are usually small, but meta-GGAs on SG-1 can shift the result by a few tenths of a kcal/mol.

15.1.3 What made a good report

The best reports had three things in common. They chose the method *before* computing, for stated reasons tied to benchmarks and to the course's evidence. Their methods paragraph was complete enough to rebuild every input. And they discussed the result honestly: the signed error, the likely sources of the remaining error, and what they would change with more computer time. A report that lands 2 kcal/mol away with a clear justification is better work than one that lands on 16.2 and can't explain why.

> ⚠ Common mistake: tuning to the reference
>
> It is tempting to try several functionals and keep the one closest to 16.2. In real consulting there is no reference value to tune to; that's why you were asked. A method chosen by matching the answer has no predictive power for the next molecule, the one without a reference.

15.2 Part 2 · What we learned

15.2.1 Two approximations, revisited

In Lecture 1 you met the idea that every calculation makes two approximations (see Lecture 1). Ten experiments later, you have seen each of them from several angles.

Approximating the wave function: the basis set. A finite basis can't represent the exact orbitals. The variational principle guarantees that a larger basis lowers the energy, and Experiment 1 showed reaction energies converging toward the complete-basis-set limit as ζ increases, at a cost that grows steeply. Experiment 6 showed the dark side: when the basis is small, its error can cancel the functional's error by luck, for one reaction and not for the next. For heavy elements, the basis comes with an effective core potential, which also brings in scalar relativistic effects (Experiments 8 and 10).

Approximating the Hamiltonian: the method. Hartree–Fock replaces electron–electron repulsion with a mean field and has no correlation. Kohn–Sham DFT is exact in principle, but every practical calculation uses an approximate E_{xc}. Jacob's ladder organizes the approximations by their ingredients (Experiment 2). Experiments 3–5 showed that the same functional can be excellent for one kind of reaction and poor for another, depending on error cancellation, dispersion, self-interaction error and static correlation. Experiments 8–10 added open shells and transition metals, where even the best functionals struggle.

Beyond those two, you met approximations that hide inside the program: the **integration grid** (Experiment 7), the **SCF solution** that may not be the lowest one (Experiment 8), the **convergence thresholds**, the **geometry** you didn't optimize, and the gap between an **electronic energy** and anything an experiment measures (Experiment 9).

15.2.2 A hierarchy of errors

The sources of error have very different typical sizes. Here they are in the order you are most likely to meet them, with the sizes you saw in this course.

Table 30: Sources of error in a DFT calculation, roughly in decreasing order of typical size.

Source of error	Typical size in this course	Where you saw it
Method (functional, or HF)	0.1–10 kcal/mol for good functionals; tens for hard cases; hundreds for HF atomizations	Experiments 3–5, 8–10
Basis set	a few kcal/mol from double- to triple-ζ; tens when a small basis is badly unbalanced	Experiments 1, 6
Missing physics (dispersion, spin–orbit, ZPE, thermal)	0.2–25 kcal/mol, depending on the reaction	Experiments 4, 9, 10
Wrong SCF solution	zero, or tens of kcal/mol: it is all or nothing	Experiment 8
Integration grid	0.01–0.5 kcal/mol in energies; more for weak interactions and frequencies	Experiment 7
SCF convergence threshold	negligible at 10^{-8}; noticeable only with loose thresholds	throughout

The general order is method > basis set > grid > SCF convergence. But the table has two warnings built in. First, the ranges overlap: a bad basis can be worse than a mediocre functional, and an unstable SCF solution can be the largest error of all. Second, several errors are **all or nothing**. A wrong spin state, a missing ECP, a forgotten dispersion correction or a wrong stoichiometric coefficient doesn't make the answer a bit worse; it makes it meaningless. Every one of them can bite, and the small ones bite exactly when the quantity you want is small, as the 0.28 kcal/mol well of Ar_2 showed.

15.2.3 The black box warning

Modern programs make a DFT calculation look easy. You draw a molecule, choose a functional from a menu, press Submit, and a number with eight decimals comes back. Nothing in that number tells you whether

- the functional is appropriate for the chemistry;
- the basis set is converged, or lucky;
- the grid is fine enough for the functional;
- the SCF found the lowest solution in the right spin state;
- the number is the quantity you meant to compare with.

Every experiment in this course was designed to expose one of these hidden choices. The program will not warn you about any of them. A quantum chemistry program is not a black box; it is a set of approximations you choose, whether you know it or not. The defaults are someone else's choices, made for someone else's problem, in some earlier version of the program. The practical rules that follow from this course are short:

1. Know which approximations you are making, and why.
2. Converge what can be converged (basis, grid, SCF), or at least check how far you are from convergence.
3. Choose the method with benchmark evidence for your kind of chemistry, not by habit or popularity.
4. Check the SCF solution and the spin state for anything open-shell or containing a metal.
5. Compare like with like: electronic energies with electronic references, enthalpies with enthalpies.

6. Report everything that changes the number.

When a problem is outside what you can check, for example strong multireference character or an unfamiliar class of transition-metal compounds, the right move is the one the experiment pages kept recommending: ask a theoretician.

15.2.4 Where the field is going

DFT has been the workhorse of computational chemistry for thirty years, and the approximations you used will remain in use for many more. Three directions are changing what is possible.

Machine-learned functionals. Instead of fitting a dozen parameters to a few hundred data points, as the Minnesota functionals did, one can train a neural network to represent the exchange–correlation energy density from the local ingredients of the density. The DM21 functional, for example, was trained on molecular data together with exact conditions for systems with fractional charge and spin, which target the delocalization and static-correlation errors of Lecture 9 [156]. Such functionals can be very accurate on their training domain. They also bring back, in a stronger form, the questions this course raised: how well they transfer to chemistry outside the training data, and how smooth, and therefore how grid-friendly, they are (Lecture 11).

Double hybrids. The fifth rung of Jacob's ladder adds a second-order perturbation correlation term computed from the virtual orbitals, as in B2PLYP [92]. In large benchmarks such as GMTKN55, the best double hybrids are the most accurate functionals overall [43], [157]. They cost about as much as MP2, more than a hybrid but far less than coupled cluster, and they need larger basis sets because the correlation part converges slowly with ζ, like the cc-pVnZ behavior you saw in Experiment 1.

Local correlation methods. The reference values in this course came from CCSD(T) at the complete-basis-set limit or from experiment. Canonical CCSD(T) scales as N^7, which limited it to small molecules. Local correlation methods exploit the fact that electron correlation is short-ranged. DLPNO-CCSD(T), for example, compresses the correlation space of each electron pair, and its cost grows nearly linearly with molecular size [158]. That makes coupled-cluster accuracy, or close to it, available for molecules with a hundred atoms or more. In practice it means two things for a DFT user: better reference values for benchmarking, and the option of checking a DFT result for your own system directly.

None of these removes the need for judgment. A machine-learned functional still needs a grid and a basis set, a double hybrid still needs a converged basis and a stable reference determinant, and a local correlation calculation has its own thresholds to report. The habits you built in this course transfer to all of them.

15.2.5 Course feedback

The last part of this lecture is for you. Before you leave, write down:

- **The muddiest point:** the idea from this course that is still the least clear to you.
- **The most useful experiment,** and why.
- **One thing to change** for next year's class.

Your instructor will collect these anonymously. They shape the next version of the course.

15.2.6 Key points

- In Experiment 10, the choices alone typically moved the answer by more than 10 kcal/mol. Dispersion was the biggest single factor; a triple-ζ basis with the def2 ECP was the sensible compromise; a small basis with -D4 can overbind.
- A good consulting report justifies the choice before computing, describes a reproducible setup, and discusses the result honestly.

- Every calculation approximates both the wave function (basis set) and the Hamiltonian (method), and also hides choices of grid, SCF solution, thresholds, geometry and the quantity compared.
- Typical error sizes: method > basis > grid > SCF convergence, but the ranges overlap, and several errors are all or nothing.
- A program is not a black box: defaults are someone else's choices.
- Machine-learned functionals, double hybrids and local correlation methods are extending what is possible, and each still needs the same care.

15.2.7 Study questions

1. Why does a dispersion correction matter so much for the palladium–hexatriene bond? Which experiment showed the same effect in a simpler system?
2. A classmate's def2-SVP result is closer to 16.2 than your def2-TZVP result with the same functional and -D4. Explain how this can happen and which result you would trust.
3. For each row of Table 30, name one experiment where that error was the largest one in the calculation.
4. Give two examples of "all or nothing" errors, and say how you would detect each.
5. Explain in your own words why defaults are a choice, using the grid as the example.
6. What problem do fractional-charge and fractional-spin conditions address in a machine-learned functional? Which class discussed that problem?
7. Why do double hybrids need larger basis sets than hybrids?
8. How could a local correlation method help you decide which functional to use for a new class of molecules?

15.2.8 Further reading

- P. Morgante and R. Peverati, "The devil in the details: a tutorial review on some undervalued aspects of density functional theory calculations" [32]: the paper this lab is built on.
- N. Mardirossian and M. Head-Gordon, "Thirty years of density functional theory in computational chemistry" [63].
- L. Goerigk et al., GMTKN55 [43], and N. Mehta, M. Casanova-Páez and L. Goerigk on double hybrids [157].
- J. Kirkpatrick et al., the DM21 machine-learned functional [156].
- C. Riplinger et al., DLPNO-CCSD(T) [158].
- A. J. Cohen, P. Mori-Sánchez and W. Yang, "Challenges for density functional theory" [57].

Bibliography

[1] E. Epifanovsky and others, "Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package," *J. Chem. Phys.*, vol. 155, p. 84801, 2021.

[2] E. Schrödinger, "Quantisierung als Eigenwertproblem," *Ann. Phys.*, vol. 384, pp. 361–376, 1926, doi: 10.1002/andp.19263840404.

[3] T. Kato, "On the eigenfunctions of many-particle systems in quantum mechanics," *Commun. Pure Appl. Math.*, vol. 10, pp. 151–177, 1957, doi: 10.1002/cpa.3160100201.

[4] M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln," *Ann. Phys.*, vol. 389, pp. 457–484, 1927, doi: 10.1002/andp.19273892002.

[5] J. C. Slater, *Phys. Rev.*, vol. 35, pp. 210–211, 1930.

[6] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Mineola, NY: Dover, 1996.

[7] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory*. Chichester: Wiley, 2000.

[8] W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, 2nd ed. Weinheim: Wiley-VCH, 2001.

[9] S. F. Boys, "Electronic wave functions. I. A general method of calculation for the stationary states of any molecular system," *Proc. R. Soc. Lond. A*, vol. 200, pp. 542–554, 1950, doi: 10.1098/rspa.1950.0036.

[10] W. J. Hehre, R. F. Stewart, and J. A. Pople, "Self-consistent molecular-orbital methods. I. Use of Gaussian expansions of Slater-type atomic orbitals," *J. Chem. Phys.*, vol. 51, pp. 2657–2664, 1969, doi: 10.1063/1.1672392.

[11] J. S. Binkley, J. A. Pople, and W. J. Hehre, "Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements," *J. Am. Chem. Soc.*, vol. 102, pp. 939–947, 1980, doi: 10.1021/ja00523a008.

[12] R. Ditchfield, W. J. Hehre, and J. A. Pople, *J. Chem. Phys.*, vol. 54, pp. 724–728, 1971.

[13] W. J. Hehre, R. Ditchfield, and J. A. Pople, *J. Chem. Phys.*, vol. 56, pp. 2257–2261, 1972.

[14] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, "Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions," *J. Chem. Phys.*, vol. 72, pp. 650–654, 1980, doi: 10.1063/1.438955.

[15] P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, vol. 28, pp. 213–222, 1973.

[16] M. M. Francl *et al.*, *J. Chem. Phys.*, vol. 77, pp. 3654–3665, 1982.

[17] F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, vol. 7, pp. 3297–3305, 2005.

[18] J. Witte, J. B. Neaton, and M. Head-Gordon, *J. Chem. Phys.*, vol. 144, p. 194306, 2016.

[19] T. H. Dunning Jr., *J. Chem. Phys.*, vol. 90, pp. 1007–1023, 1989.

[20] D. E. Woon and T. H. Dunning Jr., *J. Chem. Phys.*, vol. 98, pp. 1358–1371, 1993.

[21] R. A. Kendall, T. H. Dunning Jr., and R. J. Harrison, "Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions," *J. Chem. Phys.*, vol. 96, pp. 6796–6806, 1992, doi: 10.1063/1.462569.

[22] F. Jensen, *J. Chem. Phys.*, vol. 115, pp. 9113–9125, 2001.

[23] D. Feller, "Application of systematic sequences of wave functions to the water dimer," *J. Chem. Phys.*, vol. 96, pp. 6104–6114, 1992, doi: 10.1063/1.462652.

[24] T. Helgaker, W. Klopper, H. Koch, and J. Noga, "Basis-set convergence of correlated calculations on water," *J. Chem. Phys.*, vol. 106, pp. 9639–9646, 1997, doi: 10.1063/1.473863.

[25] A. Halkier *et al.*, "Basis-set convergence in correlated calculations on Ne, N_2, and H_2O," *Chem. Phys. Lett.*, vol. 286, pp. 243–252, 1998, doi: 10.1016/S0009-2614(98)00111-0.

[26] S. F. Boys and F. Bernardi, "The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors," *Mol. Phys.*, vol. 19, pp. 553–566, 1970, doi: 10.1080/00268977000101561.

[27] F. Jensen, *WIREs Comp. Mol. Sci.*, vol. 3, pp. 273–295, 2013.

[28] B. Nagy and F. Jensen, "Basis Sets in Quantum Chemistry," in *Reviews in Computational Chemistry*, A. L. Parrill and K. B. Lipkowitz, Eds., Hoboken, NJ: John Wiley & Sons, 2017, pp. 93–149.

[29] A. T. B. Gilbert, "IQmol Molecular Viewer."

[30] R. Peverati and D. G. Truhlar, *Philos. Trans. R. Soc. A*, vol. 372, p. 20120476, 2014.

[31] H. S. Yu, X. He, S. L. Li, and D. G. Truhlar, *Chem. Sci.*, vol. 7, pp. 5032–5051, 2016.

[32] P. Morgante and R. Peverati, "The devil in the details: A tutorial review on some undervalued aspects of density functional theory calculations," *Int. J. Quantum Chem.*, vol. 120, p. e26332, 2020.

[33] D. R. Hartree, *Math. Proc. Camb. Philos. Soc.*, vol. 24, pp. 89–110, 1928.

[34] V. Fock, *Z. Phys.*, vol. 61, pp. 126–148, 1930.

[35] C. C. J. Roothaan, "New developments in molecular orbital theory," *Rev. Mod. Phys.*, vol. 23, pp. 69–89, 1951, doi: 10.1103/RevModPhys.23.69.

[36] G. G. Hall, "The molecular orbital theory of chemical valency. VIII. A method of calculating ionization potentials," *Proc. R. Soc. London, Ser. A*, vol. 205, pp. 541–552, 1951, doi: 10.1098/rspa.1951.0048.

[37] P. Pulay, "Convergence acceleration of iterative sequences. The case of SCF iteration," *Chem. Phys. Lett.*, vol. 73, pp. 393–398, 1980, doi: 10.1016/0009-2614(80)80396-4.

[38] J. A. Pople and R. K. Nesbet, "Self-consistent orbitals for radicals," *J. Chem. Phys.*, vol. 22, pp. 571–572, 1954, doi: 10.1063/1.1740120.

[39] T. Koopmans, "Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den einzelnen Elektronen eines Atoms," *Physica*, vol. 1, pp. 104–113, 1934, doi: 10.1016/S0031-8914(34)90011-2.

[40] P.-O. Löwdin, "Quantum theory of many-particle systems. III. Extension of the Hartree–Fock scheme to include degenerate systems and correlation effects," *Phys. Rev.*, vol. 97, pp. 1509–1520, 1955, doi: 10.1103/PhysRev.97.1509.

[41] C. Møller and M. S. Plesset, "Note on an approximation treatment for many-electron systems," *Phys. Rev.*, vol. 46, pp. 618–622, 1934, doi: 10.1103/PhysRev.46.618.

[42] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, "A fifth-order perturbation comparison of electron correlation theories," *Chem. Phys. Lett.*, vol. 157, pp. 479–483, 1989, doi: 10.1016/S0009-2614(89)87395-6.

[43] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, *Phys. Chem. Chem. Phys.*, vol. 19, pp. 32184–32215, 2017.

[44] W. Kohn, "Nobel Lecture: Electronic structure of matter—wave functions and density functionals," *Rev. Mod. Phys.*, vol. 71, pp. 1253–1266, 1999, doi: 10.1103/RevModPhys.71.1253.

[45] P. Hohenberg and W. Kohn, *Phys. Rev.*, vol. 136, pp. B864–B871, 1964.

[46] M. Levy, "Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem", *Proc. Natl. Acad. Sci. USA*, vol. 76, pp. 6062–6065, 1979, doi: 10.1073/pnas.76.12.6062.

[47] L. H. Thomas, "The calculation of atomic fields," *Math. Proc. Cambridge Philos. Soc.*, vol. 23, pp. 542–548, 1927, doi: 10.1017/S0305004100011683.

[48] E. Fermi, "Un metodo statistico per la determinazione di alcune proprietà dell'atomo," *Rend. Accad. Naz. Lincei*, vol. 6, pp. 602–607, 1927.

[49] P. A. M. Dirac, "Note on exchange phenomena in the Thomas atom," *Math. Proc. Cambridge Philos. Soc.*, vol. 26, pp. 376–385, 1930, doi: 10.1017/S0305004100016108.

[50] E. Teller, "On the stability of molecules in the Thomas–Fermi theory," *Rev. Mod. Phys.*, vol. 34, pp. 627–631, 1962, doi: 10.1103/RevModPhys.34.627.

[51] W. Kohn and L. Sham, *Phys. Rev.*, vol. 140, pp. A1133–A1138, 1965.

[52] J. C. Slater, *Phys. Rev.*, vol. 81, pp. 385–390, 1951.

[53] D. M. Ceperley and B. J. Alder, "Ground state of the electron gas by a stochastic method," *Phys. Rev. Lett.*, vol. 45, pp. 566–569, 1980, doi: 10.1103/PhysRevLett.45.566.

[54] S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.*, vol. 58, pp. 1200–1211, 1980.

[55] J. P. Perdew and Y. Wang, *Phys. Rev. B*, vol. 45, pp. 13244–13249, 1992.

[56] J. P. Perdew and A. Zunger, "Self-interaction correction to density-functional approximations for many-electron systems," *Phys. Rev. B*, vol. 23, pp. 5048–5079, 1981, doi: 10.1103/PhysRevB.23.5048.

[57] A. J. Cohen, P. Mori-Sánchez, and W. Yang, *Chem. Rev.*, vol. 112, pp. 289–320, 2012.

[58] E. Caldeweyher, C. Bannwarth, and S. Grimme, *J. Chem. Phys.*, vol. 147, p. 34112, 2017.

[59] E. Caldeweyher *et al.*, *J. Chem. Phys.*, vol. 150, p. 154122, 2019.

[60] O. A. Vydrov and T. Van Voorhis, *J. Chem. Phys.*, vol. 133, p. 244103, 2010.

[61] K. Burke, "Perspective on density functional theory," *J. Chem. Phys.*, vol. 136, p. 150901, 2012, doi: 10.1063/1.4704546.

[62] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*. New York: Oxford University Press, 1989.

[63] N. Mardirossian and M. Head-Gordon, *Mol. Phys.*, vol. 115, pp. 2315–2372, 2017.

[64] J. P. Perdew and K. Schmidt, *AIP Conf. Proc.*, vol. 577, pp. 1–20, 2001.

[65] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka, "Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits," *J. Chem. Phys.*, vol. 123, p. 62201, 2005, doi: 10.1063/1.1904565.

[66] A. D. Becke, *Phys. Rev. A*, vol. 38, pp. 3098–3100, 1988.

[67] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.*, vol. 77, pp. 3865–3868, 1996.

[68] C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B*, vol. 37, pp. 785–789, 1988.

[69] J. P. Perdew, *Phys. Rev. B*, vol. 33, pp. 8822–8824, 1986.

[70] J. P. Perdew, *Phys. Rev. B*, vol. 34, p. 7406, 1986.

[71] J. P. Perdew *et al.*, *Phys. Rev. B*, vol. 46, pp. 6671–6687, 1992.

[72] A. D. Becke, "Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals," *J. Chem. Phys.*, vol. 107, pp. 8554–8560, 1997, doi: 10.1063/1.475007.

[73] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, *Phys. Rev. Lett.*, vol. 91, p. 146401, 2003.

[74] J. Sun, A. Ruzsinszky, and J. P. Perdew, *Phys. Rev. Lett.*, vol. 115, p. 36402, 2015.

[75] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun, "Accurate and numerically efficient r^2SCAN meta-generalized gradient approximation", *J. Phys. Chem. Lett.*, vol. 11, pp. 8208–8215, 2020, doi: 10.1021/acs.jpclett.0c02405.

[76] N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.*, vol. 142, p. 74111, 2015.

[77] N. Mardirossian, L. Ruiz Pestana, J. C. Womack, C.-K. Skylaris, T. Head-Gordon, and M. Head-Gordon, *J. Phys. Chem. Lett.*, vol. 8, pp. 35–40, 2017.

[78] A. D. Becke, "A new mixing of Hartree–Fock and local density-functional theories," *J. Chem. Phys.*, vol. 98, pp. 1372–1377, 1993, doi: 10.1063/1.464304.

[79] A. D. Becke, *J. Chem. Phys.*, vol. 98, pp. 5648–5652, 1993.

[80] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *J. Phys. Chem.*, vol. 98, pp. 11623–11627, 1994.

[81] C. Adamo and V. Barone, *J. Chem. Phys.*, vol. 110, pp. 6158–6170, 1999.

[82] M. Ernzerhof and G. E. Scuseria, *J. Chem. Phys.*, vol. 110, pp. 5029–5036, 1999.

[83] J. P. Perdew, M. Ernzerhof, and K. Burke, "Rationale for mixing exact exchange with density functional approximations," *J. Chem. Phys.*, vol. 105, pp. 9982–9985, 1996, doi: 10.1063/1.472933.

[84] V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, "Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes," *J. Chem. Phys.*, vol. 119, pp. 12129–12137, 2003, doi: 10.1063/1.1626543.

[85] Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, vol. 120, pp. 215–241, 2008.

[86] H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, "A long-range correction scheme for generalized-gradient-approximation exchange functionals," *J. Chem. Phys.*, vol. 115, pp. 3540–3544, 2001, doi: 10.1063/1.1383587.

[87] T. Yanai, D. P. Tew, and N. C. Handy, "A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP)," *Chem. Phys. Lett.*, vol. 393, pp. 51–57, 2004, doi: 10.1016/j.cplett.2004.06.011.

[88] N. Mardirossian and M. Head-Gordon, "\omegaB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy", *Phys. Chem. Chem. Phys.*, vol. 16, pp. 9904–9924, 2014, doi: 10.1039/C3CP54374A.

[89] R. Peverati and D. G. Truhlar, *J. Phys. Chem. Lett.*, vol. 2, pp. 2810–2817, 2011.

[90] N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.*, vol. 144, p. 214110, 2016.

[91] J. Heyd, G. E. Scuseria, and M. Ernzerhof, "Hybrid functionals based on a screened Coulomb potential," *J. Chem. Phys.*, vol. 118, pp. 8207–8215, 2003, doi: 10.1063/1.1564060.

[92] S. Grimme, *J. Chem. Phys.*, vol. 124, p. 34108, 2006.

[93] S. Grimme, *J. Comput. Chem.*, vol. 27, pp. 1787–1799, 2006.

[94] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, *J. Chem. Phys.*, vol. 132, p. 154104, 2010.

[95] R. Sabatini, T. Gorni, and S. de Gironcoli, "Nonlocal van der Waals density functional made simple and efficient," *Phys. Rev. B*, vol. 87, p. 41108, 2013, doi: 10.1103/PhysRevB.87.041108.

[96] H. S. Yu, S. L. Li, and D. G. Truhlar, *J. Chem. Phys.*, vol. 145, p. 130901, 2016.

[97] A. D. Becke, *J. Chem. Phys.*, vol. 140, p. 18A301, 2014.

[98] L. Goerigk and N. Mehta, *Aust. J. Chem.*, vol. 72, pp. 563–573, 2019.

[99] A. Karton, N. Sylvetsky, and J. M. L. Martin, *J. Comput. Chem.*, vol. 38, pp. 2063–2075, 2017.

[100] F. London, "Zur Theorie und Systematik der Molekularkräfte," *Z. Phys.*, vol. 63, pp. 245–279, 1930, doi: 10.1007/BF01421741.

[101] H. B. G. Casimir and D. Polder, "The Influence of Retardation on the London–van der Waals Forces," *Phys. Rev.*, vol. 73, pp. 360–372, 1948, doi: 10.1103/PhysRev.73.360.

[102] J. Řezáč, K. E. Riley, and P. Hobza, *J. Chem. Theory Comput.*, vol. 7, pp. 2427–2438, 2011.

[103] J. Řezáč and P. Hobza, "Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications," *Chem. Rev.*, vol. 116, pp. 5038–5071, 2016, doi: 10.1021/acs.chemrev.5b00526.

[104] S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth, "Dispersion-Corrected Mean-Field Electronic Structure Methods," *Chem. Rev.*, vol. 116, pp. 5105–5154, 2016, doi: 10.1021/acs.chemrev.5b00533.

[105] S. Grimme, S. Ehrlich, and L. Goerigk, *J. Comput. Chem.*, vol. 32, pp. 1456–1465, 2011.

[106] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, *Phys. Rev. Lett.*, vol. 92, p. 246401, 2004.

[107] K. Lee, É. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth, *Phys. Rev. B*, vol. 82, p. 81101, 2010.

[108] L. Goerigk and S. Grimme, *Phys. Chem. Chem. Phys.*, vol. 13, pp. 6670–6688, 2011.

[109] L. Goerigk, *J. Phys. Chem. Lett.*, vol. 6, pp. 3891–3896, 2015.

[110] S. Grimme, "Seemingly Simple Stereoelectronic Effects in Alkane Isomers and the Implications for Kohn–Sham Density Functional Theory," *Angew. Chem. Int. Ed.*, vol. 45, pp. 4460–4464, 2006, doi: 10.1002/anie.200600448.

[111] M. D. Wodrich, C. Corminboeuf, and P. v. R. Schleyer, "Systematic Errors in Computed Alkane Energies Using B3LYP and Other Popular DFT Functionals," *Org. Lett.*, vol. 8, pp. 3631–3634, 2006, doi: 10.1021/ol061016i.

[112] A. Karton, D. Gruzman, and J. M. L. Martin, *J. Phys. Chem. A*, vol. 113, pp. 8434–8447, 2009.

[113] P. Mori-Sánchez, A. J. Cohen, and W. Yang, "Localization and Delocalization Errors in Density Functional Theory and Implications for Band-Gap Prediction," *Phys. Rev. Lett.*, vol. 100, p. 146401, 2008, doi: 10.1103/PhysRevLett.100.146401.

[114] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, "Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy," *Phys. Rev. Lett.*, vol. 49, pp. 1691–1694, 1982, doi: 10.1103/PhysRevLett.49.1691.

[115] T. Bally and G. N. Sastry, "Incorrect Dissociation Behavior of Radical Ions in Density Functional Calculations," *J. Phys. Chem. A*, vol. 101, pp. 7923–7925, 1997, doi: 10.1021/jp972378y.

[116] H. Yu and D. G. Truhlar, *J. Chem. Theory Comput.*, vol. 11, pp. 2968–2983, 2015.

[117] T. J. Lee and P. R. Taylor, "A diagnostic for determining the quality of single-reference electron correlation methods," *Int. J. Quantum Chem.*, vol. 36, no. S23, pp. 199–207, 1989, doi: 10.1002/qua.560360824.

[118] E. R. Johnson, P. Mori-Sánchez, A. J. Cohen, and W. Yang, *J. Chem. Phys.*, vol. 129, p. 204112, 2008.

[119] K. H. Hopmann, *Organometallics*, vol. 38, pp. 603–605, 2019.

[120] H. Kruse and S. Grimme, "A geometrical correction for the inter- and intra-molecular basis set superposition error in Hartree–Fock and density functional theory calculations for large systems," *J. Chem. Phys.*, vol. 136, p. 154101, 2012, doi: 10.1063/1.3700154.

[121] H. Kruse, L. Goerigk, and S. Grimme, *J. Org. Chem.*, vol. 77, pp. 10824–10834, 2012.

[122] R. Sure, J. G. Brandenburg, and S. Grimme, "Small Atomic Orbital Basis Set First-Principles Quantum Chemical Methods for Large Molecular and Periodic Systems: A Critical Analysis of Error Sources," *ChemistryOpen*, vol. 5, pp. 94–109, 2016, doi: 10.1002/open.201500192.

[123] F. Jensen, *J. Chem. Phys.*, vol. 116, p. 3502, 2002.

[124] V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern, and L. A. Curtiss, *J. Comput. Chem.*, vol. 22, pp. 976–984, 2001.

[125] A. D. Becke, *J. Chem. Phys.*, vol. 88, pp. 2547–2553, 1988.

[126] C. W. Murray, N. C. Handy, and G. J. Laming, "Quadrature schemes for integrals of density functional theory," *Mol. Phys.*, vol. 78, pp. 997–1014, 1993, doi: 10.1080/00268979300100651.

[127] V. I. Lebedev, "Quadratures on a sphere," *USSR Comput. Math. Math. Phys.*, vol. 16, no. 2, pp. 10–24, 1976, doi: 10.1016/0041-5553(76)90100-2.

[128] P. M. W. Gill, B. G. Johnson, and J. A. Pople, "A standard grid for density functional calculations," *Chem. Phys. Lett.*, vol. 209, pp. 506–512, 1993, doi: 10.1016/0009-2614(93)80125-9.

[129] Q-Chem, Inc., "Q-Chem User's Manual: Numerical Integration Grids (XC_GRID)."

[130] S.-H. Chien and P. M. W. Gill, "SG-0: A small standard grid for DFT quadrature on large systems," *J. Comput. Chem.*, vol. 27, pp. 730–739, 2006, doi: 10.1002/jcc.20383.

[131] S. Dasgupta and J. M. Herbert, "Standard grids for high-precision integration of modern density functionals: SG-2 and SG-3," *J. Comput. Chem.*, vol. 38, pp. 869–882, 2017, doi: 10.1002/jcc.24761.

[132] S. E. Wheeler and K. N. Houk, *J. Chem. Theory Comput.*, vol. 6, pp. 395–404, 2010.

[133] N. Mardirossian and M. Head-Gordon, *J. Chem. Theory Comput.*, vol. 12, pp. 4303–4325, 2016.

[134] N. Mardirossian and M. Head-Gordon, *J. Chem. Theory Comput.*, vol. 9, pp. 4453–4461, 2013.

[135] S. Lehtola and M. A. L. Marques, "Many recent density functionals are numerically ill-behaved," *J. Chem. Phys.*, vol. 157, p. 174114, 2022, doi: 10.1063/5.0121187.

[136] K. T. Tang and J. P. Toennies, *J. Chem. Phys.*, vol. 118, pp. 4976–4983, 2003.

[137] J. Č\ižek and J. Paldus, "Stability conditions for the solutions of the Hartree–Fock equations for atomic and molecular systems. Application to the pi-electron model of cyclic polyenes," *J. Chem. Phys.*, vol. 47, pp. 3976–3985, 1967, doi: 10.1063/1.1701562.

[138] R. Seeger and J. A. Pople, *J. Chem. Phys.*, vol. 66, pp. 3045–3050, 1977.

[139] H. Fukutome, *Int. J. Quantum Chem.*, vol. 20, pp. 955–1065, 1981.

[140] R. Bauernschmitt and R. Ahlrichs, *J. Chem. Phys.*, vol. 104, pp. 9047–9052, 1996.

[141] L. Cao and U. Ryde, *Phys. Chem. Chem. Phys.*, vol. 21, pp. 2480–2488, 2019.

[142] D. Andrae, U. Häußermann, M. Dolg, H. Stoll, and H. Preuß, "Energy-adjusted ab initio pseudopotentials for the second and third row transition elements," *Theor. Chim. Acta*, vol. 77, pp. 123–141, 1990, doi: 10.1007/BF01114537.

[143] C. E. Hoyer, G. Li Manni, D. G. Truhlar, and L. Gagliardi, *J. Chem. Phys.*, vol. 141, p. 204309, 2014.

[144] S. Luo, B. B. Averkiev, K. R. Yang, X. Xu, and D. G. Truhlar, *J. Chem. Theory Comput.*, vol. 10, pp. 102–121, 2014.

[145] H. Yu and D. G. Truhlar, *J. Chem. Theory Comput.*, vol. 10, pp. 2291–2305, 2014.

[146] S. Luo and D. G. Truhlar, *J. Chem. Theory Comput.*, vol. 8, pp. 4112–4126, 2012.

[147] A. Karton, S. Daon, and J. M. L. Martin, *Chem. Phys. Lett.*, vol. 510, pp. 165–178, 2011.

[148] P. J. Linstrom and W. G. Mallard, "NIST Chemistry WebBook, NIST Standard Reference Database Number 69." National Institute of Standards, Technology, Gaithersburg, MD. doi: 10.18434/T4D303.

[149] C. J. Cramer and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, vol. 11, pp. 10757–10816, 2009.

[150] W. Zhang, D. G. Truhlar, and M. Tang, *J. Chem. Theory Comput.*, vol. 9, pp. 3965–3977, 2013.

[151] X. Xu, W. Zhang, M. Tang, and D. G. Truhlar, *J. Chem. Theory Comput.*, vol. 11, pp. 2036–2052, 2015.

[152] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, *J. Chem. Phys.*, vol. 106, pp. 1063–1079, 1997.

[153] B. Chan, P. M. W. Gill, and M. Kimura, *J. Chem. Theory Comput.*, vol. 15, pp. 3610–3622, 2019.

[154] S. Dohm, A. Hansen, M. Steinmetz, S. Grimme, and M. P. Checinski, *J. Chem. Theory Comput.*, vol. 14, pp. 2596–2608, 2018.

[155] B. B. Averkiev, Y. Zhao, and D. G. Truhlar, *J. Mol. Catal. A: Chem.*, vol. 324, pp. 80–88, 2010.

[156] J. Kirkpatrick, B. McMorrow, D. H. P. Turban, A. L. Gaunt, and others, "Pushing the frontiers of density functionals by solving the fractional electron problem," *Science*, vol. 374, pp. 1385–1389, 2021, doi: 10.1126/science.abj6511.

[157] N. Mehta, M. Casanova-Páez, and L. Goerigk, *Phys. Chem. Chem. Phys.*, vol. 20, pp. 23175–23194, 2018.

[158] C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, "Natural triple excitations in local coupled cluster calculations with pair natural orbitals," *J. Chem. Phys.*, vol. 139, p. 134101, 2013, doi: 10.1063/1.4821834.