

# **Quantum Chemistry Lab**

Francesco A. Evangelista

Emory University

Updated on September 6, 2021



# Contents

|   |           |
|---|-----------|
| <b>1 The Basics of Quantum Chemistry</b>  | <b>7</b>  |
| 1.1 What is Quantum Chemistry? . . . . .  | 7         |
| 1.2 Quantum Mechanics and the Schrödinger equation . . . . .  | 7         |
| 1.3 The Quantum Mechanics of Molecules . . . . .  | 10        |
| 1.4 Atomic units . . . . .  | 11        |
| <b>2 The Born-Oppenheimer Approximation and Potential Energy Surfaces</b>   | <b>13</b> |
| 2.1 The Born-Oppenheimer or fixed nuclei approximation . . . . .  | 13        |
| 2.2 The potential energy surface of diatomic molecules . . . . .  | 14        |
| 2.3 Potential energy surfaces of polyatomic molecules . . . . .   | 15        |
| 2.4 Where are the bonds? . . . . .  | 16        |
| 2.5 Stationary points in polyatomic molecules . . . . .   | 16        |
| <b>3 Hartree-Fock Theory</b>  | <b>19</b> |
| 3.1 Why is it difficult to solve the Schrödinger equation for systems that contain more than a few electrons? . . . . . | 19        |
| 3.2 Hartree-Fock theory and the independent particle approximation . .  | 19        |
| 3.3 Spin and the Pauli principle (antisymmetry) . . . . .   | 20        |
| 3.4 The Hartree-Fock equation . . . . .   | 21        |
| 3.5 Interpretation of the orbital energies . . . . .  | 23        |
| 3.6 Theoretical vs. experimental orbital energies . . . . .   | 24        |
| <b>4 Characterization of stationary points and vibrational analysis</b>   | <b>27</b> |
| 4.1 Characterization of stationary points via the second derivative of the energy . . . . .                             | 27        |
| 4.2 Characterization of stationary points in polyatomic molecules . . .   | 28        |
| 4.3 Vibrational frequencies . . . . .   | 28        |
| 4.4 Example: vibrational analysis of water . . . . .  | 31        |
| <b>5 Geometry optimization and transition state search</b>  | <b>35</b> |
| 5.1 Geometry optimization algorithms . . . . .  | 35        |
| 5.2 Analytical vs. numerical gradients and choice of the Hessian . . . .  | 36        |
| 5.3 Choice of coordinates . . . . .   | 37        |
| 5.4 Potential energy scans and constrained optimization . . . . .   | 38        |
| 5.5 Transition state searches . . . . .   | 39        |

|   |           |
|---|-----------|
| 5.6 Intrinsic reaction coordinate (IRC) . . . . .   | 40        |
| <b>6 Basis Sets</b>   | <b>41</b> |
| 6.1 Exact solutions to the Schrödinger equation for the hydrogen atom . . . . .   | 41        |
| 6.2 Bases, basis vectors, and basis functions . . . . .   | 42        |
| 6.3 Approximate solutions via expansion in a basis . . . . .  | 43        |
| 6.4 Gaussian-type orbitals . . . . .  | 43        |
| 6.5 Contracted Gaussian-Type Orbitals . . . . .   | 44        |
| 6.6 The linear combination of atomic orbitals approximation . . . . .   | 46        |
| 6.7 A survey of common basis sets . . . . .   | 47        |
| 6.8 The complete basis set (CBS) limit and the basis set error . . . . .  | 49        |
| <b>7 Density Functional Theory</b>  | <b>51</b> |
| 7.1 Going beyond Hartree–Fock theory: Electron correlation . . . . .  | 51        |
| 7.2 Theoretical foundations of density functional theory . . . . .  | 52        |
| 7.3 A practical guide to density functionals . . . . .  | 53        |
| 7.4 Limitations of approximate DFT functionals . . . . .  | 54        |
| 7.5 The cost of DFT and practical aspects . . . . .   | 56        |
| 7.6 Practical recommendations when choosing a set of functionals . . . . .  | 57        |
| <b>8 Open-shell species</b>   | <b>59</b> |
| 8.1 What do we mean when we talk about open-shell species? . . . . .  | 59        |
| 8.2 Spin, multiplicity, and all of that . . . . .   | 59        |
| 8.3 How do quantum chemistry codes handle (high-spin) open-shell HF and DFT computations? . . . . .                           | 61        |
| 8.4 Low-spin open shells, bond breaking processes, and other problems that require a multideterminantal description . . . . . | 63        |
| <b>9 Wave function methods</b>  | <b>65</b> |
| 9.1 Configuration interaction . . . . .   | 65        |
| 9.2 Perturbation theory . . . . .   | 67        |
| 9.3 Coupled cluster methods . . . . .   | 68        |
| 9.4 Basis sets and frozen-core treatments for correlated methods . . . . .  | 69        |
| 9.5 Cost and accuracy of wave function methods . . . . .  | 69        |
| <b>10 Thermochemistry</b>   | <b>71</b> |
| 10.1 How accurate is accurate enough? . . . . .   | 71        |
| 10.2 Types of energy computations . . . . .   | 72        |
| 10.3 Basis set incompleteness . . . . .   | 74        |
| 10.4 Treatment of the correlation energy . . . . .  | 75        |
| 10.5 Vibrational corrections and entropic terms . . . . .   | 75        |
| 10.6 Smaller terms: Corrections to the Born–Oppenheimer approximation and relativistic effects . . . . .                      | 77        |

|   |           |
|---|-----------|
| <b>10 Intermolecular interactions and solvation</b> | <b>79</b> |
| 10.1 Intermolecular interactions . . . . .          | 79        |
| 10.2 Symmetry-adapted perturbation theory . . . . . | 80        |
| 10.3 Modeling reactions in solutions . . . . .      | 81        |
| 10.4 Polarizable continuum models . . . . .         | 82        |

## List of symbols used in these notes

| Symbol                               | Meaning                                |
|--------------------------------------|--|
| $\chi_\mu(\mathbf{r})$               | Atomic orbital                         |
| $\phi_i(\mathbf{r})$                 | An orbital                             |
| $\psi_i(x)$                          | A spinorbital                          |
| $ \psi_1\psi_2 \cdots \psi_N\rangle$ | A Slater determinant for $N$ electrons |
| $\hat{H}$                            | Hamiltonian operator                   |



# 1 | The Basics of Quantum Chemistry

## 1.1 What is Quantum Chemistry?

According to Wikipedia:

Quantum chemistry, also called molecular quantum mechanics, is a branch of chemistry focused on the application of quantum mechanics in physical models and experiments of chemical systems. Understanding electronic structure and molecular dynamics using the Schrödinger equations are central topics in quantum chemistry.

Quantum chemistry, and more generally, computational chemistry is a growing subfield of chemistry. Its main purpose is to use quantum physics to predict the physical and chemical properties of molecules and materials, understand the mechanism of reactions, and connect experimental measurements (e.g., from spectroscopy) to molecular structure. The potential revolutionary impact that quantum mechanics may have on chemistry was recognized in the early days of quantum mechanics by Paul Dirac, who wrote:<sup>1</sup>

The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.

<sup>1</sup>Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, Vol. 123, No. 792 (6 April 1929)

Here we encounter one common theme of this course: that due to the difficulties of applying quantum mechanics to chemistry, we will only have access to approximate solutions. This implies that we will have methods that introduce errors, and we will have to develop an understanding what limitations each method has.

## 1.2 Quantum Mechanics and the Schrödinger equation

The basic equation at the basis of quantum mechanics is the Schrödinger equation<sup>2</sup>

<sup>2</sup>This is the **time-independent** version of the Schrödinger equation. The more general version is the time-dependent form, which can be applied to study dynamical process.

$$\hat{H}\Psi = E\Psi \quad (1.1)$$

**Hamiltonian:** the quantum mechanical operator  $\hat{H}$ .

**Wave function:**  $\Psi$  /  $\psi$  (pronounced “psahy”). Represents the state of a quantum system.

This equation contains three elements:

- The **Hamiltonian operator**  $\hat{H}$ . This is a mathematical operator, in the sense that when it is applied to a function it produces a new one. The Hamiltonian contains terms for the kinetic and potential energy.
- The **wave function**  $\Psi$ . This is a function of the coordinates of all the particle in a quantum system and contains all the information necessary to describe the state of a quantum system.
- The **energy**  $E$ . This is the energy (a real number) that corresponds to the wave function  $\Psi$ .

The Schrödinger is an **eigenvalue** equation, because it falls under the category of equations like

$$\hat{O}f(x) = \lambda f(x) \quad (1.2)$$

where the unknown function  $f(x)$  and number  $\lambda$  are such that when the operator  $\hat{O}$  is applied to  $f(x)$  we get back  $f(x)$  times a number.

■ **Example 1.2.1 (Eigenvalue equations).**

As an example consider the derivative operator  $\hat{O} = \frac{d}{dx}$ . The eigenfunctions of this operator then satisfy the following equation

$$\frac{d}{dx}f(x) = \lambda f(x) \quad (1.3)$$

This is a simple differential equation that has solution

$$f(x) = ce^{\lambda x}. \quad (1.4)$$

We can verify this by plugging in  $f(x)$  the first equation and noting that the left and right hand sides are equal

$$\frac{d}{dx}ce^{\lambda x} = \lambda ce^{\lambda x} \Rightarrow c\lambda e^{\lambda x} = c\lambda e^{\lambda x} \quad (1.5)$$

■

For a molecule, the wave function depends on the coordinates of the electrons and nuclei. For example, for one electron the wave function is a function of the electron position vector  $\mathbf{r} = (x, y, z)$  specified by the the Cartesian coordinates  $x, y, z$ ,  $\Psi(\mathbf{r})$ . From quantum mechanics, we know that the wave function should be interpreted as a **probability amplitude**, that is, if we take the modulus square the wave function we get the probability that the electron is in position  $\mathbf{r}$

$$P(\mathbf{r}) \equiv \text{probability of finding the electron in position } \mathbf{r} = |\Psi(\mathbf{r})|^2 = \Psi^*(\mathbf{r})\Psi(\mathbf{r}) \quad (1.6)$$

For this probabilistic interpretation to make sense, the sum of the probability of finding the electron in the whole space must add to one, that is, the wave function should satisfy the **normalization** condition

$$\int d\mathbf{r} P(\mathbf{r}) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz |\Psi(\mathbf{r})|^2 = 1 \quad (1.7)$$

This equation says that if we integrate (sum) the probability of finding one electron over all space, it must add up to one.

Once we have an exact or approximate wave function we can compute the **expectation value** (average value) of an operator corresponding to a quantity that we might want to determine. For example, to find out the energy we compute the expectation value of the Hamiltonian

$$E = \int \Psi^* \hat{H} \Psi \quad (1.8)$$

where the integral runs over all positions of the particle in our system.<sup>3</sup>

As pointed out by Dirac, the Schrödinger equation is hard to solve! There are only very few systems for which one can find exact solutions of the Schrödinger equation. These include a particle in a box, the harmonic oscillator, a rotating molecule. The next example discusses the particle in a box.

### ■ Example 1.2.2 (The particle in a box).

In this example we examine one of the simplest problems in quantum mechanics that can be solved exactly: a particle trapped in a one-dimensional box. We assume that the box has length  $L$ , and it runs from 0 to  $L$ . For this toy model the Hamiltonian is given by

$$\hat{H} = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2}. \quad (1.10)$$

The Schrödinger equation is given by

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x) = E\psi(x). \quad (1.11)$$

The solutions of this equation depend on a **quantum number**  $n = 1, 2, 3, \dots$ , an integer that goes from one to infinity. The energy of the  $n$ th level is

$$E_n = \frac{\hbar^2 n^2 \pi^2}{2mL^2} = \frac{\hbar^2 n^2}{8mL^2}, \quad n = 1, 2, 3 \dots \quad (1.12)$$

while the corresponding wave functions are given by

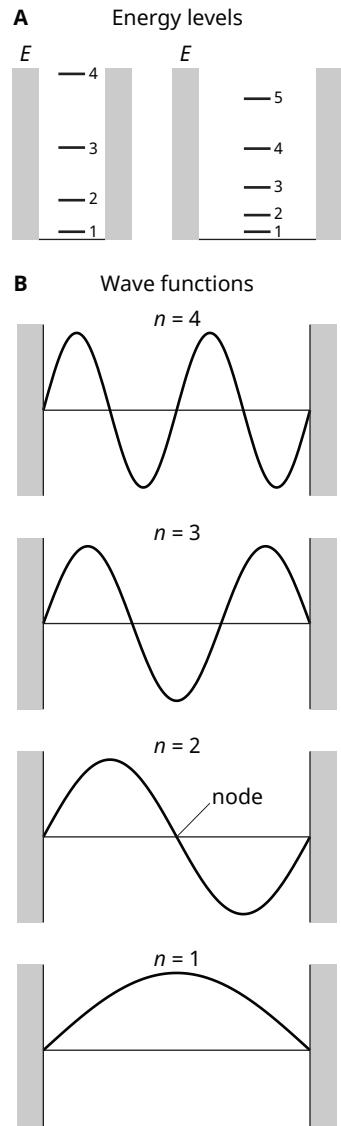
$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right). \quad (1.13)$$

The energy levels and wave functions for the particle in a box are shown in Fig. 1.1. Here we see two important features of the Schrödinger equation. Firstly, the energy levels are discrete, a phenomenon that is referred to as quantization. Since the energy is proportional to  $1/L^2$ , when we make the box wider, the spacing between the energy levels becomes smaller. Note also that the lowest energy level ( $n = 1$ ) always has a positive energy, this is usually called the **zero point energy**.

Secondly, the higher energy wave functions contain more nodes. Here by a node we mean a point where the wave function is equal to zero. This is a consequence of the fact that wave functions are orthogonal with respect to each other. You have already encountered this phenomenon when you studied the atomic orbitals of the hydrogen atom. For example, the 1s orbital has no nodes. The 2p orbitals ( $2p_x, 2p_y, 2p_z$ ) have one nodal plane, etc. ■

<sup>3</sup>This result generalizes to any operator. If the operator  $\hat{O}$  corresponds to the observable  $O$ , then the expectation value of  $O$  is given by

$$\langle O \rangle = \int \Psi^* \hat{O} \Psi \quad (1.9)$$



**Fig. 1.1:** Particle trapped in a one-dimensional box. (A) the first few energy levels as a function of the quantum number  $n = 1, 2, \dots$  for a narrow (left) and wide (right) box. (B) The first four wave functions of the particle in a box.

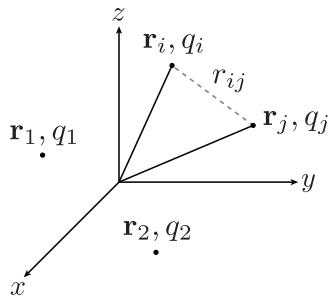
### 1.3 The Quantum Mechanics of Molecules

In an atom or a molecule, where we have more than one particle, the wave function is a function of the coordinates of all the electrons ( $\mathbf{r}_i$ ) and nuclei ( $\mathbf{R}_i$ )

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{R}_1, \mathbf{R}_2, \dots) \quad (1.14)$$

The wave function must also satisfy another important property: it must be anti-symmetric with respect to the interchange of electrons. We will return to this point later.

The Hamiltonian operator depends on the coordinates ( $\mathbf{r}_i$ ) and the charges ( $q_i$ ) of all the particles, and it can be separated into a kinetic energy operator plus a potential operator



**Fig. 1.2:** In quantum mechanics, a molecule is a collection of particles with position vector  $\mathbf{r}_i$  and charge  $q_i$ . The distance between particles  $i$  and  $j$  is  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ .

$$\hat{H} = \underbrace{\sum_i -\frac{1}{2m_i} \nabla_i^2}_{\text{kinetic}} + \underbrace{\frac{1}{4\pi\epsilon_0} \sum_{i < j} \frac{q_i q_j}{r_{ij}}}_{\text{potential (Coulomb)}} \quad (1.15)$$

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \quad (1.16)$$

where  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$  is the distance between particles  $i$  and  $j$  and the symbol  $\nabla_i^2$  is the Laplacian operator acting on particle  $i$ . Since the electron and nuclei interact via their electric charge, the potential is just the classical Coulomb potential.

In a molecule, where we have electrons and nuclei, we can separate terms that refer to electron (e) and nuclear (n) coordinates

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{ee} + \hat{V}_{en} + \hat{V}_{nn} \quad (1.17)$$

In this equation  $\hat{T}_e$  is the kinetic energy of the electrons

$$\hat{T}_e = -\frac{1}{2m_e} \sum_i^{\text{electrons}} \nabla_i^2 \quad (1.18)$$

$\hat{V}_{ee}$  is the electron-electron potential energy ( $q_i = q_j = -q_e$ ), which is always positive

$$\hat{V}_{ee} = \frac{q_e^2}{4\pi\epsilon_0} \sum_{i < j}^{\text{electrons}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (1.19)$$

and  $\hat{V}_{en}$  is the electron-nuclear potential energy (for electrons  $q_i = -q_e < 0$ , for nuclei  $q_j > 0$ ), which is always negative

$$\hat{V}_{en} = -\frac{q_e}{4\pi\epsilon_0} \sum_i^{\text{electrons}} \sum_j^{\text{nuclei}} \frac{q_j}{|\mathbf{r}_i - \mathbf{R}_j|} \quad (1.20)$$

The other terms are defined in a similar way.

In principle, to predict the energy and properties of atoms we could solve the Schrödinger equation for a molecule

$$\hat{H}\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{R}_1, \mathbf{R}_2, \dots) = E\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{R}_1, \mathbf{R}_2, \dots) \quad (1.21)$$

However, in the next chapter we will see that there is an important approximation that we can make and that simplifies this problem a bit.

## 1.4 Atomic units

When doing computations on atoms or molecules it is convenient to use **atomic units** (abbreviated a.u.), which are defined by the following conditions

$$\text{electron mass} = m_e = 1 \quad (1.22)$$

$$\text{electron charge} = e = 1 \quad (1.23)$$

$$\text{action} = \hbar = \frac{h}{2\pi} = 1 \quad (1.24)$$

$$\text{Coulomb's constant} = k_e = \frac{1}{4\pi\epsilon_0} = 1 \quad (1.25)$$

These are natural units because the charge and mass of the particles we want to describe (electrons) are close to one. In practice, it means that we will often avoid carrying around powers of ten in computations.

Another good reason for using atomic units is that computations done in the units involve numbers that are close to one, and because computers always use approximate representations for real numbers, the results will be less affected by roundoff errors. All computer programs that perform quantum chemistry computations on molecules adopt atomic units. So you will often have to convert results from atomic units to SI or other units that are commonly used in chemistry.

Two of the most useful atomic units are the **hartree** (symbol  $E_h$ ), the unit of energy, and the **bohr**, the unit of length. One hartree is a minuscule amount of energy, only about  $4.36 \times 10^{-18}$  Joules! The energy changes that we measure in chemistry typically refer to one mole of matter, so a more convenient conversion factor is

$$1 \text{ hartee} = 627.51 \text{kcal/mol} \quad (1.26)$$

This factor converts energy differences for one atom/molecule to a mole of atoms/molecules.

■ **Example 1.4.1 (Strength of a carbon hydrogen bond in methane).**

Breaking a C–H bond in methane requires about 104 kcal/mol. When converted to hartree this quantity is  $104/627.51 \text{ kcal/mol} = 0.165 E_h$ . This is (approximately) the energy difference that you would get if you performed a quantum chemistry computation in which you compare the energy of one  $\text{CH}_4$  molecule with that of the  $\text{CH}_3$  radical plus one hydrogen atom. ■

■ **Example 1.4.2 (Chemical accuracy).**

When modeling chemical reactivity, a common goal is to achieve an accuracy of 1 kcal/mol. When converted to atomic units this amount is equal to  $1/627.51 E_h = 0.0015936 E_h$ . ■

One bohr is the most probable distance between the nucleus and the electron in a hydrogen atom, and when expressed in units of angstrom it is equal to

$$1 \text{ bohr} = 0.52918 \text{ \AA} \quad (1.27)$$

Most computational chemistry programs **assume** that the geometry of a molecule is specified using in units of Å. Alternatively, the user can specify that a molecular geometry is defined in units of bohr. However, computations are typically done in units of bohr, and some times the output of a computation may be printed in these units. It is always good to double check to avoid that a geometry is specified in the correct units. A typical problem for beginners is confusing the two units. In this case, it is easy to spot the problem because something strange typically happens in a computation (all the atoms in a molecule are far apart, or scrunched up).

The following table shows the name and conversion factors between atomic units and SI units:

| Dimension | Symbol (Name)   | Value in Other Units  |
|-----------|-----------------|---|
| Length    | $a_0$ (bohr)    | $0.52918 \text{ \AA} = 0.52918 \times 10^{-10} \text{ m}$   |
| Mass      | $m_e$           | $9.1095 \times 10^{-31} \text{ Kg}$   |
| Charge    | $e$             | $1.6022 \times 10^{-19} \text{ C}$  |
| Action    | $\hbar$         | $1.05457 \times 10^{-34} \text{ J} \cdot \text{s}$  |
| Energy    | $E_h$ (hartree) | $627.51 \text{ kcal/mol}$<br>$27.211 \text{ eV}$<br>$219474.63 \text{ cm}^{-1}$<br>$4.3598 \times 10^{-18} \text{ J}$ |
| Time      |                 | $2.41889 \times 10^{-17} \text{ s} \approx 1/41.3 \text{ fs}$   |

<sup>4</sup>The quantity  $\alpha$  is also known as the fine-structure constant.

The speed of light in atomic units is  $\alpha^{-1} \approx 137$  a.u.<sup>4</sup>

## Calculation or computation?

What word should we use when we want to describe the result of some computational procedure used to determine the properties of molecule? Often the words **computation** and **calculation** are used interchangeably, however, the word computation best conveys the meaning of the results of a mathematical calculation done with a computer. Here are dictionary definitions of these two terms:

Calculation: A mathematical determination of the size or number of something

Computation: The action of mathematical calculation. The use of computers, especially as a subject of research or study.

# 2 | The Born–Oppenheimer Approximation and Potential Energy Surfaces

The previous chapter introduced the Schrödinger equation and the molecular Hamiltonian. At this point, we are going to examine an important approximation that allows us break down the complexity of the Schrödinger equation for molecules by separating the motion of electrons and nuclei. We will also formalize the concept of potential energy surface.

## 2.1 The Born–Oppenheimer or fixed nuclei approximation

One important simplification of the Schrödinger equation is to separate the motion of the electrons and the nuclei. This approach is justified by the fact that the mass of a nucleus is at least 1836 times larger than that of an electron. A good analogy is that of flies (electrons) that move around cows (nuclei). The motion of the flies does not perturb the cows, but if a cow moves then all the flies will rapidly readjust their motion to follow the cow.

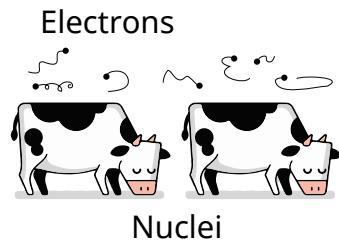
In the fixed nuclei approximation, also known as **Born–Oppenheimer (BO) approximation**, we separate the wave function of electron and nuclei in the following way

$$\Psi(\mathbf{r}, \mathbf{R}) \approx \psi_{\mathbf{R}}^{\text{el}}(\mathbf{r})\psi^{\text{nuc}}(\mathbf{R}) \quad (2.1)$$

where to keep these expression compact we have indicated the coordinates of all electrons and nuclei with a single variable, that is,  $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \dots)$  and  $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2, \dots)$ .

This equation says that the total wave function can be written as a wave function for the electrons ( $\psi^{\text{el}}$ ) times the wave function for the nuclei ( $\psi^{\text{nuc}}$ ). Note that  $\psi^{\text{nuc}}(\mathbf{R})$  only depends on the coordinate of the nuclei  $\mathbf{R}$  (nuclei are like the cows, they can almost completely ignore the exact position of the electrons). In contrast, the electron wave function  $\psi_{\mathbf{R}}^{\text{el}}(\mathbf{r})$  depends both on the coordinates of the nuclei and the electrons. The way we should think of the electronic wave function is as providing information about all the electrons **for a given position** of the nuclei  $\mathbf{R}_1, \mathbf{R}_2, \dots$  (in our analogy, the electrons are like the flies, and their position depends significantly on the position of the cows). We say that in the BO approximation the electronic wave function depends **parametrically** on the nuclear coordinates.

When we assume the Born–Oppenheimer approximation, the Hamiltonian for



**Fig. 2.1:** The fly/cow analogy of the Born–Oppenheimer approximation.

the electrons takes a simpler form

$$\hat{H}^{\text{el}} = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{en} + V_{nn} \quad (2.2)$$

The first term in this expression is just the kinetic energy of the electrons. The second term accounts for the electron-electron repulsion energy. The third term is the electron-nuclear attraction (attractive force acting on the electron). In this approximation, the last term is just a constant, which accounts for the nuclear-nuclear repulsion energy.<sup>5</sup>

The wave functions for the electronic in the BO approximation satisfies the following Schrödinger equation

$$\hat{H}^{\text{el}}\psi_{\mathbf{R}}^{\text{el}}(\mathbf{r}) = E(\mathbf{R})\psi_{\mathbf{R}}^{\text{el}}(\mathbf{r}) \quad (2.4)$$

which yields the potential energy  $E(\mathbf{R})$ , a quantity that depends on the position of the nuclei

$$E(\mathbf{R}) = E(\mathbf{R}_1, \mathbf{R}_2, \dots) \quad (2.5)$$

A useful way to think of this quantity is as describing a **potential energy surface**, that is, a high-dimensional surface that relates a molecular structure (described by nuclear coordinates  $\mathbf{R}_1, \mathbf{R}_2, \dots$ ) to its energy.

The potential energy  $E(\mathbf{R})$  is the central quantity that computational chemists aim to obtain when doing a quantum chemistry computation. This quantity should be understood as accounting for the energy of the electrons interacting with the nuclei at a given molecular geometry. The potential energy  $E(\mathbf{R})$  provides all the information we need to determine the structure of a molecule, the way it vibrates and rotates, and in most cases also the way it reacts. We will come back to some of these topics in later chapters.

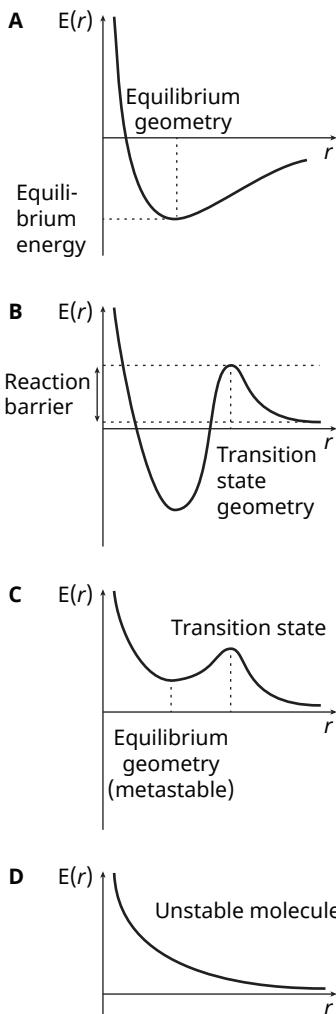
The Born-Oppenheimer approximation and the potential energy are fundamental concepts that are at the basis of how we understand and talk about molecules in chemistry. Even though you may not realize it, you are already very familiar with the BO approximation because it forms the basis of how we talk about the structure of molecules in introductory chemistry courses. When we say, for example, that a water molecule in the gas phase has a O-H bond length of 0.957 Å and a H-O-H angle of 104.5°, we are basically describing the geometric arrangements of the atoms H, H, and O with the minimum potential energy.

## 2.2 The potential energy surface of diatomic molecules

The potential energy surface for a diatomic molecule A–B is easy to visualize, because the energy depends only on the bond distance  $r = |\mathbf{R}_A - \mathbf{R}_B|$ , that is,  $E = E(r)$ , and we call it a **potential energy curve** (PEC). Fig. 2.2 shows four potential energy surfaces for a diatomic molecule. The top two plots show a stable AB molecule. In the first case (**A**), the reaction  $A + B \rightarrow AB$  is barrier-less and happens spontaneously. The bond distance  $r_e$  where the energy of AB is a minimum is the **equilibrium geometry** (often indicated with the symbol  $r_e$ ), and the corresponding energy is the **equilibrium energy**  $E(r_e)$ . In the second case (**B**), the reaction  $A + B \rightarrow AB^*$  proceeds via a **transition state**  $AB^*$  that has an energy higher than  $A + B$ . This reaction

<sup>5</sup> The nuclear-nuclear repulsion energy is given by

$$V_{nn} = \frac{1}{4\pi\epsilon_0} \sum_{i < j}^{\text{nuclei}} \frac{q_i q_j}{R_{ij}} \quad (2.3)$$



**Fig. 2.2:** The potential energy surface for a diatomic molecule. **A** and **B** show typical stable diatomic molecule. **C** and **D** show the case of a metastable molecule (stable only for a finite time) and an unstable molecule, respectively.

can happen only if A and B collide with enough energy to overcome the **reaction barrier**.

Mathematically, both equilibrium geometries and transition states correspond to **stationary points**  $r^*$  on the potential energy curve, which are characterized by a zero first derivative with respect to  $r$

$$\text{At a stationary point } r^* : \left. \frac{dE(r)}{dr} \right|_{r=r^*} = 0 \quad (2.6)$$

What distinguishes an equilibrium geometry and a transition state is the **curvature** of the potential energy curve at  $r^*$ . For equilibrium geometries, the energy is a minimum and the curvature is positive, that is, the second derivative of the potential is positive

$$\text{At a minimum } r_e : \left. \frac{d^2E(r)}{dr^2} \right|_{r=r_e} > 0 \quad (2.7)$$

A transition state instead corresponds to a negative curvature or negative second derivative

$$\text{At a transition state } r^* : \left. \frac{d^2E(r)}{dr^2} \right|_{r=r^*} < 0 \quad (2.8)$$

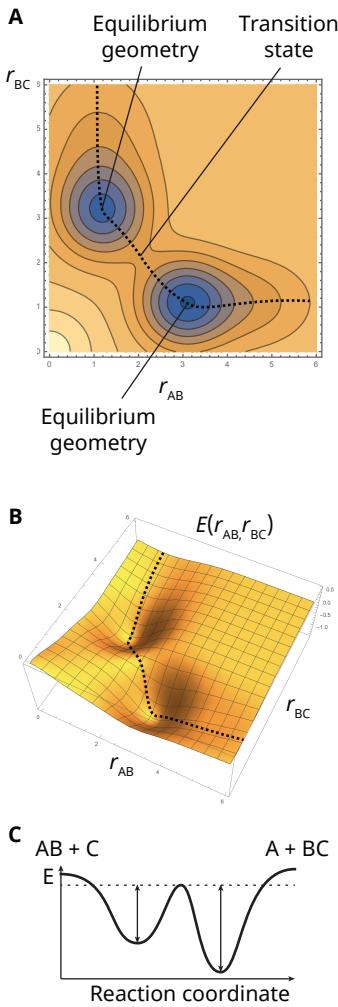
Plots **C** and **D** show two other scenarios in which the molecule AB is unstable. When there is transition state (**C**), then we call a molecule **metastable**, because it may be able to exist for a short amount of time before breaking down into A + B. If there is no transition state (**D**), then the molecule AB is unstable and the two atoms will seek to be as far as possible.

## 2.3 Potential energy surfaces of polyatomic molecules

More general potential energy surfaces are more difficult to visualize because the function  $E(\mathbf{R}_1, \mathbf{R}_2, \dots)$  corresponds to a high-dimensional hypersurface. For example, a molecule like H<sub>3</sub>CNO has potential energy surface that is 14-dimensional (13 coordinates + energy). Thinking in 3D can already be challenging, and visualizing simple objects in 4D is already outside of the ability of most people, so think how complicated it would be to visualize a curve in 13 + 1 dimensions!

A much simpler case is that of a molecule with three atoms (ABC), which corresponds to a potential energy surface in four dimensions. This number comes from realizing that there are only three degrees of freedom that can change the molecular energy, the distances A-B and B-C, and the bond angle A-B-C. However, if we keep the angle A-B-C fixed at a given angle (for example, 180°), then the resulting surface lives in only three dimensions and we can easily plot it. In this case, we can represent the potential energy surface with a contour map, in a way similar to topographic maps, which have two dimensions, and the third dimension is indicated by contour lines.

Fig. 2.3 shows the contour map for a hypothetical A—B—C system. From the PES, we can tell that there are two equilibrium structures that corresponds to energy minima. Of these two, the one with a short BC bond distance is more stable, and we would describe it as a BC molecule weakly interacting with A (A···B—C). The



**Fig. 2.3:** The potential energy surface for a linear triatomic molecule. **A** and **B** show the potential energy surface plotted using a contour plot and in 3D, respectively. A reaction path connecting  $\text{A} + \text{BC}$  to  $\text{AB} + \text{C}$  is shown, which goes through two local minima and one transition state. The energy profile for this reaction coordinate is shown in **C**.

second minimum (a local minimum) corresponds to the molecule AB weakly interacting with C ( $\text{A}-\text{B}\cdots\text{C}$ ). These two structures, are separated by a transition state  $\text{A}-\text{B}-\text{C}$ . These plots also show a the minimum energy reaction path that connects  $\text{A} + \text{BC}$  to  $\text{AB} + \text{C}$  and the corresponding energy potential. To go from  $\text{A}\cdots\text{B}-\text{C}$  to  $\text{A}-\text{B}\cdots\text{C}$ , the molecule has to overcome a reaction barrier. As you can see from the plot in **C**, the height of this barrier depends on the direction of the reaction.

## 2.4 Where are the bonds?

When we first encounter the concepts of BO approximation and potential energy surfaces, it is natural to try to embrace this new view with the help of concepts of bonding that we have acquired earlier. It is important to remember that bonds and atomic charges are interpretative concepts that are helpful to rationalize the behavior of molecules in chemistry, but do not correspond to well-defined quantum mechanical observables. By this we mean that there is no physical observable and a corresponding operator  $\hat{O}$  that we can directly link to the concepts of bonds and charges and use to extract these quantities from a quantum chemistry computation.

Therefore, it is important to be aware that when we solve the electronic Schrödinger equation of an hypothetical molecule like ABC, what we can control is only the position of the atoms A, B, and C. This implies that we cannot computed the energy of a specific Lewis structure like  $\text{A}=\text{B}^+-\text{C}^-$  or  $\text{A}-\text{B}-\text{C}$ . However, once we specify a geometry and solve the electronic Schrödinger equation, we can then analyze the electronic wave function and try to attribute it to particular Lewis structure like  $\text{A}=\text{B}^+-\text{C}^-$  or  $\text{A}-\text{B}-\text{C}$ . This process is indicated schematically below

Specify geometry  $\rightarrow$  compute  $E(\mathbf{R})$  and  $\Psi_{\mathbf{R}}(\mathbf{r})$   $\rightarrow$  analyze  $\Psi_{\mathbf{R}}(\mathbf{r})$

## 2.5 Stationary points in polyatomic molecules

In the same way we characterized the stationary points of a diatomic molecule using the first and second derivatives of the potential energy curve, we can extend the same analysis to polyatomic molecules. A stationary point is similarly defined as a geometry  $\mathbf{R}^*$  at which **the partial derivatives with respect to all coordinates are zero**

$$\text{At a stationary point } \mathbf{R}^* : \left. \frac{\partial E(\mathbf{R}_1, \mathbf{R}_2, \dots)}{\partial \mathbf{R}_i} \right|_{\mathbf{R}=\mathbf{R}^*} = 0, \quad i = 1, \dots, N_{\text{atoms}} \quad (2.9)$$

To characterize the nature of a stationary point it is necessary to examine the curvature of the PES. In the case of a local or global minimum, the stationary point must have a positive curvature along any direction. Going back to the example in Fig. 2.3, you can see that both minima look like a 2D parabola (paraboloid), and moving away from the minimum always leads to an increase in energy.

Transition states are instead characterized by having at least one direction along which the curvature is negative (negative second derivative), while in any other orthogonal direction the curvature is positive. In the example in Fig. 2.3, at the transition state, the direction with negative curvature is a line along the dashed line.

We will come back to the question of how to characterize stationary points and give a more precise mathematical definition when we study vibrations in molecules.

## Summary

- ☞ The Born–Oppenheimer approximation separates the motion of electrons and nuclei to simplify the Schrödinger equation.
- ☞ The potential energy surface  $E(\mathbf{R})$  is the solution of the electronic Schrödinger equation at a fixed nuclear configuration  $\mathbf{R} \equiv (\mathbf{R}_1, \mathbf{R}_2, \dots)$ .
- ☞ The stationary points of the potential energy surface represent either stable molecular configurations (minima) or transition states (saddle points). These can be characterized by the curvature of the  $E(\mathbf{R})$  at a stationary point.



# 3 | Hartree-Fock Theory

## 3.1 Why is it difficult to solve the Schrödinger equation for systems that contain more than a few electrons?

In this chapter we will study the simplest way to solve the electronic Schrödinger equation. However, before we turn to finding approximate solutions, we should ask: Why can't we solve the Schrödinger equation for systems that contain more than a few electrons? The problem encountered when one tries to solve the Schrödinger equation beyond a few electrons is so important that it has its own name: **the curse of dimensionality**. This term is used to indicate any problem where the computational cost increases **exponentially** with the size of the problem (for example, the number of electrons).

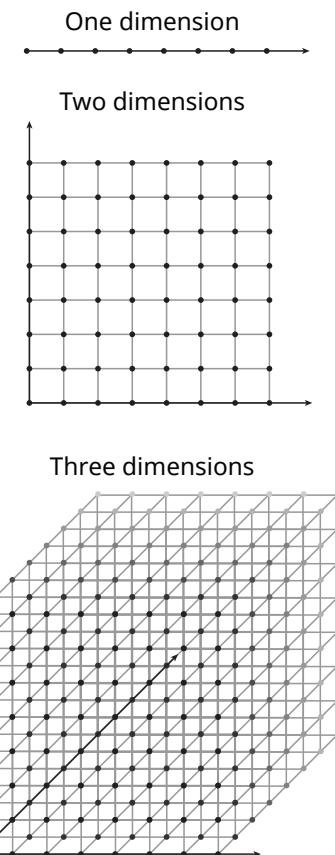
Why do we run into the curse of dimensionality when we solve the Schrödinger equation for many-electron systems? Imagine representing the wave function

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad (3.1)$$

on a computer using a grid that has 8 points along each x, y, z coordinate of each electron. To specify the wave function, we have to know the value of  $\Psi$  at all points in this multidimensional grid. For  $N$  electrons, the grid will have a total of  $8^{3N}$  points since we use 8 points for each of the  $3N$  coordinates. Such a grid is shown in Fig. 3.1 for the case of one, two, and three dimensions. Even for as little as six electrons, the exact wave function would require storing the value of  $\Psi$  at  $1.8 \times 10^{16}$  points. To store this information we would need 131,072 terabytes of memory! (at the time of writing, the typical memory capacity of a computer is less than a few terabytes of data).

## 3.2 Hartree-Fock theory and the independent particle approximation

The simplest quantum chemistry method is the Hartree-Fock approach. This method relies on an independent particle approximation, where the wave function is expressed as a product of wave functions of single electrons (orbitals). We will indicate orbitals with the lower-case Greek letter phi ( $\phi$ ). An orbital is a function of the coordinate vector  $\mathbf{r} = (x, y, z)$  for an electron. To distinguish between different orbitals, we will label them with an index and write them as



**Fig. 3.1:** Illustration of the curse of dimensionality. If we specify the value of the wave function as points on a grid, the amount of memory necessary to represent the wave function scales exponentially with the number of variables (dimension).

$$\phi_i(\mathbf{r}) \quad (3.2)$$

In the early days of quantum mechanics, Douglas Hartree proposed to approximate the wave function of  $N$  electrons as a product of the orbitals for each electron

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \approx \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2) \cdots \phi_N(\mathbf{r}_N) \quad (3.3)$$

This approximation reduces the problem of solving the Schrödinger equation for a wave function of  $3N$  coordinates into a simpler problem: we only need to determine  $N$  wave functions that depend on 3 spatial coordinates. This simplification makes Hartree's approach much more affordable than a direct solution of the Schrödinger equation!

### 3.3 Spin and the Pauli principle (antisymmetry)

The wave function form of Eq. (3.3) is still missing two important ingredients: 1) it **does not account for the spin of electrons** and 2) it **does not obey the Pauli exclusion principle**. These issues were first realized by the physicist Vladimir Fock, who proposed a way to correct Hartree's method for this missing ingredient.

Spin can be easily introduced by recognizing that the wave function of each electron must be accompanied by a spin function, that is, a label for the spin of the electron. As you have learned from previous classes, the spin of an electron can be either up (alpha,  $\alpha$ ) or down (beta,  $\beta$ ). So if we want to specify that an electron occupies the orbital  $\phi_i$  and has spin alpha, we write the wave function  $\phi_i(\mathbf{r})\alpha$ . We can represent this wave function graphically with an orbital diagram in which one electron with up spin occupies the level  $i$

$$\phi_i(\mathbf{r})\alpha \equiv \uparrow_i \quad (3.4)$$

In the case of beta electron, the graphical representation is similar but with a down arrow

$$\phi_i(\mathbf{r})\beta \equiv \downarrow_i \quad (3.5)$$

When we do not want to be specific about the spin of the electron (and thus allow for both options), we will use instead the **spinorbital** notation based on the Greek letter psi [ $\psi(x)$ ] and use the variable  $x = \{\mathbf{r}, \sigma\}$  to represent both the position vector ( $\mathbf{r}$ ) and one spin coordinate ( $\sigma$ ). For the moment, we can think of spin orbitals as always coming in alpha/beta pairs that share the same spatial orbital. To distinguish between alpha/beta orbitals we place a bar above the orbital index

$$\psi_i(x) = \phi_i(\mathbf{r})\alpha \quad (3.6)$$

$$\bar{\psi}_i(x) = \phi_i(\mathbf{r})\beta \quad (3.7)$$

Now we can be more precise and write the product wave function as

$$\Psi(x_1, x_2, \dots, x_N) \approx \psi_1(x_1)\bar{\psi}_1(x_2) \cdots \quad (3.8)$$

Equation (3.8) is still not fully correct. To make sure that the wave function satisfies the Pauli exclusion principle, we have to enforce that the sign of  $\Psi(x_1, x_2, \dots, x_N)$

changes if we exchange the labels of any two particles. To illustrate this point, consider two electrons. The wave function  $\Psi(x_1, x_2)$  depends on the position and spin of each electron  $(x_1, x_2)$ , and it must change sign when we replace  $x_2$  with  $x_1$  and vice versa, which means<sup>6</sup>

$$\Psi(x_1, x_2) = -\Psi(x_2, x_1) \quad (3.9)$$

It is easy to write down a wave function that satisfies this condition if we take a product of orbitals and make it antisymmetric with respect to the electron coordinates  $x_1$  and  $x_2$ . Such a wave function is called a **Slater determinant**<sup>7</sup> and for the two electron case is it written as<sup>8</sup>

$$\Psi_{SD}(x_1, x_2) = |\psi_i \psi_j\rangle = \frac{1}{\sqrt{2}} [\psi_i(x_1)\psi_j(x_2) - \psi_i(x_2)\psi_j(x_1)] \quad (3.10)$$

A Slater determinant can be represented graphically with a molecular orbital diagram. Each orbital is represented with a line and filled with electrons (represented by arrows). For example, the electron configuration with two electrons filling the orbital  $\phi_1$  corresponds to the following Slater determinant

$$|\psi_1 \psi_1\rangle = |\phi_1 \alpha \phi_1 \beta\rangle \equiv \begin{array}{c} \uparrow \downarrow \\ \text{---}^2 \end{array} \quad (3.11)$$

This electron arrangement is called a **closed shell**, because all the electrons are paired. It corresponds to a **singlet** electronic state.

However, we could also arrange the electrons in two different orbitals both with alpha spin in an **open-shell** configuration

$$|\psi_1 \psi_2\rangle = |\phi_1 \alpha \phi_2 \alpha\rangle \equiv \begin{array}{c} \uparrow \quad 2 \\ \uparrow \quad 1 \end{array} \quad (3.12)$$

In this state, the spin of the electrons add up and we get what is called a **triplet state**.

We have covered a lot of details in this section, so I want to make sure that the important message comes across: **in the Hartree-Fock method the wave function of many electrons is approximated with a single Slater determinant**. This approach accounts for spin and satisfies the Pauli exclusion principle. We can also conveniently represent Slater determinants graphically with orbital diagrams.

## 3.4 The Hartree-Fock equation

The main goal of the Hartree-Fock method is to find the orbitals that give the “best” possible wave function. Here by “best” we mean precisely the one that minimizes the expectation value of the energy, which is given by the integral

$$E_{SD} = \int dx_1 dx_2 \dots \Psi_{SD}^*(x_1, x_2, \dots) \hat{H} \Psi_{SD}(x_1, x_2, \dots) \quad (3.13)$$

<sup>6</sup>This wave function satisfies the Pauli principle in the sense that if the two electron have the same **space and spin** coordinates ( $x_1 = x_2 = x$ ), then the wave function is zero because the condition

$$\Psi(x, x) = -\Psi(x, x)$$

can be satisfied only if  $\Psi(x, x) = 0$ .

<sup>7</sup> The Slater determinant is named after John C. Slater, who introduced the determinant in 1929 as a means of ensuring the antisymmetry of a many-electron wave function, although the wave function in the determinant form first appeared independently in Heisenberg's and Dirac's articles three years earlier. (Source: Wikipedia)



**Fig. 3.2:** John C. Slater.

<sup>8</sup>Here I use the symbol  $|\psi_i \psi_j \dots\rangle$  to indicate a general Slater determinant.

The **variational** theorem guarantees that the energy of an approximate wave function  $\tilde{\Psi}$  is always greater than the exact energy,  $\tilde{E} \geq E_{\text{exact}}$ , and the closer we get to the exact energy  $E_{\text{exact}}$ , the better our wave function  $\tilde{\Psi}$  approximate the exact solution  $\Psi_{\text{exact}}$ .

The orbitals that minimize the energy  $E_{\text{SD}}$  satisfy the **Hartree-Fock equations**

$$\hat{f}\psi_i(x) = \varepsilon_i\psi_i(x), \quad \text{for all } \psi_i(x) \quad (3.14)$$

where  $\hat{f}$  is the **Fock operator**. This is an eigenvalue equation like the Schrödinger equation, but it is simpler because it only involves one orbital at the time (a functions of 3 coordinates only). However, it is important to remember that this equation is valid only in the Hartree-Fock approximation. The quantity  $\varepsilon_i$  is called the **orbital energy** and it is the eigenvalue corresponding to the orbital  $\psi_i$ .

The physical interpretation of the Hartree-Fock equation is that we can approximate the interaction (repulsion) of an electron in orbital  $\psi_i$  with all the remaining  $N - 1$  electrons via an average potential. The Fock operator is defined as

$$\hat{f} = \underbrace{\frac{1}{2}\nabla^2}_{\text{kinetic}} + \underbrace{-\sum_i^{\text{nuclei}} \frac{Z_i}{|\mathbf{r} - \mathbf{R}_i|}}_{\text{electron-nuclear}} + \underbrace{\int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}}_{\text{Coulomb interaction}} + \text{exchange interaction} \quad (3.15)$$

and it contains two terms that depend on the distribution of electrons. One of these two terms is the **Coulomb interaction** of an electron with all the other electrons and it is proportional to the density of electrons ( $\rho$ ). This term has a classical interpretation: it represents the Coulomb interaction of an electron with a charge distribution.<sup>9</sup> The electron density is in turn given by the sum of the square of the occupied orbitals, which for a closed-shell molecule is given by

$$\rho(\mathbf{r}) = 2 \sum_i^{\text{occupied}} |\phi_i(\mathbf{r})|^2 \quad (3.16)$$

The Hartree-Fock equations and these last two equations tell us that to find the orbital we need to know  $\hat{f}$ , but to know  $\hat{f}$  we need to know the density, which is determined by the orbitals. We can illustrate this relationship between these quantities by the diagram

$$\psi_i \rightarrow \hat{f} \rightarrow \rho \rightarrow \psi_i \quad (3.17)$$

where “ $\rightarrow$ ” means that the quantity on the left depends on the quantity on the right. In practice since to solve for  $\psi_i$  we need to know  $\psi_i$ , the Hartree-Fock equations have to be solved via a **self-consistent-field (SCF) procedure** where one first starts with a guess for the orbitals and then proceeds to improve the solution iteratively.

The SCF procedure consists of the following steps:

- 1. Forming an initial guess.** At beginning of an SCF procedure we need to start from orbitals that are close to the optimal ones. Common ways to guess the orbitals include neglecting electron repulsion or forming the average potential from a superposition of atomic densities.

<sup>9</sup>The term labeled **exchange interaction** is a bit more complicated to write down. It comes from imposing the Pauli principle and it does not have a classical interpretation.

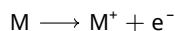
2. **Updating the Fock matrix.** Using the current set of orbitals, the average potential and the Fock operator are built.
3. **Determining the orbitals and orbital energies.** Using the current Fock matrix, the Hartree–Fock equations are solved to obtain one set of orbitals and orbital energies and an updated value for the total energy.
4. **Convergence check.** The program checks if the change in energy and orbitals with respect to the previous iteration is less than a predefined convergence threshold. If the computation has converged, we stop. Otherwise, we go back to step 2. and use the new set of orbital to compute a new Fock operator.

The following is the output of a Hartree–Fock computation on water. In this example, the iterations start with a guess of the density given by the superposition of atomic densities (SAD)<sup>10</sup>. As the iterations progress, the energy becomes lower and lower (in accordance with the variational principle) and we reach convergence in eight iterations to the final value  $-76.02665366185 E_h$ . Note that in the last step the energy is converged to  $-5.4 \times 10^{-10} E_h$ .<sup>11</sup>

```
Water/Hartree-Fock
==> Iterations <==
      Total Energy      Delta E      RMS |[F,P]|
@RHF iter SAD: -75.50773245935 -7.55077e+01 0.00000e+00
@RHF iter  1: -75.95378451304 -4.46052e-01 3.03072e-02 DIIS
@RHF iter  2: -76.00708060915 -5.32961e-02 1.73589e-02 DIIS
@RHF iter  3: -76.02624491762 -1.91643e-02 1.58426e-03 DIIS
@RHF iter  4: -76.02663310976 -3.88192e-04 3.65265e-04 DIIS
@RHF iter  5: -76.02665270339 -1.95936e-05 6.70876e-05 DIIS
@RHF iter  6: -76.02665363515 -9.31757e-07 1.06688e-05 DIIS
@RHF iter  7: -76.02665366131 -2.61654e-08 1.50538e-06 DIIS
@RHF iter  8: -76.02665366185 -5.39330e-10 3.50179e-07 DIIS
Energy and wave function converged.
```

## 3.5 Interpretation of the orbital energies

It is important to understand the precise meaning of the orbital energies  $\varepsilon_i$  because it is slightly different than our intuition as chemists. **Koopmans' theorem** shows that the orbital energies are related to the ionization energy and electron affinity of a molecule. For orbitals that are occupied by electrons, Koopmans' theorem says that  $-\varepsilon_i$  correspond to the energy necessary to remove one electron (ionize), like in the process



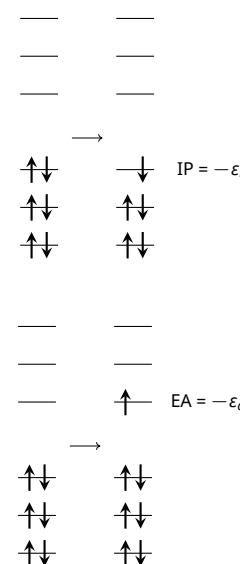
Koopmans' theorem for occupied orbitals can be stated as

$$\text{Ionization potential for electron in } \psi_i = \text{IP}_i = E^{N-1} - E^N = -\varepsilon_i. \quad (3.18)$$

<sup>10</sup>Computational chemists love acronyms. With time you will become conversant with this language.

<sup>11</sup>Quantum chemistry computations are usually converged to high accuracy. It is not uncommon to see the energy optimized to less than  $10^{-7}$ – $10^{-8} E_h$ , and in certain cases down to even  $10^{-12} E_h$ .

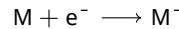
The quantity labeled “RMS |[F,P]|” measures the derivative of the energy, and it is another way to monitor the convergence (remember that at a minimum, the gradient is zero). The label “DIIS”, indicates that the code is using the direct inversion of the iterative subspace (DIIS) algorithm to accelerate the convergence.



**Fig. 3.3:** Illustration of Koopmans' theorem.

If an occupied orbital has a negative energy, then we have to put energy into a molecule to remove that electron (the IP is positive).

Similarly, for orbitals that are not occupied, the orbital energy  $\varepsilon_a$  corresponds to the energy released when an electron is added to form an anion (if the starting molecule is neutral), like in the process



In this case, Koopmans' theorem helps us quantify the electron affinity

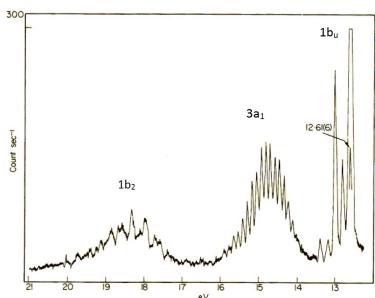
$$\text{Electron affinity for electron in } \psi_a = EA_a = E^N - E^{N+1} = -\varepsilon_a. \quad (3.19)$$

Therefore, if an unoccupied orbital has a negative energy (positive EA), a molecule will attach an electron and form a stable negative ion. If instead  $\varepsilon_a$  is positive, a molecule will not accept an electron, and its anion will be unstable with respect to self ionization (the anion will spontaneously lose an electron in a finite amount of time).

### 3.6 Theoretical vs. experimental orbital energies

The importance of Koopmans' theorem is that it gives **physical meaning to the Hartree-Fock orbital energies and orbitals**. Using Koopmans' theorem, it is possible to bridge the results of Hartree-Fock computation to experiments that measure the IP and EA of molecules. For example, **photoelectron spectroscopy** can measure the ionization potential (also called the binding energy) of electrons in different orbitals. In this spectroscopy, a molecule is ionized with a high energy photon and one electron is released with kinetic energy equal to the energy of the photon ( $h\nu$ ) minus the binding energy,  $h\nu - IP_i$ . Therefore, it allows to measure the binding energy of different orbitals. An example of the photoelectron spectrum of water in the gas phase is given in Fig. 3.4.

The example below shows the Hartree-Fock orbital energies of the occupied and unoccupied orbitals of an isolated water molecule in the gas phase.



**Fig. 3.4:** Photoionization spectrum of water in the gas phase (in eV). Source: Wikipedia Commons.

| Water/Hartree-Fock<br>Orbital Energies [Eh] |            |     |           |     |           |
|---|------------|-----|-----------|-----|-----------|
| <hr/>                                       |            |     |           |     |           |
| Doubly Occupied:                            |            |     |           |     |           |
| 1A1   | -20.550918 | 2A1 | -1.335304 | 1B2 | -0.697799 |
| 3A1   | -0.566090  | 1B1 | -0.492954 |     |           |
| <br>Virtual:                                |            |     |           |     |           |
| 4A1   | 0.185103   | 2B2 | 0.255850  | 3B2 | 0.787301  |
| 5A1   | 0.851798   | 6A1 | 1.163709  | 2B1 | 1.200353  |
| 4B2   | 1.253480   | 7A1 | 1.444918  | 1A2 | 1.475588  |
| 3B1   | 1.674083   | 8A1 | 1.867861  | 5B2 | 1.931955  |
| 6B2   | 2.446380   | 9A1 | 2.483524  | 4B1 | 3.283306  |

```

2A2      3.336170    10A1      3.506961    11A1      3.862825
7B2      4.144454

Final Occupation by Irrep:
      A1      A2      B1      B2
DOCC [      3,      0,      1,      1 ]

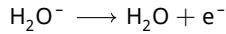
@RHF Final Energy: -76.02665366185380

=> Energetics <=

Nuclear Repulsion Energy =          9.1681932964243487
One-Electron Energy =           -123.1035625229022514
Two-Electron Energy =            37.9087155646241101
Total Energy =                  -76.0266536618537998

```

The lowest ionization potential of water is estimated according to Koopmans' theorem from the energy of the highest occupied MO (HOMO, labeled  $1b_1$ ), which has an energy of  $-0.493 E_h$ . The corresponding IP is equal to  $0.493 E_h = 13.4$  eV, a value that compares well with the experimental gas-phase IP (ca. 12.6 eV, indicated in the figure). The ionization potential for the second orbital (labeled  $3a_1$ ) is  $0.566 E_h = 15.4$  eV. This corresponds to the middle broad band in the spectrum centered around 15 eV. Note that Koopmans' theorem applied to the Hartree-Fock solution also predicts that the anion  $H_2O^-$  is unstable with respect to the autoionization process



since the energy of the first unoccupied orbital (virtual) is positive ( $0.185 E_h$ ).



# 4 | Characterization of stationary points and vibrational analysis

In this chapter we cover how to characterize stationary points that are found via geometry optimization. We will discuss harmonic vibrational analysis and the computation of vibrational frequencies.

## 4.1 Characterization of stationary points via the second derivative of the energy

One of the most important applications of quantum chemistry is to find and identify the stationary points of potential energy surfaces. Recall that a stationary point is a molecular geometry for which all first derivatives are null

$$\frac{\partial E(\mathbf{R})}{\partial \mathbf{R}_i} \Big|_{\mathbf{R}=\mathbf{R}^*} = 0, i = 1, \dots, N_{\text{atoms}} \quad (4.1)$$

Points on the PES that correspond minima are stable molecular geometries, while saddle points correspond to transition states.

How can we distinguish minima from transition states? As we discussed in the Chapter on the Born–Oppenheimer approximation, we need to know the curvature of the PES near a stationary point. It helps that near a stationary point the PES is a **quadratic function** of the nuclear coordinates. To appreciate this point consider the case of a diatomic molecule A–B with potential energy curve  $E(r)$  as a function of the bond distance  $r$ . At a stationary point  $r^*$ , we have that  $E'(r) = 0$ . Therefore, if we write the Taylor series for  $E(r)$  centered around  $r^*$  it can be simplified to

$$\begin{aligned} E(r) &= E(r^*) + \underbrace{E'(r^*)(r - r^*)}_{=0} + \frac{1}{2}E''(r^*)(r - r^*)^2 + \frac{1}{6}E'''(r^*)(r - r^*)^3 + \dots \\ &= E(r^*) + \frac{1}{2}E''(r^*)(r - r^*)^2 + \dots \end{aligned} \quad (4.2)$$

In the second line we keep only the leading terms in the quantity  $r - r^*$  [assuming that  $E''(r^*) \neq 0$ ]. This shows that near the stationary point, the PEC  $E(r)$  **looks like a parabola** centered around  $r^*$ . The curvature of this parabola depends on  $E''(r^*)$ ,

the **second derivative** of the PEC at the stationary point. If we compute  $E''(r^*)$  then we can determine if the stationary point of a diatomic is a minimum or a transition state.

## 4.2 Characterization of stationary points in polyatomic molecules

For polyatomic molecules the energy is a function of all the atomic coordinates and the Taylor series near a stationary point is given by a more complex formula

$$E(\mathbf{R}) = E(\mathbf{R})^* + \frac{1}{2} \sum_{ij}^{3N_{\text{atoms}}} \left. \frac{\partial^2 E(\mathbf{R})}{\partial R_i \partial R_j} \right|_{\mathbf{R}=\mathbf{R}^*} (R_i - R_i^*)(R_j - R_j^*) + \dots \quad (4.3)$$

In writing this equation, we use the symbols  $R_i$  and  $R_j$  to indicate the  $x, y, z$  coordinates of all atoms, which in total are  $3N_{\text{atoms}}$ . This expression is still a quadratic function in the displacements  $R_i - R_i^*$  and  $R_j - R_j^*$ , but the second derivative is replaced with the **Hessian matrix**, the matrix of mixed second derivatives

$$H_{ij} = \left. \frac{\partial^2 E(\mathbf{R})}{\partial R_i \partial R_j} \right|_{\mathbf{R}=\mathbf{R}^*} \quad (4.4)$$

To characterize the nature of a stationary point it is necessary to consider the eigenvalues of the Hessian matrix at the stationary point. If none of the eigenvalues is equal to zero, we can distinguish three cases:

- If all the eigenvalues are positive, then a stationary point is a (local) minimum.
- If some eigenvalues are positive and some are negative, then a stationary point is a saddle point. **Transition states** are characterized by only one negative eigenvalue and are also called a **first-order saddle points**.
- If the Hessian has only negative eigenvalues then the energy is a (local) maximum.

## 4.3 Vibrational frequencies

Once we compute the Hessian matrix, it is possible to compute approximate **harmonic vibrational frequencies**. Vibrations in a molecule are due to the motion of the nuclei. So to obtain information about vibrations we have to solve the nuclear Schrödinger equation assuming the Born–Oppenheimer approximation

$$[\hat{T}_n + E(\mathbf{R})]\Psi_\nu(\mathbf{R}) = E_\nu\Psi_\nu(\mathbf{R}) \quad (4.5)$$

This equation describes the quantum mechanical states of nuclei that move on the potential energy surface  $E(\mathbf{R})$ . The electrons never appear explicitly in this equation, but their effect is implicitly included in the PES (which we get from the electronic Schrödinger equation).

Solving the nuclear Schrödinger equation near a stationary point is particularly simple because the potential is quadratic

$$E(\mathbf{R}) = E(\mathbf{R})^* + \frac{1}{2} \sum_{ij} H_{ij}(R_i - R_i^*)(R_j - R_j^*) \quad (4.6)$$

For a diatomic molecule, the Hamiltonian is equivalent to that of a **quantum harmonic oscillator**. This can be more easily seen if we change coordinate from  $r$  to  $x = r - r^*$

$$\begin{aligned} \hat{T}_n + E(\mathbf{R}) &= -\frac{1}{2\mu} \frac{d^2}{dr^2} + E(r^*) + \underbrace{\frac{1}{2} E''(r^*)}_{V_0} (r - r^*)^2 \\ &= \underbrace{\left( -\frac{1}{2\mu} \frac{d^2}{dx^2} + \frac{1}{2} kx^2 \right)}_{\text{harmonic oscillator}} + V_0 \end{aligned} \quad (4.7)$$

where  $V_0$  is the energy at the bottom of the well (the energy you get when you optimize the geometry of the diatomic molecule),  $\mu = (M_A + M_B)/M_A M_B$  is the reduced mass of the diatomic molecule, and  $k = E''(r^*)$  is the force constant.

The quantum harmonic oscillator is one of those few problems for which we know analytical solutions to the Schrödinger equation. The eigenvalues of the quantum harmonic oscillator are given by (in atomic units)

$$E_v = V_0 + \sqrt{\frac{k}{\mu}} \left( v + \frac{1}{2} \right) = V_0 + \omega \left( v + \frac{1}{2} \right), \quad v = 0, 1, 2, \dots \quad (4.8)$$

where  $\omega$  is the angular frequency of the oscillator and  $v$  is the vibrational quantum number. The ground state of the harmonic oscillator (the minimum energy) is obtained for  $v = 0$  and it is equal to

$$E_0 = V_0 + \frac{1}{2}\omega \quad (4.9)$$

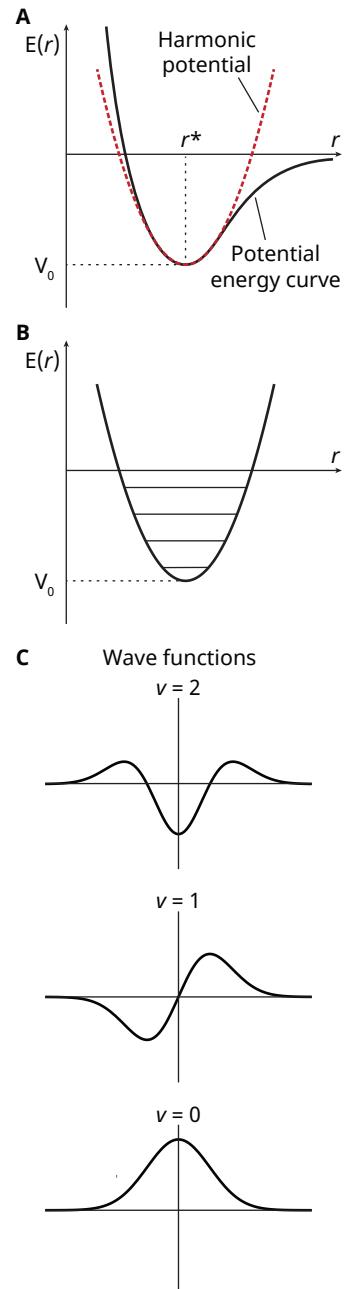
Note that this quantity is always **higher** than the energy at the bottom of the potential ( $V_0$ ) by an amount equal to  $\frac{1}{2}\omega$ ! This amount of energy is called the **zero-point energy**, and it is due to the quantum nature of vibrations.

For polynuclear molecules, the solution to the nuclear Schrödinger equation can be found by diagonalization of the mass-weighted Hessian matrix

$$\tilde{H}_{ij} = \frac{H_{ij}}{\sqrt{M_i M_j}} \quad (4.10)$$

Diagonalizing the mass-weighted Hessian matrix means finding a set of eigenvectors  $\mathbf{u}^{(a)}$  and corresponding eigenvalues  $\lambda_a$  that satisfy the equation

$$\sum_j \tilde{H}_{ij} u_j^{(a)} = \lambda_a u_i^{(a)} \quad (4.11)$$



**Fig. 4.1:** The potential energy surface of a diatomic molecule near the equilibrium bond distance ( $r_{eq}$ ) can be approximated by a quadratic function (parabola, see A). B and C show the energy levels and wave functions of the diatomic molecule vibrating with a quadratic potential.

Each eigenvector/eigenvalue pair corresponds to a molecular vibrational mode  $a$ , which we call a **normal vibrational mode**. Each mode behaves as an independent harmonic oscillator with angular frequency

$$\omega_a = \sqrt{\lambda_a} \quad (4.12)$$

and has an associated quantum number  $v_a = 0, 1, 2, \dots$

Once the mass-weighted Hessian matrix is diagonalized, the eigenvalues of the nuclear Schrödinger equation (nuclear energy levels) are given by

$$E_v = V_0 + \sum_{v_a} \omega_a \left( \frac{1}{2} + v_a \right) \quad (4.13)$$

where the quantities  $v_a$  are the quantum numbers for each vibrational mode. For a general molecule, the zero-point energy is given by

$$E_0 = V_0 + \frac{1}{2} \sum_{v_a} \omega_a \quad (4.14)$$

which again we see is higher than the energy at the bottom of the well.

The harmonic vibrational analysis introduced here can be performed with the majority of quantum chemistry codes. There are three main applications of a vibrational analysis

- The **sign** of the eigenvalues  $\lambda_a$  tell us information about the curvature of the PES at the stationary point. If all the  $\lambda_a$  are positive the stationary point is a minimum. **This corresponds to the case where all the frequencies are real**, since  $\omega_a = \sqrt{\lambda_a}$ . The presence of one or more **negative** eigenvalues indicates that there is a vibrational mode with a negative curvature. In this case a stationary point is a transition state. This case corresponds to finding an imaginary frequency since if an eigenvalue  $\lambda_a$  is negative, its square root will be an imaginary number,  $\omega_a = i\sqrt{|\lambda_a|}$ .
- The **value** of the individual vibrational frequencies  $\omega_a$  **approximate the vibrational frequencies measured in IR absorption spectroscopy**. The eigenvectors of the mass-weighted Hessian tell us how the atoms are displaced in a vibrational mode, and can therefore tell us if a given mode is a bond stretching mode, a bending mode, etc.
- The **sum** of the vibrational frequencies tell us information about the zero-point energy of a molecule. Accurate thermodynamic computations **must** include the zero-point energy to include nuclear vibrational effects and ultimately achieve good agreement with experiment. We will come back to this topic later.

An important thing to remember is that a harmonic vibrational analysis is meaningful only if:

1. The geometry used to compute the frequencies is **optimized**

2. The computation of the Hessian is **performed at the same level of theory**

For example, if you compute the geometry with the B3LYP density functional using the def2-TZVP basis, then a subsequent frequency computation **must be done at the same level of theory starting from the optimized geometry**.

## 4.4 Example: vibrational analysis of water

The example below shows the vibrational analysis of the water molecule using Hartree-Fock and the def2-TZVP basis. The vibrational analysis shows that six modes are classified as “low-frequency” because they are close to zero (post-proj low-frequency mode). Some of these numbers are imaginary, and that’s due to small numerical noise. These six low-frequency modes are trivial modes corresponding to three translations (along the  $x$ ,  $y$ , or  $z$  direction) and three rotations (around the  $x$ ,  $y$ , or  $z$  axis), hence the label “TR”.

Modes 7, 8, and 9 corresponds to three molecular vibrational modes. Starting from the highest frequency mode, we have

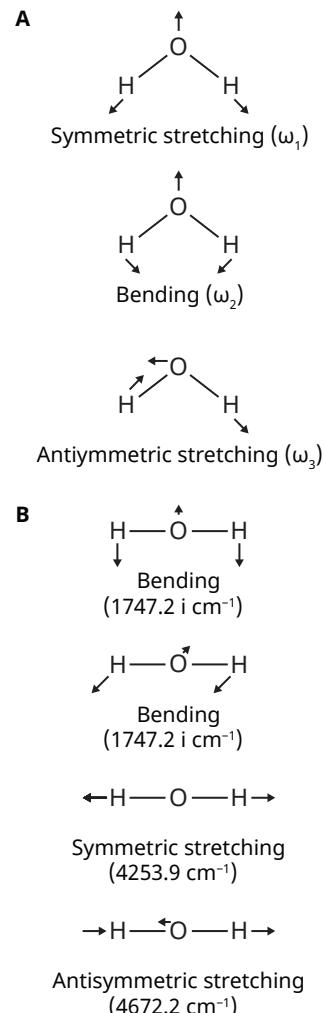
- Mode 9 ( $\omega_3 = 4213.5 \text{ cm}^{-1}$ ), an antisymmetric stretching of the two O–H bonds.
- Mode 8 ( $\omega_1 = 4111.5 \text{ cm}^{-1}$ ), a symmetric stretching of the two O–H bonds.
- Mode 7 ( $\omega_2 = 1734.7 \text{ cm}^{-1}$ ), a H–O–H bending mode.

Each normal mode has also an associated symmetry. The  $A_1$  modes correspond to **totally symmetric modes**. This means that if we transform the modes according to any of the symmetries of water, the normal mode does not change. Mode 9 is of type  $B_2$ , which means that if we rotate the molecule by  $180^\circ$  along the vertical axis that includes the oxygen atom, then the normal mode changes sign.

The calculation also tells us the IR activity (measured in  $\text{km/mol}$ ) of each normal mode. This information allows us to predict or relate our results to an experimental spectrum. All three normal modes are active, with mode 7 and 9 being the most intense.

```
=> Harmonic Vibrational Analysis <=
...
post-proj low-frequency mode: 0.0001i [cm^-1] (TR)
post-proj low-frequency mode: 0.0001i [cm^-1] (TR)
post-proj low-frequency mode: 0.0000i [cm^-1] (TR)
post-proj low-frequency mode: 0.0000i [cm^-1] (TR)
post-proj low-frequency mode: 0.0000i [cm^-1] (TR)
post-proj low-frequency mode: 0.0001 [cm^-1] (TR)

Vibration          7           8
Freq [cm^-1]      1734.7180   4111.5086
Irrep             A1          A1
Reduced mass [u]  1.0835     1.0444
Force const [mDyne/A] 1.9210   10.4022
Turning point v=0 [a0] 0.2531   0.1674
```



**Fig. 4.2:** Normal modes of water at the equilibrium geometry (**A**) and at the linear transition state (**B**) computed with Hartree-Fock theory and the def2-TZVP basis.

|                        |                   |                   |
|------------------------|-------------------|-------------------|
| RMS dev v=0 [a0 u^1/2] | 0.1863            | 0.1210            |
| IR activ [km/mol]      | 101.9411          | 17.0610           |
| Char temp [K]          | 2495.8730         | 5915.5454         |
| <hr/>                  |                   |                   |
| 1 O                    | -0.00 -0.00 -0.07 | 0.00 0.00 0.05    |
| 2 H                    | 0.00 0.42 0.56    | 0.00 0.59 -0.39   |
| 3 H                    | -0.00 -0.42 0.56  | -0.00 -0.59 -0.39 |
| <hr/>                  |                   |                   |
| Vibration              | 9                 |                   |
| Freq [cm^-1]           | 4213.5468         |                   |
| Irrep                  | B2                |                   |
| Reduced mass [u]       | 1.0839            |                   |
| Force const [mDyne/A]  | 11.3375           |                   |
| Turning point v=0 [a0] | 0.1624            |                   |
| RMS dev v=0 [a0 u^1/2] | 0.1195            |                   |
| IR activ [km/mol]      | 76.6343           |                   |
| Char temp [K]          | 6062.3558         |                   |
| <hr/>                  |                   |                   |
| 1 O                    | -0.00 -0.07 0.00  |                   |
| 2 H                    | -0.00 0.57 -0.42  |                   |
| 3 H                    | 0.00 0.57 0.42    |                   |

The information listed at the bottom shows how each atom moves in each of the normal modes. For each atom, each triplet of numbers under a normal mode indicate how much is the atom displaced in the  $x$ ,  $y$ , and  $z$  directions. This information can be used to analyze the nature of the normal modes in a molecule.

If we now repeat the same computation for linear water (after optimizing the geometry) we find seven low-frequency modes, two of which have a large imaginary frequency,  $\omega = 1747.2 \text{ i cm}^{-1}$ . The presence of these two imaginary modes tells us that linear water is a transition state. The two imaginary frequencies are two orthogonal bending modes that connect linear H–O–H to the bent ground state. If we were to follow these modes and optimize the geometry, we would obtain back the bent structure.

==> Harmonic Vibrational Analysis <==

```
...
post-proj low-frequency mode: 1747.1869i [cm^-1] (V)
post-proj low-frequency mode: 1747.1869i [cm^-1] (V)
post-proj low-frequency mode: 0.0000i [cm^-1] (TR)
post-proj low-frequency mode: 0.0000 [cm^-1] (TR)

Vibration          1          2
Freq [cm^-1]      1747.1869i     1747.1869i
Irrep
Reduced mass [u]   1.1259       1.1259
```

|                                    |           |           |
|------------------------------------|-----------|-----------|
| Force const [mDyne/ $\text{\AA}$ ] | -2.0250   | -2.0250   |
| Turning point v=0 [ $a_0$ ]        | 0.0000    | 0.0000    |
| RMS dev v=0 [ $a_0 u^{1/2}$ ]      | 0.0000    | 0.0000    |
| IR activ [km/mol]                  | 551.4304  | 551.4304  |
| Char temp [K]                      | 0.0000    | 0.0000    |
| <br>                               |           |           |
| Vibration                          | 8         | 9         |
| Freq [cm $^{-1}$ ]                 | 4253.9467 | 4672.2019 |
| Irrep                              | Ag        | B1u       |
| Reduced mass [u]                   | 1.0078    | 1.1259    |
| Force const [mDyne/ $\text{\AA}$ ] | 10.7453   | 14.4807   |
| Turning point v=0 [ $a_0$ ]        | 0.1676    | 0.1513    |
| RMS dev v=0 [ $a_0 u^{1/2}$ ]      | 0.1190    | 0.1135    |
| IR activ [km/mol]                  | 0.0000    | 829.5184  |
| Char temp [K]                      | 6120.4822 | 6722.2582 |

## Summary

- ☞ The potential energy surface near stationary points can be approximated by a quadratic function of the energy.
- ☞ The Hessian of the energy (the matrix of second partial derivatives) provides information about the curvature of the PES. Its eigenvalues allow us to distinguish between (local) minima (all eigenvalues positive) and transition states (one or more negative eigenvalues).
- ☞ The solutions of the nuclear Schrödinger equation near a stationary point (assuming a quadratic approximation to the energy) are called **normal modes**.
- ☞ Normal modes tell information about the nature of stationary points, IR spectroscopy, and the zero-point energy.



# 5 | Geometry optimization and transition state search

In this chapter we discuss some of the basic aspects of algorithms for geometry optimization. We will also take a look at the problem of finding transition states.

## 5.1 Geometry optimization algorithms

The main goal of geometry optimization algorithms is to find stationary points on a PES. Recall that at a stationary point the gradient of the energy with respect to coordinate  $R_i$ , which we indicate with  $g_i$ , is equal to zero

$$g_i = \left. \frac{\partial E(\mathbf{R})}{\partial R_i} \right|_{\mathbf{R}=\mathbf{R}^*} = 0, i = 1, \dots, 3N_{\text{atoms}} \quad (5.1)$$

This expression refers to all the  $3N_{\text{atoms}}$  **Cartesian coordinates**, which means that the  $R_i$  stands for any of the X, Y, or Z coordinates of the atoms.

How do geometry optimization algorithms find stationary points? Consider the potential energy curve of a diatomic molecule,  $E(r)$ , which depends on the bond distance  $r$ . Suppose that we have guessed that the bond length is going to be close to some value  $r_0$ . In the previous section we have seen that  $E(r)$  can be approximated by a Taylor series as

$$\begin{aligned} E(r) &= E(r_0) + E'(r_0)(r - r_0) + \frac{1}{2}E''(r_0)(r - r_0)^2 + \dots \\ &= E(r_0) + g(r - r_0) + \frac{1}{2}h(r - r^*)^2 + \dots \end{aligned} \quad (5.2)$$

where  $g = E'(r_0)$  and  $h = E''(r_0)$  are the gradient and hessian, respectively. Note that in this case we **are not at a stationary point**, so the gradient will not be zero,  $g \neq 0$ .

How can get closer to the stationary point? One of the simplest algorithms is the **Newton-Raphson method** (NR). Remember that at a stationary point  $r^*$ , the energy satisfies the condition  $E'(r^*) = 0$ . This suggests that we can try to find the value of  $r$  such that the first derivative of the **Taylor approximation** is equal to zero. If we take the derivative of the Taylor series of  $E(r)$  we get

$$E'(r) = g + h(r - r_0) + \dots \quad (5.3)$$

If we set this expression to zero and ignore the quadratic terms and higher (...), we obtain the condition

$$g + h(r - r_0) = 0 \Rightarrow r = r_0 - \frac{g}{h} \quad (5.4)$$

This equation tells us at what bond distance we will find the stationary point, **if the potential energy curve is quadratic**. In practice, the potential contains terms beyond quadratic, but if we are close enough to a stationary point they will only amount to a small correction. In this case we can repeat the Newton-Raphson procedure and recompute the gradients and Hessians and continue to update the geometry until the gradient is close enough to zero. As we get closer to the stationary point, cubic and higher terms in the Taylor expansion become smaller, and the Newton-Raphson procedure predicts a more geometry.

For a polyatomic molecule, this procedure can be easily generalized and it leads to the following rule to update the geometry

$$R_i = R_{0,i} - \sum_j^{3N_{\text{atoms}}} (H^{-1})_{ij} g_j \quad (5.5)$$

where  $(H^{-1})_{ij}$  is the inverse Hessian matrix, while  $R_i$  and  $R_{0,i}$  are the new and old values of the  $i$ -th coordinate, respectively.

Convergence of the energy optimization is determined by monitoring several variables.

- The energy change after one step in the optimization,  $\Delta E = E(\mathbf{R}) - E(\mathbf{R}_0)$ .
- The maximum force (the maximum value of  $g_i$ ) and the root mean square (RMS) force,  $\sqrt{\frac{1}{3N_{\text{atoms}}} \sum_i g_i^2}$ .
- The maximum displacement (the maximum value of  $R_i - R_{0,i}$ ) and the RMS deviation,  $\sqrt{\frac{1}{3N_{\text{atoms}}} \sum_i (R_i - R_{0,i})^2}$ .

The reason for using all these metrics is that we want to guarantee that both the energy and geometry are well optimized. In certain cases, for example, when the potential energy surface is nearly flat, the optimization steps only produce small changes in the energy but large displacements. In this case monitoring only the energy would prematurely stop the optimization.

Note that other optimization algorithms exist. When analytic energy gradients are not available, it is possible to optimize the energy using **derivative-free algorithms** like the Nelder-Mead method. Algorithms like **gradient descent** optimize the geometry along the direction of the gradient, but determine the step without the Hessian.

## 5.2 Analytical vs. numerical gradients and choice of the Hessian

Molecular optimization are most easily performed when the gradient can be computed by direct differentiation of the energy. For example consider the function

$f(x) = \exp(-x^2)$ . We can immediately compute the **analytic gradient** of  $f(x)$  by applying the rules of calculus

$$f'(x) = -2x \exp(-x^2) \quad (5.6)$$

However, suppose we did not know how to take the derivative of  $\exp$  or powers of  $x$ . In this case we can always apply the definition of the derivative

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \quad (5.7)$$

Since we cannot compute the limit numerically, we can always approximate it by **finite difference**, that is, by using a small value of  $h$

$$f'(x) \approx \frac{f(x+h) - f(x)}{h} + \text{error of order } h \quad (5.8)$$

The error we make using this formula is proportional to  $h$ , so if we make  $h$  sufficiently enough<sup>12</sup> we can produce an accurate approximation to  $f'(x)$ .<sup>13</sup> Analytic gradients are always preferred because one evaluation of the gradients costs a bit more than a single energy computation. Obtaining energy gradients via finite difference can be very expensive, because we need to evaluate the energy one or more times for each coordinate we optimize. This means that the relative cost of computing a gradient grows with the size of a molecule, and it is about  $3N_{\text{atoms}}$  that of a single energy computation.

Because computing the Hessian is quite expensive, geometry optimization often employs approximate or empirical Hessians. If a Hessian is not available—which is typically the default assumption of quantum chemistry codes—one may use an empirical hessian based on a ‘ball-and-sticks’ description of atoms in a molecule. If the Hessian of the initial geometry is available, this may be used to speed up the convergence of a geometry optimization, especially more challenging cases where empirical Hessians are incorrect. Programs may update the Hessian during the optimization process, or give the option to recompute the Hessian at each step (which is very expensive). Some times a good option is to start a geometry optimization with an Hessian computed at a very low level of theory. Since the Hessian provides information about the curvature of the PES, just getting a qualitatively accurate Hessian can be enough to improve the convergence in troubles cases and transition state searches.

## 5.3 Choice of coordinates

Geometry optimization using Cartesian coordinates is not the only possible choice. Any set of coordinates that fully span the same degrees of freedom of Cartesian coordinates ( $3N_{\text{atoms}}$ , linearly independent) may be used to perform a geometry optimization.

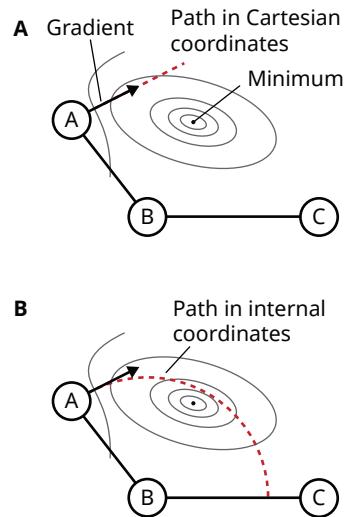
In particular, coordinates based on the bond distances and angles offer a very natural way to optimize the geometry of molecules. Some examples of internal coordinates are interatomic distances, angles, out-of-plane angles, and torsion angles.

<sup>12</sup>Here  $h$  must be still large enough to avoid rounding errors.

<sup>13</sup>There are better formulas for approximating the first derivative. For example, the

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}$$

introduces an error proportional to  $h^2$ , so if we use  $h = 0.001$ , the error will be of the order of  $10^{-6}$ .



**Fig. 5.1:** Comparison of optimization of the coordinates of atom A in the triatomic A-B-C in Cartesian (**A**) and internal coordinates (**B**). The red line shows the path in which the atom A can move in the two coordinate systems.

For example, **interatomic distance** between atoms A-B ( $r_{AB}$ ) is defined in terms of the Cartesian positions of atoms A and B [ $\mathbf{r}_A = (x_A, y_A, z_A)$ ]

$$r_{AB} = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2} = |\mathbf{r}_{AB}| \quad (5.9)$$

where  $\mathbf{r}_{AB} = \mathbf{r}_A - \mathbf{r}_B$ . Similarly, the **angle** between atoms A-B-C ( $\theta_{ABC}$ ) is defined as the angle between the vectors  $\mathbf{r}_{AB} = \mathbf{r}_A - \mathbf{r}_B$  and  $\mathbf{r}_{CB} = \mathbf{r}_C - \mathbf{r}_B$

$$\cos \theta_{ABC} = \frac{\mathbf{r}_{AB} \cdot \mathbf{r}_{CB}}{|\mathbf{r}_{AB}| |\mathbf{r}_{CB}|} \quad (5.10)$$

Optimization in internal coordinates requires the generation of a complete set of internal coordinates. Most codes employ **redundant internal coordinates** by including more coordinates than the number strictly needed. Redundant internal coordinates lead to faster convergence to stationary points, especially for more challenging systems with cyclic structure.

What is the advantage of using internal coordinates? As shown in Fig. 5.1, internal coordinates allow to step into curvilinear directions, which sometimes may offer a more rapid convergence. In this example, the gradient is the same, but the step taken by the optimizer is different depending on the coordinate system used. In Cartesian coordinates, atom A is displaced along a linear path that follows the direction of the gradient. This path brings the atom closer to the minimum, but it is not optimal. In contrast, when working internal coordinates, atom A follows a curvilinear path that corresponds to changing the angle A-B-C while keeping the distance A-B fixed. Stepping along this curved path can bring atom A closer to the minimum.

In which case it is convenient to use Cartesian coordinates? One example where internal coordinates may fail is when optimizing the structure of weakly interacting molecules. Since typical implementations automatically generate a set of internals, restricting the coordinates to atoms that are not too far apart, there might not be an internal coordinate that directly controls the distance or orientation between two molecules. When this happens, an optimization algorithm may take a long time to optimize a molecule. Cartesian coordinates work better in this case. It is also possible to mix Cartesian and internal coordinates. This set of coordinates can lead to more robust optimizations.

## 5.4 Potential energy scans and constrained optimization

We are often interested in mapping out the energy of a molecule as its geometry changes. In this case one often performs a series of energy computations in which a small set of geometric parameters (Cartesian or internal) is varied over a range. This type of computation is called a **potential energy scan**. Potential energy scans are often used to compute bond dissociation curves, to compute the intermolecular interaction energy between two molecules, or to study how the energy changes as a function of a change in conformation.

A straightforward potential energy scan of polyatomic molecules yields unrelaxed energies, by which we mean that the value of the coordinates that are kept fixed are not optimal. However, it is possible to relax all other coordinates during a

scan with a **constrained optimization**. For example, suppose you are interested in computing the energy of hydrogen peroxide (HOOH) as a function of the H-O-O-H dihedral angle  $\tau$  using the values 0, 5, 10, ..., 180°. As the dihedral angle changes, the optimal value of the O-H and O-O bond distances and the H-O-O angle will vary, so they need to be optimized. In a **relaxed potential energy scan** you would loop over each value of  $\tau$ , and for a given  $\tau$  perform an optimization of all other coordinates while keeping the dihedral fixed. Figure 5.2 shows the result of a scan of the H-O-O-H dihedral angle using relaxed and unrelaxed geometries. These energies are identical at the value of  $\tau$  that minimize the energy, but they differ for smaller and larger values of  $\tau$ . Most importantly, the unrelaxed energy is higher than the relaxed one, and the difference is as large as 1 kcal/mol when  $\tau = 0$ .

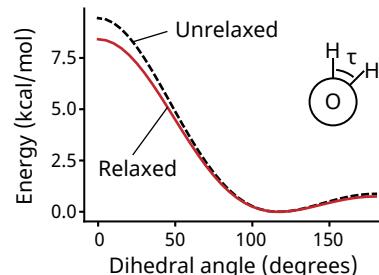
## 5.5 Transition state searches

Searching for a transition state (TS) is significantly more different than finding minima on the PES. There are several reasons that make transition state searches more difficult. Transition states are often significantly different than equilibrium molecular geometries and so are difficult to guess. There can also be multiple transition states that connect reactants to products making the search for the lowest energy transition state more difficult.

The easiest way to find a transition state is to have a very good guess of what its geometry looks like! This might be easy for computing the barrier to rotation in a molecule, or for simple reactions like SN2. When a geometry is sufficiently close to a transition state, and the Hessian has one negative eigenvalue that is well aligned with the reaction coordinate, then the Newton-Raphson method is likely to find the transition state. This is the case because we did not make assumption regarding the nature of the stationary point in the NR method. This approach to finding transition states is called a **local search**. It is mandatory in this case to compute an exact Hessian before performing the transition state search, because if the Hessian comes from a guess that incorrectly assumes all eigenvalues are positive, then the NR method could go in the direction of a minimum.

For relatively simple molecules, transition states may be found by performing a PES scan. For example, the scan shown in Fig. 5.2 has two transition states, that connect two minima on the potential energy surface of HOOH (approximately at ±110°). For certain reactions (e.g., SN2), one may be able to find the transition state via a scan of the energy as a function of one of the bond lengths.

**Semi-global** methods are more sophisticated transition state search algorithms that require the user to input a structure for the reactant and optionally that of the product. Some algorithms are open-ended, meaning that they explore the PES in search of a nearby transition state. Algorithms that use both the reactant and product geometry try to identify a path that connects these two structures and contains the transition state. There are a myriad of methods to find transitions states and the best way to learn how to use them is to read the details of the implementation in the manual of the computational program you plan to use.



**Fig. 5.2:** Hydrogen peroxide (HOOH). Potential energy scan as a function of the dihedral angle  $\tau$ . The relaxed curve is obtained by constraining the dihedral angle and optimizing all other coordinates. The unrelaxed curve is obtained by fixing the O-H and O-O bond lengths and the H-O-O bond angle at value that correspond to the minimum energy geometry.

## 5.6 Intrinsic reaction coordinate (IRC)

Once you have identified a transition state how do you know that it is the correct one? An intrinsic reaction coordinate is a path that starts at a transition state and has backward and forward branches that connect to minima. By computing the IRC, it is possible to verify that the transition state connects the reactant and product that we are studying.

The IRC is a path in coordinate space  $\mathbf{R}(s)$  parameterized by the variable  $s$ . This path satisfies the steepest descent differential equation

$$\frac{d\mathbf{R}(s)}{ds} = -\frac{\mathbf{g}(s)}{|\mathbf{g}(s)|} \quad (5.11)$$

where  $\mathbf{g}(s)$  is the gradient vector at the geometry  $\mathbf{R}(s)$ . Therefore, computing the IRC requires only the energy gradient.

Unfortunately there is not a unique way to defined the IRC, and the path you obtain depends on the set of coordinates used by the optimization algorithm. However, if you plan on doing an IRC computation to verify that you have found a transition state that connects the correct reactant and products, the differences in IRC path will not matter.

## Summary

- ☞ The most efficient geometry optimization algorithms are based on the Newton-Raphson method. They employ both the gradient and Hessian of the energy.
- ☞ Different coordinate sets may be used to perform the optimization. Internal coordinates perform better than Cartesian coordinates in most situations.
- ☞ Potential energy scans may be used to study the energy as a function of a geometric variable.
- ☞ Transition state algorithm are not black-box methods. Most of the time a successful TS search starts from a geometry that is very close to a transition state. An intrinsic reaction coordinate computation can be used to verify that a transition state connects to the correct reactant and product.

# 6 | Basis Sets

In the previous chapter we introduced the Hartree–Fock method. The goal of this chapter is to discuss some practical aspects of how Hartree–Fock computations actually run on a computer and the way they are approximated. We will encounter the idea of using a linear combination of basis functions to approximate the orbitals  $\phi(\mathbf{r})$  in systematically improvable way. Lastly, we will look at some applications of these ideas.

## 6.1 Exact solutions to the Schrödinger equation for the hydrogen atom

As it was mentioned earlier, we know exact solutions of the Schrödinger only for a handful of models. An example that should be familiar to you is the hydrogen atom. For this system, the solutions are known to depend on three quantum numbers  $n$ ,  $l$ , and  $m_l$ , and that the energy levels (eigenvalues) are given by (in atomic units)

$$E_n = -\frac{1}{2n^2} E_h, \quad n = 1, 2, \dots \quad (6.1)$$

The wave function  $\psi_{nlm_l}$  is usually expressed in **spherical coordinates**  $(r, \theta, \varphi)^{14}$  as a product of a function that depends on the electron distance  $R_{nl}(r)$  and a term that depends on the angles  $\theta$  and  $\varphi$ , which we write as  $Y_l^{m_l}(\theta, \varphi)$

$$\psi_{nlm_l}(r, \theta, \varphi) = R_{nl}(r)Y_l^{m_l}(\theta, \varphi) \quad (6.2)$$

For example, the wave function for the ground state of hydrogen, the 1s orbital, is given by

$$1s = \psi_{1,0,0}(r, \theta, \varphi) = R_{1,0}(r)Y_0^0(\theta, \varphi) = 2e^{-r} \frac{1}{\sqrt{4\pi}} = \frac{1}{\sqrt{\pi}} e^{-r} \quad (6.3)$$

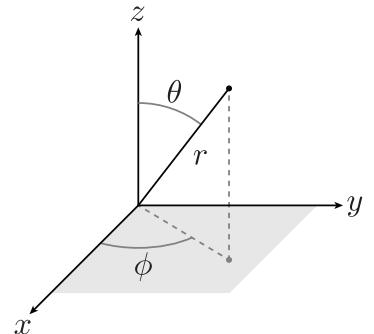
while the 2p<sub>z</sub> orbital is given by

$$2p_z = \psi_{2,1,0}(r, \theta, \varphi) = R_{2,1}(r)Y_1^0(\theta, \varphi) = \frac{1}{4\sqrt{2\pi}} r e^{-r/2} \cos \theta. \quad (6.4)$$

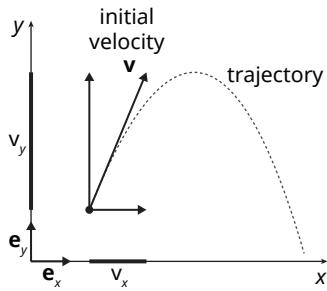
When we turn to more complicated systems, like atoms with more than one electron and molecules, the Schrödinger equation becomes intractable and we have to turn to approximate numerical methods to solve it (recall Dirac's quote on this point). This is what we will look into in the next section.

<sup>14</sup> If you are curious, spherical coordinates are connected to Cartesian coordinates via the following equations

$$\begin{aligned} x &= r \sin \theta \cos \varphi, \\ y &= r \sin \theta \sin \varphi, \\ z &= r \cos \theta \end{aligned}$$



**Fig. 6.1:** Definition of the spherical coordinates  $r$ ,  $\theta$ , and  $\varphi$ .



**Fig. 6.2:** The trajectory of a particle with initial velocity vector  $\mathbf{v} = (v_x, v_y)$ . This vector may be decomposed in the basis of orthogonal unit vectors  $\mathbf{e}_x$  and  $\mathbf{e}_y$  as  $\mathbf{v} = v_x \mathbf{e}_x + v_y \mathbf{e}_y$ .

<sup>15</sup> A linear combination of vectors/functions/etc. ( $f_i$ ) is a sum

$$f_1 c_1 + f_2 c_2 + \dots + f_k c_k$$

where the quantities  $c_1, c_2, \dots$  are numbers (they can be real or complex).

<sup>16</sup> A basis can have orthogonal or nonorthogonal vectors. However, bases that are orthogonal are typically preferred.

One way to think about this is to discretize the integral as a sum. The integral of a function  $f(x)$  is the limit of  $\Delta x \rightarrow 0$  of the sum

$$\int_a^b dx f(x) = \Delta x \sum_{k=1}^n f(x_k)$$

where the points  $x_k$  form a grid with spacing  $\Delta x$ . The integral of a product of two functions  $u^*(x)v(x)$  is then also given by the limit of the following expression

$$\int_a^b dx u^*(x)v(x) = \Delta x \sum_{k=1}^n u^*(x_k)v(x_k)$$

If we identify the value of the functions at the points  $x_k$  as the components of a vector, for example,  $v(x_1) = v_1, v(x_2) = v_2, \dots$ , then the above expression can be written as

$$\int_a^b dx u^*(x)v(x) = \Delta x [u_1 v_1 + u_2 v_2 + \dots]$$

This looks very similar to the definition of the dot product of two vectors of arbitrary dimension ( $\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \dots$ ).

## 6.2 Bases, basis vectors, and basis functions

An important concept when constructing numerical approximations is that of a basis. To build some intuition about the concept of a basis you can think back to your introductory physics courses and Newton's equations. Consider the problem of describing the trajectory of a particle under the effect of gravity. To solve this problem you need to specify the position and initial velocity of the particle. The position can be specified by the  $x$  and  $y$  coordinates, and the initial velocity by the velocity in  $x$  and  $y$  directions ( $v_x, v_y$ ). These quantities are essentially vectors, which we indicate with the symbols  $\mathbf{r} = (x, y)$  and  $\mathbf{v} = (v_x, v_y)$ . Figure 6.2 shows the position and velocity vectors.

Every vector can be written as a **linear combination**<sup>15</sup> of vectors that are linearly independent and span the full space, which we call a **basis**. For example, the velocity vector  $\mathbf{v}$  can be expressed as a sum of a component pointing in the  $x$  direction ( $\mathbf{e}_x$ ) and one pointing in the  $y$  direction ( $\mathbf{e}_y$ )

$$\mathbf{v} = v_x \mathbf{e}_x + v_y \mathbf{e}_y, \text{ with } \mathbf{e}_x = (1, 0) \text{ and } \mathbf{e}_y = (0, 1) \quad (6.5)$$

The vectors  $\mathbf{e}_x$  and  $\mathbf{e}_y$  are called a basis, because any vector in two dimensions  $\mathbf{a}$  can be written as a sum of coefficients multiplied by the basis vectors

$$\mathbf{a} = a_x \mathbf{e}_x + a_y \mathbf{e}_y \quad (6.6)$$

The vectors  $\mathbf{e}_x$  and  $\mathbf{e}_y$  have another special property: they are orthogonal, which means that their dot product is zero<sup>16</sup>

$$\mathbf{e}_x \cdot \mathbf{e}_y = 0. \quad (6.7)$$

In the more general case, the velocity is a three dimensional vector, and must be described as a sum of three terms

$$\mathbf{v} = v_x \mathbf{e}_x + v_y \mathbf{e}_y + v_z \mathbf{e}_z \quad (6.8)$$

Interestingly, this idea can also be extended to functions. For example, suppose we want to approximate some function  $g(x)$ . We could start from choosing two simple functions  $f_1(x)$  and  $f_2(x)$  and approximate  $g(x)$  as

$$g(x) = c_1 f_1(x) + c_2 f_2(x) \quad (6.9)$$

We can think of this equation as being analogous to taking a combination of basis vectors, with the difference being that we are combining the functions  $f_1(x)$  and  $f_2(x)$ . This is not just an analogy. We use the same language because you can think of a function as a vector with infinite dimensions and most of the concepts that apply to vectors (like length, being orthogonal, etc.) can be generalized to functions.

When “fitting” the function  $g(x)$  in this way, it is convenient to work with functions that are **orthogonal**. But how does one define the idea of orthogonality for functions? A natural generalization is the overlap integral ( $\langle f_1 | f_2 \rangle$ )<sup>17</sup>

$$\langle f_1 | f_2 \rangle = \int_{-\infty}^{\infty} dx f_1^*(x) f_2(x) \quad (6.10)$$

From this definition, we define two functions to be orthogonal if their dot product (inner product) is null,  $\langle f_1 | f_2 \rangle = 0$ .

## 6.3 Approximate solutions via expansion in a basis

We can now look into how the idea discussed in the previous section can be applied to approximate solutions of the Schrödinger equation. The main point is to use a **linear combination of basis functions** to approximate the wave function. This is a very common strategy for studying differential equations that are too difficult to solve in analytical form. In Eq. (6.9) we considered the case of a basis composed by two functions. However, in the more general case, we approximate  $g(x)$  as a sum of a finite number ( $K$ ) of fixed functions  $f_i(x)$ , called basis functions. Each basis function is multiplied by the coefficients  $c_i$  and the approximation is given by

$$g(x) \approx c_1 f_1(x) + c_2 f_2(x) + \cdots + c_K f_K(x) = \sum_{i=1}^K c_i f_i(x) \quad (6.11)$$

You can think of this approximation as fitting the function  $g(x)$  with the basis functions, where the unknowns are the fitting coefficients  $c_i$ . An important point is that, if the basis  $\{f_i(x)\}$  is chosen judiciously, the approximate solution should converge rapidly to the exact  $g(x)$  when the number of basis functions ( $K$ ) is increased.

This process is illustrated in Fig. 6.3, where the function  $\exp(x) \sin(x)$  is approximated with a basis of polynomials of  $x$ . In this example, we stop at the polynomial of order  $K$  (here we also include the constant term  $x^0 = 1$ )

$$g(x) \approx c_0 + c_1 x + \cdots + c_K x^K = \sum_{i=0}^K c_i x^i \quad (6.12)$$

As you can see, the first two approximations are not very good, and by the time we use a forth-order polynomial we cannot easily distinguish  $\exp(x) \sin(x)$  from its approximation.

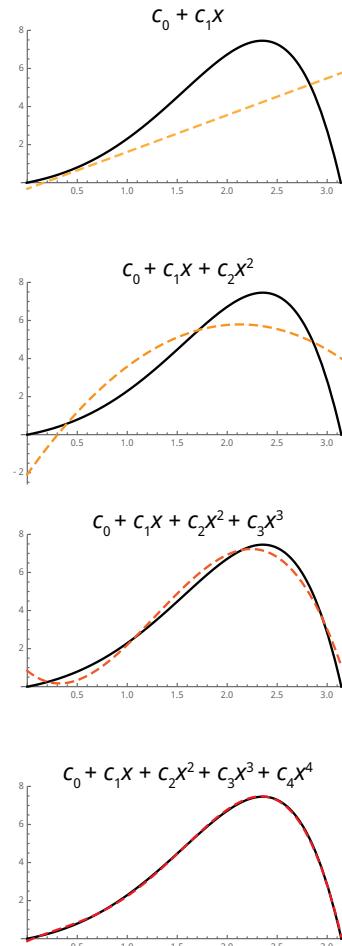
## 6.4 Gaussian-type orbitals

To understand how we use bases in quantum chemistry we will start by learning how atomic orbitals are approximated with a Gaussian basis. Since the early days of computational chemistry it has become common practice to approximate atomic orbitals using a basis of **Gaussian functions**. For example, if we want to approximate the  $1s$  orbital of the hydrogen atom,  $1s = \frac{1}{\sqrt{\pi}} e^{-r}$ , an expansion in Gaussian functions consists of the following approximation

$$\frac{1}{\sqrt{\pi}} e^{-r} \approx \sum_{\mu} C_{\mu} \exp(-a_{\mu} r^2) \quad (6.13)$$

At this point we should ask: **Why would we want to approximate a function that we already know how to write in closed form?** The answer is that all quantum chemistry methods need to know integrals of atomic orbitals like

$$\int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\phi_i^*(\mathbf{r}_1) \phi_j^*(\mathbf{r}_2) \phi_k(\mathbf{r}_1) \phi_l(\mathbf{r}_2)}{r_{12}} \quad (6.14)$$



**Fig. 6.3:** Example of expansion of the function  $\exp(x) \sin(x)$  in a basis of polynomials  $x^i$ . By the time we include fourth powers of  $x$  this function is approximated well in the entire range  $0 \leq x \leq \pi$ .

<sup>18</sup>This is due to the Gaussian product theorem, which states that the product of two multidimensional Gaussians is still a Gaussian function.

These integrals are difficult to compute with function of the form  $e^{-ar}$ , but they turn out to be doable if one instead uses functions of the form  $\exp(-ar^2)$ .<sup>18</sup>

For each type of atomic orbital (s, p, d, etc.) one can write a corresponding **Gaussian-type orbital** obtained by multiplying a 3D spherical Gaussian function  $\exp(-ar^2)$  times a polynomial in the coordinates  $x, y$ , and  $z$

$$x^l y^m z^n e^{-ar^2} \quad (6.15)$$

As we have seen, s orbitals are obtained from a combination of Gaussians

$$e^{-ar^2} \quad (6.16)$$

while p orbitals can be represented with a product of  $x, y$ , or  $z$  and a Gaussian

$$xe^{-ar^2}, ye^{-ar^2}, ze^{-ar^2} \quad (6.17)$$

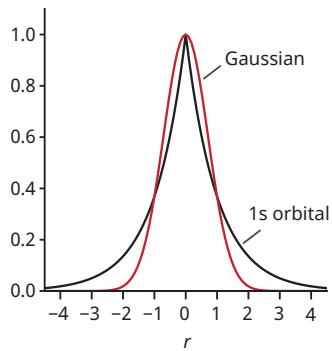
Atomic orbitals of d type can be formed by quadratic polynomials times a Gaussian. In this case there are six combinations

$$xye^{-ar^2}, yze^{-ar^2}, xze^{-ar^2}, x^2 e^{-ar^2}, y^2 e^{-ar^2}, z^2 e^{-ar^2} \quad (6.18)$$

We call these 6 d orbitals **Cartesian** d functions, to distinguish them from the 5 **spherical** or **pure angular momentum** d functions that are obtained when we solve the Schrödinger equation

$$xye^{-ar^2}, yze^{-ar^2}, xze^{-ar^2}, (x^2 - y^2)e^{-ar^2}, (2z^2 - x^2 - y^2)e^{-ar^2} \quad (6.19)$$

Similarly, there are more Cartesian f functions (10) than spherical f function (7), etc. In a quantum chemistry code, integrals are first computed using Cartesian and then these are combined to form spherical functions. Some codes might allow you to use Cartesian or spherical functions, but you should always use the default.



**Fig. 6.4:** Plot of the STO-6G Gaussian basis set and the six Gaussian primitives for the hydrogen atom.

## 6.5 Contracted Gaussian-Type Orbitals

The Gaussians functions defined in Eq. (6.15) are also called **primitives**. Figure 6.4 reports a comparison of the shape of the hydrogen 1s orbital and a 1s Gaussian primitive. As you notice, the Gaussian and the 1s orbital are both peaked at the origin. However, the Gaussian has a rounded top at  $r = 0$  and decays too quickly when  $r > 1$ . To represent the correct shape of atomic orbitals several primitives are combined together using a fixed set of coefficients to form numerically accurate approximations to atomic orbitals. We call these function **contracted Gaussians** and to avoid confusing them for orbitals, they will be denoted with the Greek letter chi ( $\chi$ ). For example, a  $p_z$  contracted Gaussian centered at the origin is approximated as

$$\chi_{p_z}(\mathbf{r}) = \sum_{\mu} b_{\mu} z \exp(-a_{\mu} r^2) \quad (6.20)$$

The quantities  $a_{\mu}$  are called the **exponents** of the basis, while the coefficients  $b_{\mu}$  are called **contraction coefficients**.

A **Gaussian basis set** is a collection of contracted Gaussian functions with optimized exponents and contraction coefficients. Making a basis sets is not trivial and there are many articles in the literature describing how to build basis set. The most commonly used basis sets are collected in the **basis set exchange** on-line repository (<https://www.basissetexchange.org>).

It is instructive to take a look at an example of a Gaussian basis used by a computer program. The following shows the **STO-6G** basis set for the hydrogen atom. This basis approximates one atomic orbital of hydrogen using six Gaussians (hence the “6G”). This basis includes only one atomic orbital for each hydrogen. The first line lists the element name, while the second one is a tag that indicates that the next few lines show the exponent and contraction coefficients for a s-type (“S”) orbital approximated with 6 Gaussians. The left column shows the exponent of the Gaussians, while the second column the corresponding contraction coefficients

```
H 0
S   6   1.00
      0.3552322122E+02      0.9163596281E-02
      0.6513143725E+01      0.4936149294E-01
      0.1822142904E+01      0.1685383049E+00
      0.6259552659E+00      0.3705627997E+00
      0.2430767471E+00      0.4164915298E+00
      0.1001124280E+00      0.1303340841E+00
```

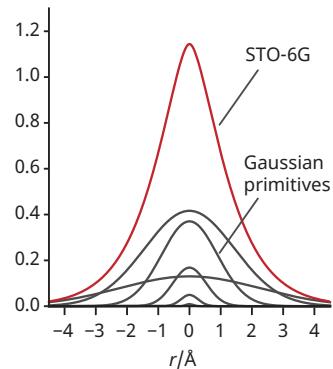
This basis corresponds to the function

$$\chi(\mathbf{r}) = 0.009163596281 \exp(-35.52322122 r^2) + \dots \quad (6.21)$$

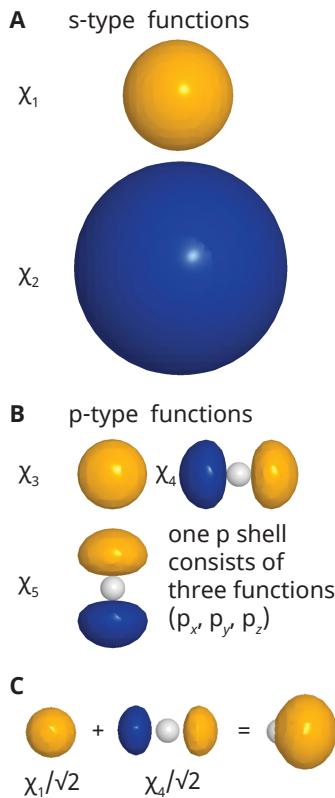
which is plotted in Fig. 6.5. This figure also shows the six Gaussian primitives that are combined together to obtain the STO-6G basis function. The STO-6G function looks like a Gaussian but it is more pointy at the origin and decays more slowly for large values of  $r$ . **Even though the STO-6G basis uses six Gaussians, this basis consists of only one function.**

The STO-6G basis is not a very flexible basis, since it does not allow the orbital of the hydrogen atom to shrink, grow, or polarize. For example, if we consider the H<sup>-</sup> anion, the STO-6G basis does not allow the occupied orbital to expand due to the increased electron-electron repulsion experienced due to the extra negative charge. In practice, computation basis include more functions than the strict valence atomic orbitals. For example, the def2-SVP basis for hydrogen is defined as

```
H      0
S   3   1.00
      13.0107010      0.19682158E-01
      1.9622572      0.13796524
      0.44453796      0.47831935
S   1   1.00
      0.12194962      1.0000000
P   1   1.00
      0.8000000      1.0000000
```



**Fig. 6.5:** Plot of the STO-6G Gaussian basis set and the six Gaussian primitives for the hydrogen atom.



**Fig. 6.6:** Plot of the def2-SVP Gaussian basis set for the hydrogen atom. (A) split-valence s-type functions and (B) p-type polarization functions. The combination of a s-type and p-type function is a s-like function polarized along the y direction (C)

This basis contains two s-type contracted Gaussians (SV = split valence, meaning two sets of valence orbitals) plus a set of three p orbitals ( $p_x, p_y, p_z$ ), which are called **polarization functions** (P = polarization). Figure 6.6 shows these five basis functions. By combining orbitals  $\chi_1$  and  $\chi_2$  it is possible to form spherical orbitals of different size. The p orbitals instead allow to polarize the wave function along any direction in space as shown in part C of this figure.

Gaussian basis sets can be built in such a way that the accuracy of the orbitals can be systematically improved. The following table shows the energy of the hydrogen atom computed with a family of basis sets (cc-pVXZ, with  $X = D, T, Q, 5, 6$ ) that contains an increasing number of Gaussian functions (this basis contains more than just s functions). As you can see, the cc-pVDZ basis is already very accurate (the absolute energy error is less than 0.001 mE<sub>h</sub>) and the error can be reduced by almost one order of magnitude as we go from one basis to next one in this family.

Table 6.1: Convergence of the energy of the hydrogen atom as a function of the computational basis. The exact solution corresponds to the analytic solution of the Schrödinger equation ( $E = -0.5 E_h$ ).

| Basis   | Basis Functions | Energy ( $E_h$ ) | Error ( $E_h$ ) |
|---------|-----------------|------------------|-----------------|
| cc-pVDZ | 5               | -0.499278403     | 0.000721597     |
| cc-pVTZ | 14              | -0.499809811     | 0.000190189     |
| cc-pVQZ | 30              | -0.499945569     | 0.000054431     |
| cc-pV5Z | 55              | -0.499994535     | 0.000005465     |
| cc-pV6Z | 91              | -0.499999245     | 0.000000755     |
| Exact   | $\infty$        | -0.500000000     | 0.000000000     |

## 6.6 The linear combination of atomic orbitals approximation

The Gaussian basis functions introduced in the previous section are the building blocks used to construct approximate orbitals (atomic and molecular). A sensible choice for approximating the shape of molecular orbitals is to use a basis that is as close as possible to the atomic orbitals of isolated atoms. The intuition here is that a lot of the details of the structure of atoms is preserved in molecules. So it makes sense to use atomic orbitals as building blocks for approximating molecular orbitals.

This idea is called the **linear combination of atomic orbitals** (LCAO). You have already encountered LCAO when you applied molecular orbital theory to simple molecules like H<sub>2</sub>, where the bonding orbital is the in-phase combination (superposition) of two 1s atomic orbitals, or when you were introduced to the concept of hybrid orbitals like the sp<sup>3</sup> hybrids of carbon in methane. This is not the only way quantum chemistry computations are done,<sup>19</sup> but it is by far the most common approach used in most quantum chemistry programs.

Mathematically, the LCAO approximation expresses a generic orbital  $\phi_i$  as a linear combination of Gaussian basis functions  $\chi_\mu$ , so that

$$\phi_i(\mathbf{r}) \approx \sum_{\mu} C_{\mu}^{(i)} \chi_{\mu}(\mathbf{r}) \quad (6.22)$$

<sup>19</sup>Other examples include using plane waves [ $\exp(i\mathbf{k} \cdot \mathbf{r})$ ] and numerical grids to approximate the orbitals.

The quantity  $C_\mu^{(j)}$  is a vector of coefficients for orbital  $\phi_j$ . **When you run a Hartree-Fock computation, this is the quantity that is optimized by the SCF procedure.**

For very simple molecules, the coefficients  $C_\mu^{(j)}$  are simply determined by symmetry. As an example, we will look at the orbitals of the H<sub>2</sub> molecule using the STO-6G basis. For convenience, we will label the two hydrogen atoms 1 and 2. Each hydrogen atom carries one basis function, which we label as  $\chi_1$  and  $\chi_2$ . These functions are shown in Fig. 6.7.

From these basis functions we can form two atomic orbitals. The first one is a in-phase combination (g, from the German word *gerade*) of  $\chi_1$  and  $\chi_2$ , times a normalization factor  $N_g$

$$\phi_g = N_g(\chi_1 + \chi_2), \text{ which corresponds to } C_1^{(1)} = N_g, C_2^{(1)} = N_g \quad (6.23)$$

The second orbital is the out-of-phase combination (u, *ungerade*)

$$\phi_u = N_u(\chi_1 - \chi_2), \text{ which corresponds to } C_1^{(2)} = N_u, C_2^{(2)} = -N_u \quad (6.24)$$

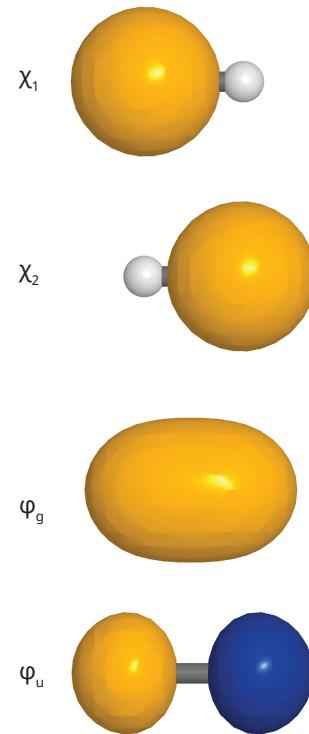
These are the orbitals that you will get if you do a Hartree-Fock computation on H<sub>2</sub>. The orbital  $\phi_g$  is a **bonding orbital** while  $\phi_u$  is an **antibonding orbital**. Since  $\phi_g$  has no nodes, its energy is lower than that of  $\phi_u$ .

## 6.7 A survey of common basis sets

The basis set exchange website counts a total of 589 basis sets! So how do you go about choosing the right basis set for your computation? In choosing the right basis set we always have to balance two competing factors: the accuracy of the result and the cost of a computation. When one is willing to tolerate a given average error, it is possible to choose a basis using previous results in the literature and use it in a study. However, if your goal is to get the best possible result, the it might be necessary to use a range of basis sets. By repeating computations using basis sets of increasing size, we can gauge the so-called **basis set incompleteness error**, the error that arises from approximating the orbitals using a finite basis set.

For computations on large molecules with Hartree-Fock or DFT it is customary to employ a fixed basis set. Basis sets like Pople's 6-31G, 6-311G, 6-31G\*,<sup>20</sup> 6-31G\*\*, etc. are often used in DFT computations on large molecules. However, I would discourage you from using this family of basis sets because it is now obsolete and some serious problems have been discovered (e.g., when used with the MP2 method this basis gives a nonplanar structure for benzene!).

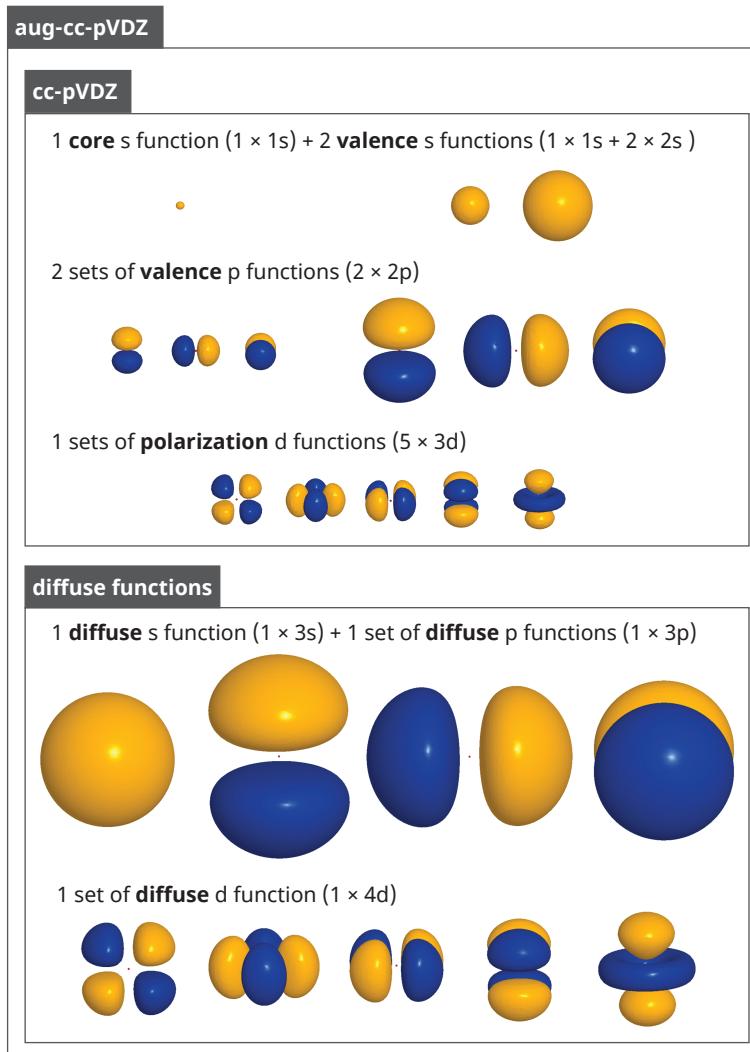
A more modern choice is the family of Karlsruhe basis sets (def2-SVP, def2-TZVP, def2-QZVP, def2-TZVPD, ...). For routine DFT computations, a def2-TZVP basis should provide nearly-converged results. The "TZV" and "QZV" stand for triple zeta and quadruple zeta valence (for the Greek letter zeta,  $\zeta$ , often used to indicate the exponent of a Gaussian instead of the letter  $\alpha$  used in these notes), that is three or four sets of valence orbitals. "P" and "D" indicate polarization and **diffuse functions**, respectively. Diffuse functions are extra Gaussian basis functions with a small exponent, which are added to a basis to give it more flexibility in the description of



**Fig. 6.7:** Hydrogen molecule. Plot of the STO-6G basis functions ( $\chi_1$  and  $\chi_2$ ) and bonding (gerade,  $\phi_g$ ) and antibonding (ungerade,  $\phi_u$ ) molecular orbitals.

<sup>20</sup>The 6-31G\* basis set is similar to the def2-SVP basis. It is defined for the atoms H through Zn. It uses two sets of valence orbitals plus polarization functions (d-type polarization functions on atoms Li-Ca and f-type polarization functions on atoms Sc-Zn).

orbitals that are diffuse. These functions increase the accuracy of computations on anions, long-range interactions, and excited states.



**Fig. 6.8:** Oxygen atom. Plot of the cc-pVDZ and aug-cc-pVDZ contracted Gaussian-type basis functions. The aug-cc-pVDZ basis is built from the cc-pVDZ by adding an extra set of diffuse functions.

For wave function methods like MP2, CCSD, CCSD(T), etc., Dunning's correlation-consistent basis sets (cc-pV $X$ Z, with  $X = D, T, Q, 5, 6$ ) are recommended and represent the state-of-the-art. These are families of basis sets that systematically converge to the full basis set limit. Figure 6.8 shows the basis functions included in the cc-pVDZ basis for oxygen. This basis includes a double set of valence orbitals (2s and 2p) and adds one set of d polarization functions. The cc-pVDZ basis is likely to provide reasonable, but potentially inaccurate results. A cc-pVTZ basis is generally recommended

for routine applications. For highly-accurate benchmark results it is necessary to use the cc-pVQZ or cc-pV5Z basis sets. There are many variants of the Dunning's correlation-consistent basis sets.

- The basis sets aug-cc-pVXZ include **diffuse functions** and are useful to describe Rydberg states, long range interactions, and anions. Figure 6.8 shows the extra diffuse functions included in the aug-cc-pVDZ basis.
- Dunning's basis sets (cc-pVXZ) are meant to correlate only valence electrons, so core electrons should be frozen in computations that use them. To account for core electron correlation it is necessary to use a **core-valence basis set**, usually indicated with the symbol cc-pCVXZ, or the newer weighted core-valence sets (cc-pwCVXZ).
- Other variants of Dunning's basis sets include **basis extended for period 3 atoms** [cc-pV(X+d)Z], **pseudopotential basis sets** for heavy elements [cc-pVXZ-PP], and **basis sets for relativistic computations** [cc-pVXZ-DK].

This is by no means a comprehensive overview of basis sets. A good practice is to use the literature as a starting point for getting a feeling for what basis sets are used by other researchers. Ideally, before working on a project you should benchmark the convergence of the energy, the molecular geometry, or other properties that you are interested in computing using different basis sets.

Note, that **it is common practice in Quantum Chemistry to cite the original reference to the basis sets employed in a computation**. Crafting a basis set for a large number of elements in the periodic table takes a lot of work and the people that invested their time doing so deserve credit for it.

## 6.8 The complete basis set (CBS) limit and the basis set error

To analyze the error introduced by using a finite basis we define the **complete basis set (CBS) limit** as the limit of the energy (or any other property) for the basis size that goes to infinity.

$$E_{\text{CBS}} = \lim_{K \rightarrow \infty} E(\text{basis of size } K) \quad (6.25)$$

The underlying assumption is that with a **complete basis** we can approximate any well-behaved function. It is convenient to introduce the idea of the CBS limit and not just call this energy exact, because even if we use a complete basis, we are still introducing other sources of error that will make our results deviate from experiment.

The error due to the finite basis set is called the **basis set incompleteness error** (BSIE). For any method, the BSIE can be formally defined as the difference between the energy computed with a finite basis  $E_{\text{finite basis}}$  and the CBS limit ( $E_{\text{CBS}}$ )

$$\Delta E_{\text{BSIE}} = E_{\text{finite basis}} - E_{\text{CBS}} \quad (6.26)$$

We will now take a look at the convergence of the Hartree-Fock energy as a function of the basis set for the  $\text{H}_2^+$  molecule (see Table 6.2). This system has only one electron and, therefore, a Hartree-Fock computation is only affected by the error

that arises from using a finite basis set. In this case the energy converges rapidly, like in the case of the hydrogen atom. However, the rate at which the energy converges is slower. For example, the change in energy when going from the cc-pVDZ to the cc-pVTZ basis is only  $0.53 \text{ m}E_h$  for H, but this number is  $1.98 \text{ m}E_h$  for  $\text{H}_2^+$ . Comparing the energy to the best result for  $\text{H}_2^+$  (cc-pV6Z basis), we conclude that cc-pVTZ is already sufficiently accurate to make quantitative predictions, and the cc-pVQZ energy is only  $0.1 \text{ m}E_h$  (ca. 0.15 kcal/mol) off from the best result.

| Basis   | Basis Functions | Energy ( $E_h$ ) |
|---------|-----------------|------------------|
| cc-pVDZ | 10              | $-0.600264667$   |
| cc-pVTZ | 28              | $-0.602244426$   |
| cc-pVQZ | 60              | $-0.602520583$   |
| cc-pV5Z | 110             | $-0.602619758$   |
| cc-pV6Z | 182             | $-0.602632085$   |

Table 6.2: Convergence of the energy of the  $\text{H}_2^+$  molecule at a bond distance  $r_{\text{HH}} = 2$  bohr as a function of the computational basis.

# 7 | Density Functional Theory

In the previous chapters, we discussed the Hartree–Fock method. This approximation is insufficient to make accurate predictions of molecular properties because it neglects **electron correlation** effects. There are two main ways to include correlation effects in quantum chemistry computations: 1) density functional theory and 2) correlated wave function methods. In this chapter we will discuss density functional theory (DFT), an approach that in the past several decades has revolutionized the field of quantum chemistry. DFT is an approximate method, but in many cases it yields results of sufficient accuracy to make reliable predictions.

## 7.1 Going beyond Hartree–Fock theory: Electron correlation

From a classical point of view, electrons experience Coulomb repulsion, and consequently one would expect that their motion is influenced by their mutual interactions. From a probabilist point of view, this means that it is more likely that two electrons are found at a large distance than at a shorter one. In this case we say that the probability distribution of the electrons is **correlated**. From the point of view of quantum mechanics, this means that the probability of finding the electrons in positions/spin equal to  $x_1, x_2, \dots$ , which we indicate with  $P(x_1, x_2, \dots)$ , cannot be written as a product of individual electron probabilities

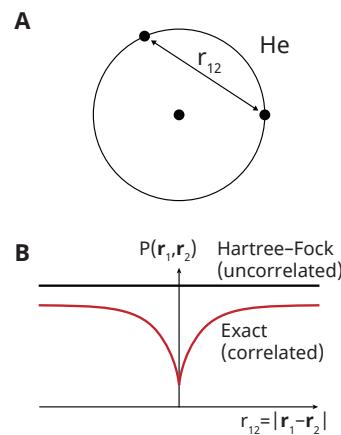
$$P(x_1, x_2, \dots) = |\Psi(x_1, x_2, \dots)|^2 \neq P_1(x_1)P_2(x_2) \cdots \quad (7.1)$$

The Hartree–Fock method does fully account for electron correlation.<sup>21</sup> This can be seen in Fig. 7.1, where if we plot the probability of finding the electrons in the helium atom as a function of their distance—while keeping their distance from the nucleus fixed—we find out that the Hartree–Fock wave function is a constant. This is because for the case of Hartree–Fock the probability of finding the two electrons is the product of two independent probabilities. Figure 7.1 also shows the exact wave function, which displays a dip in the probability due to Coulomb repulsion.

When doing quantum chemistry computations, we are mostly concerned with the effects of electron correlation on the energy. Therefore, we define the electron correlation energy ( $E_{\text{corr}}$ ) as the difference between the exact and the Hartree–Fock energies

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}} \quad (7.2)$$

Because of the variational principle, the **correlation energy is always negative**.



**Fig. 7.1:** Two electrons in the helium atom at equal distance from the nucleus (A). The probability of finding two electrons [ $P(\mathbf{r}_1, \mathbf{r}_2)$ ] at a given radius as a function of their distance  $r_{12}$  (B).

<sup>21</sup>Hartree–Fock theory includes a special type of correlation for electrons with same spin, called Fermi correlation.

## 7.2 Theoretical foundations of density functional theory

One of the simplest way to include the effects of electron correlation is via density functional theory (DFT). The basic idea of DFT is to avoid solving the  $N$ -electron Schrödinger equation, which requires us to determine the wave function  $\Psi(x_1, x_2, \dots)$ . Instead, DFT tries to capture the properties of the ground state only using the **electron density**  $\rho(x)$ . This quantity is derived from the full electron wave function by squaring and integrating over all particle coordinates except for the first coordinate ( $x_1$ )

$$\rho(x_1) = N \int dx_2 dx_3 \dots |\Psi(x_1, x_2, \dots)|^2 \quad (7.3)$$

Since  $\rho(x_1)$  is a function of 3 spatial coordinates and one spin coordinate, it is much easier to represent and compute.

DFT was put on solid theoretical grounds by the work of Hohenberg and Kohn (1964). These authors established the existence of a universal **functional** of the density  $E[\rho]$ , such that the energy of the ground state can be obtained by minimizing  $E[\rho]$  with respect to the density

$$\min_{\rho} E[\rho] = E_{\text{exact}} \quad (7.4)$$

As shown in Fig. 7.2, functional  $F[f]$  is a mathematical operation that converts a function  $f(x)$  into a number

$$F : F[f(x)] \rightarrow \mathbb{R} \quad (7.5)$$

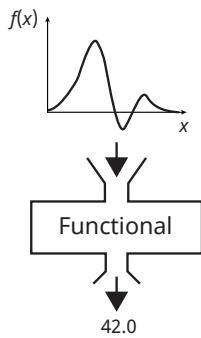
In many cases functionals are integrals of the input function. The following two examples are functionals that convert  $f$  to a number

$$F[f(x)] = \int dx |f(x)|^2, \quad F[f(x)] = \int dx_1 dx_2 \frac{f(x_1)f(x_2)}{|x_1 - x_2|}, \quad (7.6)$$

Going back to DFT, the density  $\rho(x)$  that minimizes the functional  $E[\rho]$  corresponds to the exact ground state density ( $\rho_{\text{exact}}$ ). Unfortunately, the work of Hohenberg and Kohn only proved the existence of the exact functional. **The exact form of this functional is unknown**, at it is likely that we may not be able to write it down in a closed form.

A second important development was the introduction of the **Kohn-Sham method**, which is a practical way to apply DFT theory. The Kohn-Sham method is so fundamental that it has now become a synonym for density functional theory, and any computational code that implements DFT is basically performing a computation with the Kohn-Sham method. In the Kohn-Sham method, we reintroduce orbitals with the purpose of obtaining a better numerical approximation to the energy. Once we know the orbitals, the density can be obtained as a sum of the orbital squared (you can think of this as adding up the probability distribution for each electron)

$$\rho(x_1) = \sum_i^{\text{electrons}} |\psi_i(x_1)|^2 \quad (7.7)$$



**Fig. 7.2:** A functional is a mathematical operation that converts a function,  $f(x)$ , into a number.

The KS DFT functional is then a functional of the orbitals ( $E_{\text{KS}}[\{\psi_i\}]$ ) and the density and contains four contributions

$$E_{\text{KS}}[\{\psi_i\}] = T[\{\psi_i\}] + V_{\text{en}}[\rho] + J[\rho] + E_{\text{xc}}[\rho] \quad (7.8)$$

These terms represent four different contributions to the energy

- $T[\{\psi_i\}]$ : The kinetic energy functional. This function is known exactly and it uses information from the orbitals to compute the kinetic energy.
- $V_{\text{en}}[\rho]$ : The electron-nucleus interaction energy. This term accounts for the Coulomb interaction of the negative electron charge with the positive charge of the nuclei. This energy contribution is given by a functional that is known exactly and it depends only on the electron density.
- $J[\rho]$ : The Coulomb functional. This function takes into account the classical Coulomb interaction of the electron density with itself.
- $E_{\text{xc}}[\rho]$ : The exchange-correlation (xc) functional. This is the part of the functional that accounts for exchange and correlation effects. With exchange, we mean corrections to the energy due to the fact that the wave function must be antisymmetric (exchange terms are already included in Hartree-Fock theory). We add exchange because this is a purely quantum mechanical contribution that is not contained in  $J[\rho]$ . With correlation, we instead mean all corrections to the energy due to the fact that electrons avoid each other due to Coulomb repulsion. In practice, this is the place where we push all the contributions to the energy that we do not exactly how to model.

## 7.3 A practical guide to density functionals

Because the exact exchange-correlation functional is unknown, practical implementations of DFT have to use approximate mathematical functionals that are optimized by fitting DFT results to experimental or theoretical results. Typically, density functionals are a combination of many terms. The exchange-correlation functional is divided into two separate terms that account for exchange and correlation separately

$$E_{\text{xc}}[\rho] = E_x[\rho] + E_c[\rho] \quad (7.9)$$

The simplest form of density functionals is the **local density approximation** (LDA). LDA functionals only depend on the density and example is the Dirac exchange functional

$$E_x^{\text{Dirac}} = -C_x \int dx [\rho(x)]^{4/3} \quad (7.10)$$

LDA functionals consider only the density at one point and can describe well systems with a homogeneous electron density. However, they tend to perform poorly for molecules because the density of molecules is far from being homogeneous! One way to go beyond LDA is the **generalized gradient approximation** (GGA). GGA functionals depend both on the density  $\rho(x)$  and its gradient  $\nabla\rho(x)$ . The gradient term helps improve the quality of the DFT functional by keeping into account the

fact that  $\rho(x)$  is not homogeneous. For example, Becke's GGA exchange functional is given by

$$E_x^{B88} = -\beta \int dx [\rho(x)]^{4/3} \frac{|\nabla \rho(x)|}{\rho(x)} \frac{1}{1 + 6\beta \sinh^{-1}(|\nabla \rho(x)|/\rho(x))} \quad (7.11)$$

where  $\beta$  is a parameter. There are several commonly used GGA functionals, including BLYP, PBE,

A significant improvement in the performance of GGA can be obtained by adding a small fraction of the exchange contribution from Hartree–Fock theory. This family of functionals is called **hybrids** and it is computationally more expensive than regular GGAs. The most popular among the GGA functionals is **Becke's 3-parameter functional (B3LYP)**, which is defined as

$$E_{xc}^{\text{B3LYP}} = E_x^{\text{LDA}} + a_0(E_x^{\text{HF}} - E_x^{\text{LDA}}) + a_x(E_x^{\text{GGA}} - E_x^{\text{LDA}}) + E_c^{\text{LDA}} + a_c(E_c^{\text{GGA}} - E_c^{\text{LDA}}), \quad (7.12)$$

This functional mixes exchange and correlation functions from the LDA and GGA, and contains a contribution (20%) from Hartree–Fock exchange  $E_x^{\text{HF}}$ . Another common type of functional is PBE0, which mixes the PBE functional with Hartree–Fock exchange (25%).

Another way to improve the performance of GGA functionals is to add terms that depend on the second derivatives of the density. This choice leads to **meta-GGA** functionals, which contain terms proportional to  $\nabla^2 \rho(x)$ . Examples of meta-GGA include the TPSS functional, the Minnesota family of functionals (M05-L, M06-L, ...), and the B97M-V functional. meta-GGA functionals can be combined with Hartree–Fock exchange to give hybrid meta-GGAs (like M06).

One last class of functionals are **range-separated hybrid** functionals like  $\omega$ B97M-V. In these functionals, the amount of Hartree–Fock exchange included in the functional depends on the distance between electrons ( $r_{12}$ ). At short distances, these functionals use less than 100% of HF exchange, while they include the full amount at infinite electron separation. This splitting separate the exchange energy into a short-range (SR) plus a long-range (LR) term

$$E_x = E_x^{\text{SR}} + E_x^{\text{LR}} \quad (7.13)$$

Range-separated functionals play an important role in excited-state extensions of DFT (time-dependent DFT, TDDFT) and provide greater flexibility in the optimization of functionals.

## 7.4 Limitations of approximate DFT functionals

Despite significant efforts to optimize approximated density functionals, the functionals currently available cannot be applied across all problems in chemistry. A recent review by Mardirossian and Head–Gordon [*Mol. Phys.*, **115**, 2315 (2017)] summarizes the current status of DFT in the following way

Despite the ‘systematic’ improvement offered by additional physical ingredients, there are three major limitations to the exchange-correlation

functionals described above that cannot be remedied by the inclusion of local ingredients such as  $\rho$ ,  $\nabla\rho$ , and  $\nabla^2\rho$ : (1) self-interaction error (SIE), (2) long-range dynamic correlation (dispersion), and (3) strong correlation.

The **self-interaction error** (SIE) is an artifact of approximate density functionals and leads to incorrect energies even for systems containing one electron. For one-electron systems, there is no electron-electron interaction energy and in an exact functional the last two terms of Eq. (7.8) should cancel out, that is  $J[\rho] + E_{xc}[\rho] = 0$ . Note that in Hartree-Fock theory the electron-electron interaction energy can be written as

$$J[\rho] + E_x^{\text{HF}}[\rho], \quad (7.14)$$

and **this term is identically zero for one-electron systems** (in other words, Hartree-Fock theory is exact for all one-electron systems). To provide an example of the SIE, we will consider the energy of the hydrogen atom. Using the cc-pV6Z basis set, the Hartree-Fock method gives an energy equal to  $-0.499999 E_h$ , which is nearly converged to the exact energy  $-0.5 E_h$ . BLYP (a GGA functional) yields a higher energy ( $-0.497908 E_h$ ), while B3LYP (a hybrid GGA) gives an energy that falls below the exact value ( $-0.502440 E_h$ ). Range separated functionals are also not immune to the SIE, with the  $\omega$ B97M-V functional giving an error even larger than BLYP ( $-0.494675 E_h$ ). Therefore, hybrid density functionals that include a fraction of HF exchange suffer less from the SIE. Range-separated functionals are also less susceptible to the SIE error. Another way to illustrate the SIE is by considering the  $H_2^+$  molecule. At infinite distance, the energy of  $H_2^+$  should be identical of that of a H atom ( $-0.5 E_h$ ) plus a  $H^+$  ion ( $0 E_h$ ). Instead energies from approximate density functionals for  $H_2^+$  with a 100 Å H-H distance are  $-0.606551 E_h$  for BLYP,  $-0.587652 E_h$  for B3LYP, and  $-0.534163 E_h$  for  $\omega$ B97M-V. These values deviate significantly from the exact value due to the SIE.

**Dispersion corrections** are another important recent development in DFT. Dispersion forces (also known as London or van der Waals forces) arise from long-range correlation of electrons. Energy corrections due to dispersion interactions are quite weak, and they rapidly decay as  $R^{-6}$ , where  $R$  is the distance between two atoms. Due to the long-range nature of dispersion interactions, density functionals that employ only local information about the density (its value and first few derivatives) have difficulties capturing this effect. Neglecting the dispersion energy can introduce some important errors. For example,

A pragmatic way to introduce dispersion corrections in DFT has been to add the classical dispersion interaction of atoms to DFT computations. For example, in the  $D_n$  corrections by Grimme and co-workers, the Kohn-Sham functional is augmented with a classical interaction term between atoms that contains terms proportional to  $R_{ij}^{-6}$  and  $R_{ij}^{-8}$ , where  $R_{ij}$  is the distance between atoms  $i$  and  $j$ . Because these terms diverge when  $R_{ij}$  approaches zero, the dispersion energy is damped with a function that depends on  $R_{ij}$ . The B3LYP-D3 approach is now considered an important alternative to B3LYP and is a good starting point if you want to describe systems that contain important contributions from long-range dispersion interactions. Some examples of such systems include two non-bonded fragments, the interaction of bulky

aliphatic groups, and interactions involving aromatic rings.

Another alternative is to use a density functional that can account for dispersion interactions. Such functionals are called **nonlocal** because the potential experienced by the electrons is modified by a term that depends on the density at a far distance. The range-separated  $\omega$ B97M-V functional accounts for dispersion interactions via the VV10 functional by Vydrov and Van Voorhis.

The problem of **strong correlation** arises when considering bond-breaking processes and open-shell species with near-degenerate partially occupied orbitals. The simplest example is the dissociation curve of  $H_2$ . Approximate DFT methods predict a dissociation energy of  $H_2$  that is too large. Strong correlation is an advanced topic and will be considered more in detail in a later chapter of these notes.

## 7.5 The cost of DFT and practical aspects

The **cost** of DFT computations is affected by several factors. For a basis set containing  $K$  elements, most vanilla implementations of DFT for quantum chemistry have a cost that scales as  $K^4$ . A common technique to increase the speed of DFT is via **density fitting** (DF) (also known as the resolution of the identity) or Cholesky decomposition. Density fitting is a method to approximate the **two-electron integrals** ( $\phi_i\phi_j|\phi_k\phi_l$ ) that enter in DFT (and Hartree–Fock) theory. These integrals are defined as

$$\langle \phi_i\phi_j | \phi_k\phi_l \rangle = \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\phi_i^*(\mathbf{r}_1)\phi_j(\mathbf{r}_1)\phi_k^*(\mathbf{r}_2)\phi_l(\mathbf{r}_2)}{r_{12}} \quad (7.15)$$

There are  $K^4$  of these type of integrals, because for each index ( $i, j, k, l$ ) we can choose the orbital in  $K$  different ways. In density fitting, these two-electron integrals are approximated as the product of quantities that depend only on three indices ( $L_{ij}^Q$ )

$$\langle \phi_i\phi_j | \phi_k\phi_l \rangle = \sum_Q^M L_{ij}^Q L_{kl}^Q \quad (7.16)$$

To compute the quantities  $L_{ij}^Q$ , density fitting uses an **auxiliary basis set**. Therefore, to use density fitting you must also specify an auxiliary basis set. For many families of basis sets (cc-pVXZ and Karlsruhe) there are corresponding auxiliary basis sets (usually they have the same name, but they carry the suffix -JK). Typically the size of the auxiliary basis set is about 3 times larger than a conventional basis set. The main advantage of density fitting is to reduce the cost of DFT computations that employ local functionals from  $K^4$  to  $K^3$ . Density fitting also reduces the cost of hybrid functionals because it lowers the memory and disk storage requirements (since only about  $3K^3$  integrals have to be computed and stored).

In addition to selecting a basis set, DFT computations require choosing an **integration grid**. Integration grids are necessary to evaluate integrals of the exchange-correlation functional (and its corresponding potential). These integrals involve complex functions that cannot be simplified like in the case of integrals of Gaussian functions in Hartree–Fock theory. Numerical integration can take a significant time in DFT

computations, especially if special techniques are used to reduce the cost of other steps, like the approximation of the integrals via density fitting.

DFT integral grids are used to perform integration in three dimensions. For each atom in a molecule a local grid of points is formed, and these grids are patched together to form a single grid. Each atomic grid is like an onion, being composed by many layers of spherical grids. Each layer of the onion contains a certain number of points (Lebedev-Laikov type), for example psi4 uses 302 as a default. The total number of layers is typically chosen to be 50–100, and the default value in psi4 is 75. Default grids are usually adequate for all functional types, **except for meta-GGAs, which require finer grids**, otherwise they give spurious oscillations in the potential energy surface (see Johnson et al. *Chem. Phys. Lett.* **394**, 334, 2004).

## 7.6 Practical recommendations when choosing a set of functionals

How should you choose an approximate density functionals for a computational project? The answer to this question depends on the system under consideration. The best strategy is to use a benchmark study to determine the accuracy of a functional for a particular type of problem. Benchmarks provide statistics for the average error a functional produces for certain types of reactions, or classes of compounds.

A good approach is to start with pure (as in non-hybrid) density functionals like B97, PBE, and BLYP. Pure functionals are cheaper than hybrid functionals as they have a computational cost that scales as the cubic power of the number of electrons ( $N^3$ ) instead of the quartic scaling ( $N^4$ ) of hybrids. Pure functionals are convenient to perform an initial optimization of the molecular geometry.

Next, one can generally recommend to use the B3LYP functional. This functional has been used extensively and it provides accurate relative energies for large part of organic chemistry. Many studies use B3LYP with the 6-31G\* basis. **A better option** is the combination of B3LYP with the def2-SVP or better the def2-TZVP basis set. These Karlsruhe basis sets exist for a wide range of atoms and they come together with **auxiliary basis sets** for density-fitting computations (see below).

It is always good practice to use more than one functional. The review by Mardirossian and Head-Gordon contains the following recommendation

The most promising functional considered is  $\omega$ B97M-V, a range-separated hybrid meta-GGA with VV10 nonlocal correlation, designed using a combinatorial approach. From the local GGAs, B97-D3, revPBE-D3, and BLYP-D3 are recommended, while from the local meta-GGAs, B97M-rV is the leading choice, followed by MS1-D3 and M06-L-D3. The best hybrid GGAs are  $\omega$ B97X-V,  $\omega$ B97X-D3, and  $\omega$ B97X-D, while useful hybrid meta-GGAs (besides  $\omega$ B97M-V) include  $\omega$ M05-D, M06-2X-D3, and MN15. Ultimately, today's state-of-the-art functionals are close to achieving the level of accuracy desired for a broad range of chemical applications, and the principal remaining limitations are associated with systems that exhibit significant self-interaction/delocalisation errors and/or strong correlation effects.

This is a good list of functionals to consider when starting a computational project. You may want to benchmark these functionals on the systems that you plan to study, and keep into account the cost associated with the various complexity of each functional.

# 8 | Open-shell species

So far we have only discussed the electronic structure of closed-shell molecules and glossed over open shell species. In this chapter we will look at how Hartree–Fock and Kohn–Sham theory can be applied to open-shells. We will first classify open-shell species and then discuss the theoretical formalisms available to do computations. Lastly, we will discuss classes of open-shell systems that cannot be computed accurately with Hartree–Fock and Kohn–Sham theory.

## 8.1 What do we mean when we talk about open-shell species?

**Closed-shell atoms or molecules** contain an even number of electrons and are well described by a single electron configuration in which orbitals are either doubly occupied or empty. Some examples are the Ne atom, the water molecule, most stable organic molecules, etc.

**Open-shell species** are instead characterized by having unpaired electrons and have radical character. Since unpaired electrons can have alpha or beta spin, there are different types of open-shells. If all the unpaired electrons have the same spin (say all alpha), then we talk of a high-spin open shells. High-spin open shell systems have nonzero spin which makes the paramagnetic. An example of this is the OH<sup>·</sup> radical (doublet) with one unpaired electron, or oxygen (O<sub>2</sub>) and methylene (CH<sub>2</sub>), which both have a ground state with two unpaired electrons (a triplet diradical). Open-shell character can also arise if electron pairs are broken preserving the spin of each electron (one alpha and one beta electron per pair), leading to low-spin states. For example, the ozone molecule (O<sub>3</sub>) or *p*-benzyne are considered to be species with pronounced open-shell character because they have partial diradical character (but no net spin). Many transition state compounds also display low-spin open shell states.

## 8.2 Spin, multiplicity, and all of that

To understand how computation on open-shell species work, it is necessary to grasp some elementary facts about the quantum mechanics of spin. Spin is a vector quantity like angular momentum, so we can associate three Cartesian components to it, to which correspond three operators  $\hat{S}_x$ ,  $\hat{S}_y$ , and  $\hat{S}_z$ . If we want to describe the spin of an atom or a molecule, quantum mechanics allows us to specify two quantities:

| A               | B                                | C   |
|-----------------|----------------------------------|---|
| Closed shell    | Open shell (high spin)           | Open shell (low spin)   |
| —               | —                                | —   |
| —               | —                                | —   |
| —               | ↑                                | ↑↓  |
| ↑↓              | ↑                                | ↑   |
| ↑↓              | ↑↓                               | ↑↓  |
| ↑↓              | ↑↓                               | ↑↓  |
| ↑↓              | ↑↓                               | ↑↓  |
| CH <sub>4</sub> | O <sub>2</sub> , CH <sub>2</sub> | <i>p</i> -benzyne   |
|                 |                                  |  |

**Fig. 8.1:** Illustration of the main electron configuration for closed-shell molecules (**A**), high-spin open shells (**B**), and low-spin open shells (**C**). For the low-spin open shell case (**C**) only one Slater determinant is shown.

1. The square of the total spin, to which corresponds the operator  $\hat{S}^2 = \hat{S}_x^2 + \hat{S}_y^2 + \hat{S}_z^2$ . This operator measures the magnitude of spin (like you would measure the length of a vector). For electrons, the eigenvalue corresponding to this operator is  $s(s+1)$  (in atomic units), where  $s$  is the **total spin quantum number**. For an electron  $s = 1/2$  and the total spin squared is equal to  $3/4$ .
2. One consequence of the rules of spin in quantum mechanics, is that eigenstates cannot be characterized by knowing all the components of spin. We can only pick one component, and typically we choose the projection of spin onto the z axis. This quantity corresponds to the operator  $\hat{S}_z$ . The eigenvalues of this operator are controlled by the quantum number  $m_s$  which ranges from  $-s$  to  $+s$  in increments of one.

<sup>22</sup>Ignoring relativistic effects.

In the absence of a magnetic field,<sup>22</sup> a level with total spin quantum number  $s$  has  $2s+1$  degenerate states. The number of degenerate states for a given spin multiplet is called the **multiplicity**. For example, a single electron has  $s = 1/2$ , which means that it can have two allowed values of  $m_s$

$$m_s = \begin{cases} +1/2 & \text{alpha spin} \\ -1/2 & \text{beta spin} \end{cases} \quad (8.1)$$

<sup>23</sup>Singlet = one state, doublet = two states, triplet = three states, quartet = four states, quintet = five states, sextet = six states, ...

We say that the **multiplicity of this state is two**, and call it a doublet.<sup>23</sup> The same result applies to molecules with one electron like  $\text{H}_2^+$ .

When there are more than one electron the spin of a state is still characterized by two quantum numbers that we write as capitals  $S$  and  $M_S$ . The total spin squared quantum number can be any of

$$S = 0, 1/2, 1, 3/2, \dots \quad (8.2)$$

and the z-projection of spin quantum number can take any of the values

$$M_S = -S, -S+1, -S+2, \dots, S-2, S-1, S \quad (8.3)$$

the **multiplicity** of a state, defined as the number of degenerate spin states corresponding to one value of quantum number  $S$  is given by

$$\text{multiplicity} = 2S + 1 \quad (8.4)$$

The values of  $S$  and  $M_S$  depend on the electron state. The total projection of spin on the z axis is computed as the difference between the number of alpha and beta electrons divided by two

$$M_S = \frac{1}{2} (N_\alpha - N_\beta) \quad (8.5)$$

For example, in the case of a high-spin open-shell species like  $\text{CH}_2$ , with five alpha and three beta electrons, the value of  $M_S = (5 - 3)/2 = 1$ .

**For the high-spin case** when the unpaired electrons are all alpha or all beta,  $M_S$  and  $S$  are related by

$$S = |M_S| \quad (8.6)$$

The case of low-spin states is a bit more involved, and we will not consider it here.

■ **Example 8.2.1 (The nitrogen atom).**

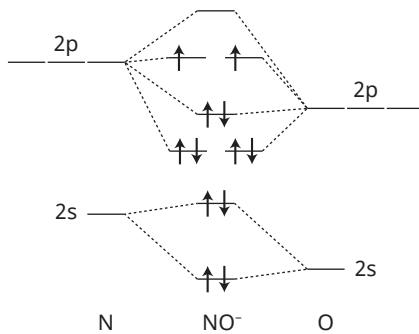
The nitrogen atom has an electron configuration  $1s^2 2s^2 2p^3$ . According to Hund's rule, the ground state of N is a high-spin state with three alpha electrons in the  $2p$  orbitals. Consequently, the N atom has 5 alpha electrons and 2 beta electrons. The  $M_S$  value for this state is  $(5 - 2)/2 = 3/2$ . This state has the highest value of  $M_S$  compatible with the electron configuration  $1s^2 2s^2 2p^3$  (because all the unpaired electrons are alpha), therefore it must correspond to a component of a multiplet with  $S = 3/2$ .<sup>24</sup> The multiplicity of the ground state of the N atom is then  $2 \times 3/2 + 1 = 4$ , and the state is a **quartet**. So to run a computation on the N atom you need to use the following input geometry (in xyz coordinates)

```
0 4
N 0.0 0.0 0.0
```

<sup>24</sup>If this was not the case we would be able to construct a state with a higher value of  $M_S$ , say  $5/2$ . But to get this value of  $M_S$  we would have to have six alpha electrons and 1 beta electron. This state is incompatible with the  $1s^2 2s^2 2p^3$  electron configuration. Try it out.

■ **Example 8.2.2 (The  $\text{NO}^-$  molecule).**

The  $\text{NO}^-$  molecule is isoelectronic with the oxygen molecule and contains a total of 16 electrons. The MOs of  $\text{NO}^-$  are shown below together with the electron occupation (the core  $1s$ -like orbitals are not shown here)



Since the last two electrons fill in a pair of degenerate  $\pi$  orbitals, electron repulsion favors a high spin open shell configuration with electrons in different orbitals. This high spin state of  $\text{NO}^-$  has 9 alpha electron and 7 beta electrons, which corresponds to a  $M_S$  value equal to  $(9 - 7)/2 = 1$ . Since for this high spin state  $S = 1$ , the multiplicity is three and the state is a triplet. To run a computation on the  $\text{NO}^-$  molecule you need to use the following input geometry (in Z-matrix coordinates)

```
-1 3
N
0 1 <NO distance>
```

## 8.3 How do quantum chemistry codes handle (high-spin) open-shell HF and DFT computations?

What is important to realize, is that for HF and DFT computations, most quantum chemistry codes assume that we are dealing with the high-spin open-shell

**case.** When you specify the input geometry you are also providing the charge and multiplicity of your molecule. The charge is the sum of the positive nuclear charges ( $Z_i$ ) minus the number of alpha/beta electrons ( $N_\alpha, N_\beta$ )

$$\text{charge} = Q = \sum_i^{\text{nuclei}} Z_i - (N_\alpha + N_\beta) \quad (8.7)$$

The multiplicity is instead  $2S + 1$ , and since for a high-spin open-shell species  $S = |M_S|$ , we have that

$$\text{multiplicity} = 2M_S + 1 = (N_\alpha - N_\beta) + 1 \quad (8.8)$$

From these two equations, a quantum chemistry codes computes the number of alpha and beta electrons.

In Hartree–Fock theory, there are two ways to perform computations on open shell systems that differ in the way the alpha and beta orbitals are treated. In the **restricted open-shell** (RO) formalism, alpha and beta orbitals occupy orbitals that have the same spatial function. Therefore, the RO formalism enforces the condition

$$\varphi_{ia}(\mathbf{r}) = \varphi_{ib}(\mathbf{r}) \quad (8.9)$$

for all orbitals. One of the advantages of the ROHF approach is that the resulting Slater determinant is an eigenfunction of spin. This means that it has an average value of  $S$  that is integer or a half-integer number.

In an **unrestricted** (U) formalism, the alpha and beta orbitals have different spatial part

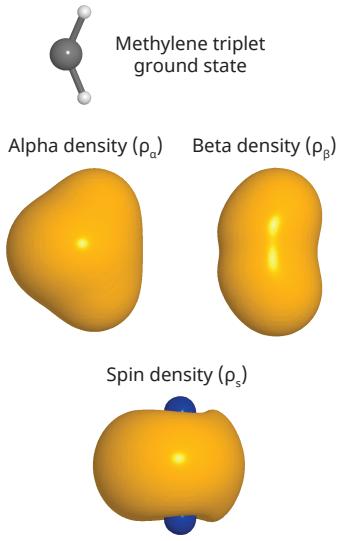
$$\varphi_{ia}(\mathbf{r}) \neq \varphi_{ib}(\mathbf{r}) \quad (8.10)$$

Note that **for a single electron**, these two approaches give the same energy. With more than one electron these two methods give different results. The UHF energy is usually lower than or equal than the ROHF energy. This is consequence of **spin contamination** by electronic states with spin higher than the value one would expect ( $S = |M_S|$ ).

In DFT, one may use either the RO or the U formalism. However, only the unrestricted Kohn–Sham approach is formally justified.<sup>25</sup> Since the Kohn–Sham approach should yield the exact density of alpha and beta electrons ( $\rho_\alpha$  and  $\rho_\beta$ ), one can obtain the exact **spin density** ( $\rho_s$ ) that measures the excess or lack of alpha electrons in space. The spin density is defined as

$$\rho_s(\mathbf{r}) = \rho_\alpha(\mathbf{r}) - \rho_\beta(\mathbf{r}) \quad (8.11)$$

and for an open-shell species can be positive or negative (it is identically zero for closed-shell molecules). In a ROKS approach the spin density is constrained to be positive ( $\rho_s \geq 0$ ), while the UKS approach may yield both positive and negative values of  $\rho_s$ . In general, open-shell molecules will display regions where the spin density is positive and negative, and therefore, only the UKS formalism can correctly describe open shells. An example is given by the triplet state of the methylene, shown



**Fig. 8.2:** Alpha, beta, and spin densities of methylene ( $\text{CH}_2$ ). Yellow (blue) surfaces indicate positive (negative) spin density.

<sup>25</sup>Some implementations of DFT include the ROKS formalism. However, you should never use it.

in Fig. 8.2. In this molecule the spin density is negative on the carbon atom (indicating an excess of alpha electrons, in yellow), while the hydrogen atom carry an excess of beta electrons (negative spin density, in blue). This is perhaps surprising because one would expect that the extra two alpha electrons would be distributed all over the molecular frame.

Even the ROHF formalism cannot describe molecules with negative spin density, but as we will see later, there are methods to systematically improve the wave function starting from ROHF. Therefore, ROHF remains a valid approach for computing high-spin open shells.

## 8.4 Low-spin open shells, bond breaking processes, and other problems that require a multideterminantal description

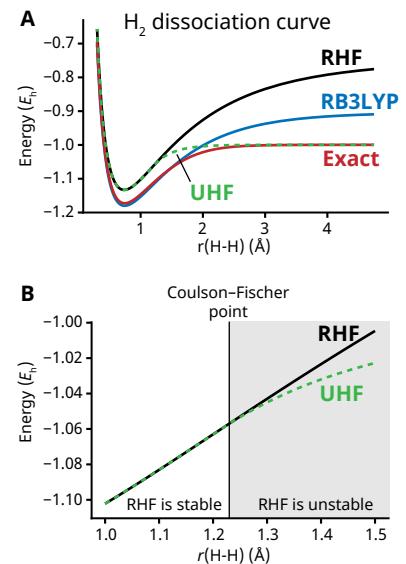
As we have seen, high-spin open shell molecules can be treated both with HF and DFT (as well as correlated methods). There are cases, like low-spin open shells and bond breaking processes, that cannot be studied with these methods. These systems are associated with a type of correlation called **static, non-dynamical, or strong**, which cannot be accurately described by a single Slater determinant. Static correlation is also problematic for Kohn-Sham DFT using approximate functionals.

The simplest example of static correlation is homolytic bond breaking in diatomics. The top panel of Fig. 8.3 shows the dissociation curve of  $\text{H}_2$  computed with restricted HF, restricted Kohn-Sham (using the B3LYP functional), and the exact solution (computed with the full configuration interaction method, FCI). Both the RHF and RB3LYP results lead to dissociation energies that are significantly higher than the exact values. Specifically, the RHF dissociation energy is almost twice as large as the exact value, and while the exact dissociation curve is already flat for bond lengths greater than 3 Å, the HF curve keeps rising and never flattens. Why is the RHF (and RB3LYP) dissociation curve incorrect? One can formally show that the single Slater determinant approximation used in HF leads  $\text{H}_2$  to dissociate to the wrong states. At infinite distance, the HF solution converges to an electronic state that is a 50/50 combination of two **covalent** terms plus two spurious **ionic** terms

$$\begin{aligned} \text{restricted Hartree-Fock} = & \underbrace{25\% (\text{H}^\uparrow + \text{H}^\downarrow) + 25\% (\text{H}^\downarrow + \text{H}^\uparrow)}_{\text{covalent configuration}} \\ & + \underbrace{25\% (\text{H}^+ + \text{H}^-) + 25\% (\text{H}^- + \text{H}^+)}_{\text{ionic configuration}} \end{aligned} \quad (8.12)$$

This solution found by HF is incorrect, because the ionic configuration should not contribute to the wave function when the two hydrogens are at infinite distance. The ionic configuration is responsible for the higher dissociation energy due to the Coulomb interaction of the oppositely charged ions. Because this is a long-range interaction, it affects the shape of the potential at large distances causing the curve to not flatten out.

**In the exact solution** the weights of the covalent and ionic configurations **are allowed to change**. You can think of the exact solution as being a superposition



**Fig. 8.3:** Dissociation curve of  $\text{H}_2$  computed with the aug-cc-pVTZ basis set and various computational methods. (A) Comparison of restricted Hartree-Fock (RHF), restricted Kohn-Sham using the B3LYP functional (RB3LYP), full configuration interaction (Exact), and unrestricted Hartree-Fock (UHF). (B) detail of the transition from a stable to an unstable RHF solution at the Coulson-Fischer point.

<sup>26</sup>Recently, approximate density functionals for static correlation have been proposed. These are not widely available.

<sup>27</sup>When you run a UHF computation, you are not guaranteed to find this lower-energy UHF solution. Calculations must use no symmetry and it might be necessary to do a stability analysis to guide the SCF procedure to find the lower energy solution.

of the covalent and ionic configurations, but with a variable amount of those two terms. At large H–H distances the correct description of  $H_2$  includes only the covalent configuration. In principle, Kohn-Sham theory should be able applicable to compute the  $H_2$  dissociation curve. However, current approximate functionals are not designed to treat static correlation.<sup>26</sup>

If one applies unrestricted HF to the  $H_2$  molecule one obtains the green curve in Fig. 8.3 A. For small H–H distances, the RHF and UHF curves are identical. However, at about 1.23 Å the two curves start to become different. We call this the **Coulson–Fischer point**.<sup>27</sup> Past the Coulson–Fischer point, the UHF energy is lower and it goes to the correct energy limit (about  $-1 E_h$ ) in the dissociation limit. This suggests that one may use UHF to study cases where static correlation is important. Why does UHF predict the correct energy in the dissociation limit? An analysis of the UHF orbitals shows that for large H–H distances the wave function is composed by a single term in which an electron with alpha spin is localized on one atom and an electron with beta spin is localized on the opposite atom

$$\text{unrestricted Hartree–Fock} = \underbrace{100\% (H^\uparrow + H^\downarrow)}_{\text{covalent term}} \quad (8.13)$$

This is called a **symmetry-broken** solution and technically it is not a good wave function because there is only one of the two covalent configurations (the one missing has the electron spins flipped). Once can show that the symmetry-broken is a combination of a singlet ( $S = 0$ ) and a triplet ( $S = 1$ ) solution. Indeed, the average value of the total spin operator  $S(S+1)$  for the broken-symmetry solution is 1, a value that is in between that of a singlet state (0) and a triplet state (2).

Although unrestricted HF and DFT are often used to treat low-spin open shells, a **proper treatment of these systems requires multiconfigurational self-consistent theory** (MCSCF). In MCSCF, the wave function is approximated by a liner combination of Slater determinants  $\Phi_I$  multiplied by coefficients  $C_I$

$$\Psi_{\text{MCSCF}} = \sum_I C_I \Phi_I \quad (8.14)$$

In MCSCF, both the orbitals used to construct the determinants  $\Phi_I$  and the  $C_I$  coefficients are variables optimized to find the minimum energy solution. **A particularly popular form of MCSCF is the complete-active-space SCF method** (CASSCF). This is the starting point for many methods that can be applied to bond breaking processes, open-shell transition metal compounds, and electronically excited states.

# 9 | Wave function methods

In previous chapters we have discussed Hartree–Fock and Kohn–Sham theory. Now we introduce more advanced wave function methods that introduce correlation effects beyond the level of Hartree–Fock theory in a systematically improvable way.

All the wave function methods considered here are **post-Hartree–Fock methods**. They start from the Hartree–Fock wave function and correct it for the missing electron correlation via a correction term ( $\Psi_{\text{corr}}$ )

$$\Psi = \Psi_{\text{HF}} + \Psi_{\text{corr}} \quad (9.1)$$

We will consider three strategies to compute  $\Psi_{\text{corr}}$ : 1) the configuration interaction approach, 2) perturbation theory, and 3) coupled cluster methods. Approaches that are based on Eq. (9.1) are referred to as **single-reference**, as they are based on the assumption that the HF reference is qualitatively correct. Therefore, they inherit many of the limitations of HF theory and are mainly applicable near the equilibrium geometry. Wave function methods specifically designed for low-spin open shells and multideterminantal electronic states exist, but are outside of the scope of this course.

## 9.1 Configuration interaction

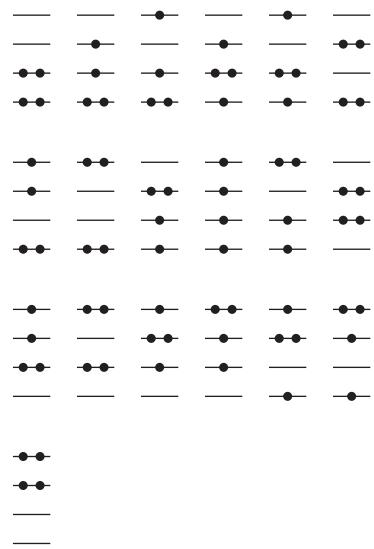
One way to understand how wave function methods work is to consider what the exact solution to the Schrödinger equation looks like in the most general case. We have already encountered the idea of expanding a solution in a fixed basis when we considered how to approximate the orbitals in Hartree–Fock theory. In this case we wrote an orbital  $\phi_i(\mathbf{r})$  as a linear combination of basis functions  $\chi_\mu(\mathbf{r})$  multiplied by the coefficients  $C_\mu$

$$\phi(\mathbf{r}) \approx \sum_\mu C_\mu \chi_\mu(\mathbf{r}) \quad (9.2)$$

The same idea can be used to approximate the wave function of a system with more than one electron. The main difference is that now the correct way to write this approximation is to **combine all possible Slater determinants** obtained by distributing electrons in a given set of orbitals. Figure 9.2 show all possible arrangements of four electrons in four orbitals. This idea is called **full configuration interaction** (FCI). Expressed mathematically, the CI wave function is a sum of all possible Slater

$\bullet$  =  $\uparrow$  or  $\downarrow$   
 $\bullet\bullet$  =  $\uparrow\downarrow$

Four electrons in four orbitals



**Fig. 9.1:** All possible arrangements of four electrons in four orbitals. Each arrangement may correspond to more than one Slater determinant when all possible spins are assigned to the unpaired electrons.

determinants  $\Phi_I$  multiplied by the coefficients  $C_I$

$$\Psi_{\text{CI}} = \sum_I^{\text{dets}} C_I \Phi_I \quad (9.3)$$

where the index  $I$  enumerates all possible arrangements of the electrons. In the FCI method, the unknown quantity that has to be determined numerically is the value of the coefficients  $C_I$ .

It is convenient to take the Hartree–Fock determinant as a reference from which all other determinants are obtained by **exciting electrons**. The image below shows the FCI wave function expressed in this way. Determinants obtained by promoting one electron from an occupied to an unoccupied orbital in the HF reference are called **single excitations** or **singly excited determinants** (S). Similarly, determinants obtained by promoting two electrons are called double excitations (D), those obtained by promoting three electrons triple excitations (T), etc. For a system containing  $N$  electrons, the FCI wave function contains excitations up to at most order  $N$ .

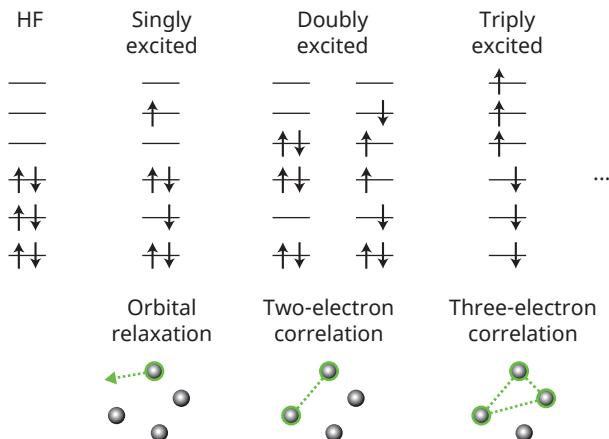


Table 9.1: Water. Convergence of the energy of truncated CI methods to the FCI limit as a function of the maximum excitation level. These results use the DZ basis and  $r_{\text{OH}} = 1.0 \text{ \AA}$ ,  $\theta_{\text{H-O-H}} = 104.5$  degrees.

| Method    | Determinants | $E_{\text{corr}}$ ( $E_h$ ) |
|-----------|--------------|-----------------------------|
| HF        | 1            | -76.005679                  |
| CIS       | 1            | -76.005679                  |
| CISD      | 880          | -76.148415                  |
| CISDT     | 9808         | -76.149595                  |
| CISDTQ    | 61318        | -76.156443                  |
| CISDTQP   | 225318       | -76.156558                  |
| CISDTQPH  | 518124       | -76.156696                  |
| CISDTQPH7 | 807052       | -76.156698                  |
| FCI       | 1002708      | -76.156699                  |

These different classes of excitations describe different effects. Single excitations mainly account for orbital relaxation effects, that is, they correct for the way orbitals change in the presence of correlation. Double excitations are typically the largest contribution to the correlation energy and account for pairwise correlation effects. Triply and higher-excited determinants are responsible for correlation effects of three or more electrons and usually are a small but play an important role.

**One issue with the FCI expansion is that its size grows exponentially with the number of electrons!** Whenever a method has exponential scaling, it means that you will quickly run out of computational resources. To illustrate this point, let's assume that a method that runs in  $2^N$  seconds, where  $N$  is the number of electrons. A computation on four electrons would take 16 seconds, one on eight electrons a bit more than four minutes, one on 16 electrons would take 18 hours, and a computation on 32 electrons would take 136 years! That is why we cannot just apply the FCI method to problem in chemistry.

When the FCI expansion is truncated to a certain excitation level, the corresponding approximation is called **truncated configuration interaction** (CI). One property

of truncated CI methods it that **they are variational**, which means that any truncated CI gives an energy that is greater than or equal to FCI

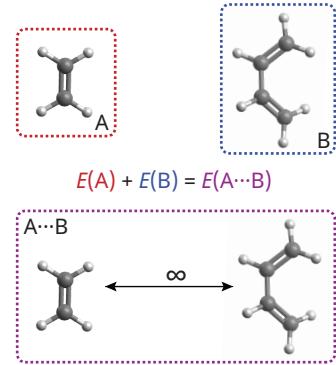
$$E_{\text{Truncated CI}} \geq E_{\text{FCI}} \quad (9.4)$$

Important approximation are CI with singles (CIS) a method used to study excited states, and CI with singles and doubles (CISD). The convergence of the CI energy as a function of the truncation level for the water molecule is shown in Table 9.1. Note that CIS is equivalent to HF (they give the same energy), but as we increase the excitation level the energy systematically converges to the FCI value. To achieve an error of less than 1 mE<sub>h</sub> it is necessary to go up to quadruple excitations (CISDTQ).

Unfortunately truncated configuration interaction methods are not practically useful methods to study molecules because **they are not size consistent**. Consider two molecules, A and B. A computational method is said to be size consistent if the sum of the energy of A and B computed separately is equal to the energy of A and B at infinite separation (A ··· B)

$$E(A) + E(B) = E(A \cdots B) \quad (9.5)$$

**Truncated CI methods fail to satisfy this basic requirement**, and that is problematic if we are interested in computing energy changes for reactions that break bonds.



**Fig. 9.2:** Illustration of the size consistency condition [Eq. (9.5)]. The sum of the energy of the molecules A and B must be equal to the energy of A and B at infinite distance.

## 9.2 Perturbation theory

Another systematic path towards the exact solution of the Schrödinger equation is perturbation theory. In perturbation theory we divide the Hamiltonian into an easy part (zeroth-order) that we can solve exactly ( $\hat{H}_0$ ) plus a perturbation ( $\hat{V}$ ). Then we ask the question: Suppose that we know how to solve the Schrödinger equation for the easy part, how does the solution change if we add a tiny bit of the perturbation? Mathematically, this can be done by introducing a small parameter  $\lambda$  defined in the range [0,1], so that the Hamiltonian now depends on  $\lambda$

$$\hat{H}_\lambda = \hat{H}_0 + \lambda \hat{V} \quad (9.6)$$

Perturbation theory allows to derive perturbative corrections to the Hartree–Fock energy that are expressed as a series

$$E = E_{\text{HF}} + E^{(2)} + E^{(3)} + E^{(4)} + \dots \quad (9.7)$$

where  $E^{(2)}$  is the second-order correction,  $E^{(3)}$  the third order correction, and so on.

The most widely used form of perturbation theory is the **Møller–Plesset perturbation method** (MP), defined by the choice  $\hat{H}_0 = \hat{f}$ , where  $\hat{f}$  is the Fock operator that arises in Hartree–Fock theory. **Contrary to truncated CI, the MP methods are size consistent at every order.** The **second-order MP energy** (MP2) is defined as

$$E_{\text{MP2}} = E_{\text{HF}} + E^{(2)} \quad (9.8)$$

while third-order MP (MP3) contains an extra term

$$E_{\text{MP3}} = E_{\text{HF}} + E^{(2)} + E^{(3)} \quad (9.9)$$

MP2 and MP3 are often used to study weak molecular interactions and they are generally not accurate enough to predict thermochemistry. Some variants of MP methods have been introduced recently, including methods that scale the energy contribution in MP2 and the MP2.5 method, which is defined as the average of the MP2 and MP3 energies

$$E_{\text{MP2.5}} = \frac{1}{2}(E_{\text{MP2}} + E_{\text{MP3}}) \quad (9.10)$$

### 9.3 Coupled cluster methods

<sup>28</sup>Recall that the Taylor series for the exponential function  $e^x$  is

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \dots$$

One of the most accurate wave function methods are based on coupled cluster (CC) theory. The CC formalism uses an exponential parameterization of the wave function<sup>28</sup>

$$\Psi_{\text{CC}} = e^{\hat{T}}\Phi_{\text{HF}} = \left(1 + \hat{T} + \frac{1}{2}\hat{T}^2 + \dots\right)\Phi_{\text{HF}} \quad (9.11)$$

where  $\Phi_{\text{HF}}$  is the Hartree–Fock determinant and  $\hat{T}$  is an operator (cluster operator) that excites electrons from the occupied to the unoccupied Hartree–Fock orbitals. The cluster operator may be separated according to the number of electrons excited as

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots \quad (9.12)$$

where  $\hat{T}_1$  applied to  $\Phi_{\text{HF}}$  generates single excitations,  $\hat{T}_2$  generates double excitations, etc. However, as shown in Eq. (9.11), the CC wave function contains powers of the operator  $\hat{T}$ , for example, terms that are quadratic in  $\hat{T}$  lead to several combination of excitation operators

$$\frac{1}{2}\hat{T}^2 = \frac{1}{2}(\hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots)^2 = \frac{1}{2}\hat{T}_1^2 + \hat{T}_1\hat{T}_2 + \frac{1}{2}\hat{T}_2^2 + \dots \quad (9.13)$$

Product terms like  $\hat{T}_1^2$  generate double excitations,  $\hat{T}_1\hat{T}_2$  triple excitations, and so on. Therefore, even if  $\hat{T}$  is truncated, the CC wave function contains excited determinants of all orders. The exponential structure of CC theory has an important property: **It guarantees that truncated CC methods are size consistent.**

The most widely used CC methods are CC with singles and double (CCSD) defined by the truncation

$$\text{CCSD: } \hat{T} \approx \hat{T}_1 + \hat{T}_2 \quad (9.14)$$

and CCSD plus triples ( $\hat{T}_3$ ) treated as a perturbation [CCSD(T)], which is based on the truncation

$$\text{CCSD(T): } \hat{T} \approx \hat{T}_1 + \hat{T}_2 + \hat{T}_3^{(2)} \quad (9.15)$$

where  $\hat{T}_3^{(2)}$  generates triply excited determinant and is obtained by a perturbative analysis (second order). The CCSD(T) method is considered to be the **gold standard**

of quantum chemistry, because for many applications it can predict **relative** energies with an accuracy equal to or less than 1 kcal mol<sup>-1</sup>.

More accurate truncation schemes based on CC theory exist like CC with singles, double, and triples (CCSDT), and an analogous perturbative correction [CCSDT(Q)] that includes quadruple excitations. These methods play an important role in highly accurate predictions of the thermochemistry and spectroscopy of small molecules in fields like atmospheric chemistry, combustion chemistry, and astrochemistry.

## 9.4 Basis sets and frozen-core treatments for correlated methods

Contrary to HF and DFT, wave function methods **require larger basis sets**. For typical wave function computations a triple- $\zeta$  basis (e.g., cc-pVTZ) is the minimum recommended, and quadruple- $\zeta$  basis are necessary to make highly accurate predictions. There are alternative ways to deal with the more stringent basis set requirement of correlated methods that will be discussed in the next chapter.

Another decision that the user must make when setting up a correlated computation is the treatment of core electrons. Most properties that chemists are interested in computing are mostly influenced by the electronic structure of the outer most electron shell, the **valence electrons**. Core electrons—like the pair of electrons occupying the 1s orbital of carbon—are largely unaffected by changes in the valence shell when bonds are formed or electrons are acquired or lost. This observation justifies simplified treatments of core electrons that are less expensive than all-electron computations.

There are two ways you can approach correlated computations:

1. The simplest approach is to **freeze the core electrons after they have been optimized at the Hartree–Fock level**. This approximation neglects the contributions of core electrons to the correlation energy. **Some basis sets like the cc-pVXZ family have been designed specifically to treat only the valence electrons in frozen-core computations**, and should not be used in all-electron computations.
2. The more expensive **all-electron computations** do not freeze the core electrons. These computations must use basis sets that contain extra p, d, etc. functions to correlate the inner core electrons. For example, the cc-pCVXZ basis sets are designed to treat core correlation in post-Hartree–Fock computations. These basis sets should only be used in all-electron computations.

## 9.5 Cost and accuracy of wave function methods

As in the case of DFT, selecting a wave function method that is appropriate for a given computational study requires finding an optimal compromise between cost and accuracy. Wave function methods have been traditionally more expensive than DFT. For example, while DFT computation have a computational cost that scales as

the fourth power of the number of electrons ( $N$ ),  $N^4$ , wave function methods scale at least as  $N^5$  or higher.

Table 9.2 shows a comparison of the cost and properties of the correlated methods considered in this chapter. note that when combined with density-fitting, some of the methods will result in lower memory cost. Density fitting will generally increase the speed of computations; however, in most cases it does not change the scaling of the CPU cost as a function of system size.

Table 9.2: Comparison of the properties of various quantum chemistry methods.

| Method       | CPU cost    | Memory cost | Size consistent |
|--------------|-------------|-------------|-----------------|
| Hartree-Fock | $N^4$       | $N^4$       | Yes             |
| DFT          | $N^4$       | $N^4$       | Yes             |
| CISD         | $N^6$       | $N^4$       | No              |
| MP2          | $N^4 + N^5$ | $N^4$       | Yes             |
| MP3          | $N^6$       | $N^4$       | Yes             |
| CCSD         | $N^6$       | $N^4$       | Yes             |
| CCSD(T)      | $N^7$       | $N^4$       | Yes             |

Table 9.3: Energy of the water molecule computed with various wave function methods and the cc-pVDZ basis set (the 1s-like orbital of oxygen was kept doubly occupied in all computations). All results are based on the experimental geometry of water ( $r_{\text{OH}} = 0.958 \text{ \AA}$ ,  $\theta_{\text{HOH}} = 104.4776 \text{ degrees}$ ).

| Method        | Energy     | % of $E_{\text{corr}}$ |
|---------------|------------|------------------------|
| Hartree-Fock  | -76.026761 | 0.00                   |
| CISD          | -76.229969 | 94.55                  |
| CISDT         | -76.232911 | 95.92                  |
| CISDTQ        | -76.241369 | 99.85                  |
| MP2           | -76.228443 | 93.84                  |
| MP3           | -76.235439 | 97.09                  |
| MP4           | -76.240679 | 99.53                  |
| CCSD          | -76.238010 | 98.29                  |
| CCSD(T)       | -76.241048 | 99.70                  |
| FCI (= exact) | -76.241683 | 100.00                 |

It is interesting to compare the absolute accuracy of different correlated methods. Table 9.3 reports the total energy and the percentage of the correlation energy recovered by each method for the water molecule using a cc-pVDZ basis. For this molecule, the correlation energy is  $-0.21492 E_h$  (about 135 kcal mol $^{-1}$ ). All methods recover more than 90% of the correlation energy, but there are significant differences in the relative performance. Among the methods that include up to doubles, CCSD recovers 98.3 % vs. 94.6 % for CISD. In CI adding triples only leads to recovering 95.9 % of the correlation energy. Instead, CCSD(T) already recovers 99.70 % of  $E_{\text{corr}}$ . Only when full quadruples are added in CI, the energy is comparable to that of CCSD(T). Why is the CCSD method so accurate? As the CISDTQ computation show, quadruple excitations are the largest contribution to  $E_{\text{corr}}$  after doubles (singles and triples play a minor role). In CCSD the wave function contains quadruple excitations from the quadratic terms of the form  $\frac{1}{2} \hat{T}_2^2$  and these are already a good approximation of the quadruples. The MP series also shows quick convergence to the FCI limit. However, **there are instances when the MP series does not converge**. Therefore, CC methods are generally preferred to MP perturbation theory.

Many implementations of wave function methods can now employ density-fitted two-electron integrals. This greatly reduces the storage cost of these methods. For example, density-fitted MP2 has a cost that is negligible compared to Hartree-Fock theory. Nevertheless, in recent years there have been major efforts dedicated to developing **local** correlation methods that reduce the cost of conventional wave function methods by exploiting the local nature of (dynamical) electron correlation. For large systems, **linear scaling methods** aim to achieve a scaling linear in  $N$ . Linear-scaling versions of many of the methods discussed in this chapter have been developed and are implemented in various codes (e.g., Orca, Molpro).

# 10 | Thermochemistry

In this chapter we will discuss all the ingredients that are required to make accurate predictions of energy difference and transition barriers. We will talk about how to reduce the errors in computations and the additional corrections that we can include and those that are typically neglected.

## 10.1 How accurate is accurate enough?

One of the major applications of computational chemistry is the prediction of energy differences, in particular for chemical reactions. In this chapter we will analyze some of the most important contributions to the energy that should be included beyond the total electronic energy that you get at the end of a computation.

However, before we look into how to achieve high accuracy, we should ask: **What degree of accuracy should quantum chemistry computations aim to achieve?**

To answer this question we could reason about the type of information we may want to extract from a quantum chemistry computation. For example, if we are interested in evaluating the ratio of the population of two conformers, we might want to know what level of accuracy is necessary to compute the correct order of magnitude of this number. The probability that a molecule in thermal equilibrium is in a conformation with energy  $E_i$  is proportional to  $\exp(-E_i/k_B T)$ , where  $k_B = 1.38066 \times 10^{-23}$  J/K is the Boltzmann constant. The ratio of the number two conformers with energies  $E_i$  and  $E_j$  is then equal to

$$\frac{N_i}{N_j} = e^{-(E_i - E_j)/k_B T} \quad (10.1)$$

Here we assume that there are only two relevant conformations, and this equation needs to be modified for more general cases. **At room temperature (298 K) and converting to energy per mole, one can express the ratio in base 10 as**

$$\frac{N_i}{N_j} = 10^{-(E_i - E_j)/1.36 \text{ kcal/mol}} \quad (10.2)$$

Suppose that our computed value of  $E_i - E_j$ , which we will indicate with  $\tilde{E}_i - \tilde{E}_j$ , contains errors due to the approximations we used. We can write the approximate energy difference as  $\tilde{E}_i - \tilde{E}_j = E_i - E_j + \varepsilon$ , where  $\varepsilon$  is the error. Then the conformers ration we compute is

$$\frac{\hat{N}_i}{\hat{N}_j} = 10^{-(\tilde{E}_i - \tilde{E}_j)/1.36 \text{ kcal/mol}} = 10^{-(E_i - E_j + \varepsilon)/1.36 \text{ kcal/mol}} = \frac{N_i}{N_j} 10^{-\varepsilon/1.36 \text{ kcal/mol}} \quad (10.3)$$

This simple computation shows that an error equal to 1.36 kcal mol<sup>-1</sup> introduces an error in the predicted ration of a factor 10. Therefore, if we want to compute the relative population within an order of magnitude from the true answer energy computations need to achieve an **error for the energy difference** that is less than 1.36 kcal mol<sup>-1</sup>. This is an important result because it tells us that what is not necessary to achieve an absolute energy error of 1.36 kcal mol<sup>-1</sup>. If errors in two different computations (e.g., arising from the use of a small basis, an incomplete or approximate treatment of electron correlation) cancel out, we can still achieve accurate predictions of relative populations.

The above argument also applies to the determination of reaction constants. This is the case because the Arrhenius equation for the reaction constant  $k$  of a reaction with activation energy  $E_a$  is given by

$$k = A e^{-E_a/k_B T} \quad (10.4)$$

This equation is similar to the one that we encountered above and leads to estimating that an error of 1.36 kcal mol<sup>-1</sup> in the activation energy gives a reaction rate that is within a factor of 10 from the correct answer.

In the quantum chemistry literature, an error of less than 1 kcal mol<sup>-1</sup> is often referred to as **chemical accuracy**. This energy error is commonly used as a criterion for judging if a method can make predictions that are sufficiently accurate for applications in chemistry.

## 10.2 Types of energy computations

The accuracy of predictions made with quantum chemistry methods relies heavily on a systematic cancellation of errors. One of the most challenging type of computation is the atomization energy of a molecule, defined as the difference between the sum of the energy of the isolated atoms (in their ground state) that form a molecule and the energy of a molecule. For example, the atomization energy of CO<sub>2</sub> is the change in energy for the reaction



These type of computations are difficult because the resulting molecules are open-shells and the electron have significantly rearranged going from the molecules to the isolated atoms.

Atomization energies (and dissociation energies) can be computed in two different ways:

1. By taking the energy difference of the energy of the **isolated** atoms and the molecule at its equilibrium geometry. For example, for CO<sub>2</sub> this approach corresponds to computing the energy difference

$$E(\text{C}) + 2E(\text{O}) - E(\text{CO}_2) \quad (10.6)$$

2. By taking the energy difference of the energy of the dissociated molecule and the molecule at its equilibrium geometry. In this case, the computation on the

dissociated molecule is done by computing the atoms together but placing them at a very far distance (say, greater than 100–1000 Å). In this approach one computes the difference

$$E(O \cdots C \cdots O) - E(CO_2) \quad (10.7)$$

where  $E(O \cdots C \cdots O)$  stands for the energy of two isolated O and one C atoms (one computation).

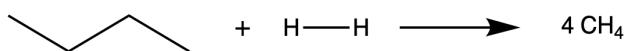
Ideally, both methods should give the same result. However, since most atoms are open-shell species you cannot directly apply methods like DFT to compute the energy of  $E(O \cdots C \cdots O)$ . Therefore, the first route is preferable, although you can still run into issues with the computation of the individual atoms.

In most cases, we are interested in describing reactions in which only a few bonds are broken and formed. This situation is easier for quantum chemistry, because computing similar systems leads to better error cancellation. Chemical reactions can be classified according to the type and number of bonds that are broken/formed. A classification often used for organic reactions is

- **Isogyreric reactions**, which conserve the number of electron pairs. For example, the dissociation reaction

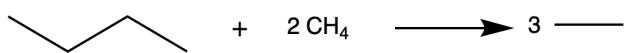


is **not isogyreric** because the electron pair that corresponds to one of the O-H bonds is broken. The following reaction for the decomposition of butane is isogyreric because the number of pairs is preserved<sup>29</sup>

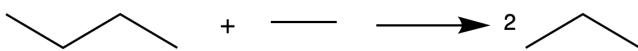


<sup>29</sup> This example is taken from Wheeler, Houck, v. R. Schleyer, and Allen, *J. Am. Chem. Soc.* 2009, **131**, 2547.

- **Isodesmic reactions**, which conserve the number of bonds and their type (the number of C-C bonds, C=C bonds, C-H bonds, etc.). The following reaction is isodesmic because the three C-C bonds and 18 C-H bonds are preserved



- **Homodesmotic reactions**, which conserve the number of bonds and the hybridized bond types (the number of sp<sup>3</sup>-sp<sup>3</sup> C-C bonds, sp<sup>3</sup>-H C-H bonds, etc.). For example, this reaction preserves the number of sp<sup>3</sup>-sp<sup>3</sup> C-C bonds (4) and contains an equal number of sp<sup>3</sup> C atoms with 3 hydrogens attached (4) and sp<sup>3</sup> C atoms with 2 hydrogens attached (2)



Computations performed on these three type of reactions are progressively more accurate because more features are conserved when going from the reactant to the products and there is better error cancellation.

### 10.3 Basis set incompleteness

One of the most obvious source of errors in quantum chemistry computations is due to the use of incomplete basis sets. This error has been discussed in the chapter on basis sets. Here we are interested in discussing how this error depends on the method and what procedures can be used to reduce it.

An important distinction is between methods like HF and DFT vs. correlated wave function methods like MP2, CCSD, and CCSD(T). In the case of HF and DFT, when one uses a systematic basis the energy error due to basis set incompleteness **decays exponentially** with the highest angular momentum of a basis. This means that the energy error is proportional to  $\exp(-\beta X)$ , where  $X$  is the maximum angular momentum of a basis. Therefore, moderately large basis sets (triple or quadruple zeta) are already sufficient to achieve well converged energies (with respect to the basis). For correlated wave function methods, the convergence of the energy is instead **polynomial**, approximately proportional to  $X^{-3}$ . This error decays significantly slower than in the case of DFT, and in practice it means that very large bases are needed to converge the energy of correlated wave function methods.

A simple way to reduce the basis set incompleteness error is to fit an energy expression to computations with finite basis sets and then **extrapolate the energy** to the limit  $X \rightarrow \infty$ . In the case of Hartree–Fock theory, this can be done by fitting the energy to the form

$$E_{\text{HF}}(X) = E_{\text{HF}}(\infty) + \alpha \exp(-\beta X) \quad (10.9)$$

This equation has three unknowns: 1) the complete basis set energy  $E(\infty)$ , and the two constants  $\alpha$  and  $\beta$ . Therefore it is necessary to perform three computations to extrapolate this energy. A simpler form exist, where the value of  $\beta$  is assumed to be a constant, and that requires only two energy computations. In the case of correlated methods, one fits the energy increment due to correlation (the correlation energy), and this quantity is fitted to the equation

$$E_{\text{corr}}(X) = E_{\text{corr}}(\infty) + \alpha X^{-3} \quad (10.10)$$

Table 10.1 shows the convergence of various contributions to the total energy for basis sets of increasing size and the results obtained by extrapolating the energies. Note that the HF contribution computed with the cc-pVTZ basis is only 0.16 kcal mol<sup>-1</sup> away from the complete basis set (CBS) limit. The combined contribution from CCSD and CCSD(T) computed with the same basis differs from the CBS answer by 0.36 kcal mol<sup>-1</sup>. When using a cc-pV5Z basis, the HF energy is converged to the CBS limit, while the contribution from the correlation energy is still 0.08 kcal mol<sup>-1</sup> off.

A more sophisticated to reduce the dependency of the correlation with respect to  $X$  is to use **explicitly correlated methods**. These are modified versions of MP2, CCSD, CCSD(T), etc. that explicitly include terms in the wave function that depend on the interelectronic distance  $r_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$ . These methods have a basis set error that scales approximately as  $X^{-7}$ . This class of  $r_{12}$  methods is implemented in only a few quantum chemistry codes because of the added complexity required by the extra terms.

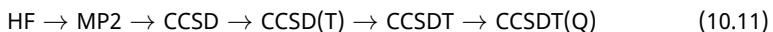
Table 10.1: Convergence of the fragmentation energy (in kcal mol<sup>-1</sup>) of NCCO to CN + CO. Each column shows the contribution going from the previous level of theory to the next one.

| Basis     | HF     | CCSD  | CCSD(T) |
|-----------|--------|-------|---------|
| cc-pVDZ   | +20.31 | -7.93 | +0.49   |
| cc-pVTZ   | +20.87 | -7.55 | +0.74   |
| cc-pVQZ   | +20.62 | -7.44 | +0.84   |
| cc-pV5Z   | +20.71 | -7.41 | +0.88   |
| cc-pV6Z   | +20.71 | -7.38 | +0.89   |
| CBS limit | +20.71 | -7.35 | +0.90   |

## 10.4 Treatment of the correlation energy

The second most important aspect to consider when aiming to achieve high accurate computational results is electron correlation. Correlated methods offer a natural hierarchy to converge computations towards the exact limit (FCI). However, it is important to note that the exact solution to the Schrödinger can be achieved only by improving both the basis and the treatment of correlation effects. This point is illustrated in Fig. 10.1.

To converge the correlation energy towards the FCI limit, it is common to employ coupled cluster methods. One way to monitor the convergence toward the exact solution is to perform computations at different level of theory and compare the **increment** in the energy or any other target property from one level to the next. For highly-accurate ( $\text{sub-kcal mol}^{-1}$ ) one must include up to quadruple excitations in the coupled cluster treatment, which often done by performing the following series of computations



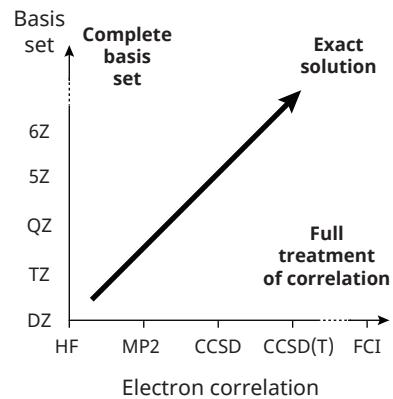
In the case of DFT computations there is no systematic way to improve an answer and the best pragmatic approach is to use functionals that have been extensively benchmarked.

## 10.5 Vibrational corrections and entropic terms

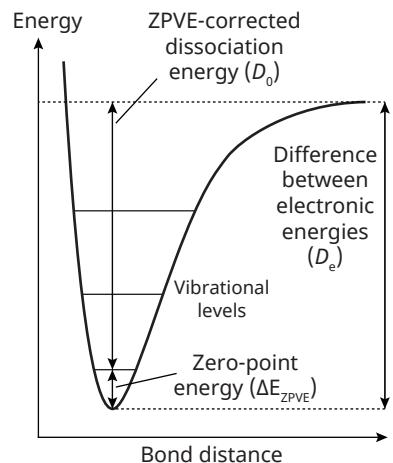
Up to this point we have only considered how to improve the approximation of the electronic wave function. An important class of corrections to the energy that should be included in computations of the thermodynamic properties of molecules should include **corrections due to the zero-point vibrational energy (ZPVE)**. Vibrational corrections account for the energy of the nuclei. In the Born–Oppenheimer approximation this energy is computed separately by solving the nuclear Schrödinger equation.

Figure 10.2 shows the importance of zero-point vibrational corrections for the case of a diatomic molecule. When computing the dissociation energy it is important to keep into account the fact that the energy of the molecule is not the bottom of the potential energy well, which corresponds to the electronic energy, the energy of the electrons alone. The correct energy is the bottom of the well energy plus the zero-point vibrational energy. In the case of the dissociated molecule, the ZPVE is zero (because the nuclei have zero kinetic energy) and so the energy contains only the electronic contribution. ZPVE corrections to the dissociation energy of a bond can be of the order of one  $\text{kcal mol}^{-1}$  and can play an important role if there are significant molecular rearrangements in a reaction.

As we have seen before, one approximate solution to the nuclear Schrödinger equation is based on the assumption that nuclei vibrate around their equilibrium positions in a quadratic potential (harmonic approximation). This computation is routinely performed when computing the vibrational frequency of a molecule. As



**Fig. 10.1:** To reach the exact solution of the (electronic) Schrödinger equation it is necessary to simultaneously improve both the quality of the basis set (by including more basis functions) and the treatment of the correlation energy (for example, by including higher excited determinants).



**Fig. 10.2:** Dissociation curve for a diatomic molecule. The dissociation energy corrected for the zero-point energy ( $D_0$ ) is the difference between the energy of the dissociated molecule and the energy of the ground vibrational level.

we have seen before, the ZPVE correction in the harmonic approximation is given by

$$E_0 = V_0 + \frac{1}{2} \sum_{v_a} \omega_a \quad (10.12)$$

where  $\omega_a$  is the energy spacing between vibrational levels for the normal mode  $a$  (this quantity is related to the vibrational frequency of a normal mode).

When computing the energy change for a reaction one need to account for the ZPVE for both the reactant and products. This point is illustrated in Fig. 10.3, where we consider the case of a reaction that proceeds via a transition state in one dimension. The change in energy corresponding to this reaction is given by the sum of the electronic energy difference ( $\Delta E^{\text{el}}$ ) plus the difference between the zero-point vibrational energies ( $\Delta ZPVE$ )

$$\Delta E = E_{\text{prod}} - E_{\text{react}} = \underbrace{E_{\text{prod}}^{\text{el}} - E_{\text{react}}^{\text{el}}}_{\Delta E^{\text{el}}} + \underbrace{\text{ZPVE}_{\text{prod}} - \text{ZPVE}_{\text{react}}}_{\Delta ZPVE} \quad (10.13)$$

In the case of transition state computations, one of the vibrational frequencies is imaginary. **This mode is excluded from the computation of the ZPVE correction for the transition state.**

Once the vibrational frequencies are known, it is possible to compute various thermodynamic properties. Since at finite temperature a population of non-interacting molecule in the gas phase is vibrationally, rotationally, and translationally excited (according to the Boltzmann distribution) there are thermal corrections for each of these contributions. For example, psi4 defines the internal (thermal) energy  $U(T)$  as the sum of the electronic energy, the ZPVE, and vibrational, rotational, and translational energies

$$U(T) = E^{\text{el}} + E^{\text{ZPVE}} + E^{\text{vib}} + E^{\text{rot}} + E^{\text{trans}} \quad (10.14)$$

This energy depends on the temperature and it is usually computed at 298.15 K. At  $T = 0$  K the internal energy is equal just to the sum of the electronic and ZPVE

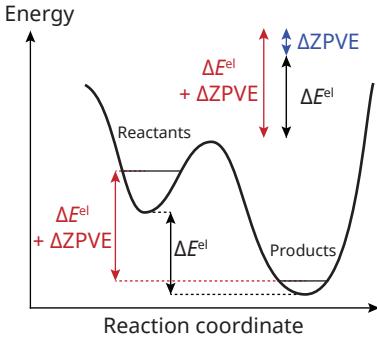
$$U(0) = E^{\text{el}} + E^{\text{ZPVE}} \quad (10.15)$$

**Quantum computational studies typically report both the electronic energy  $E^{\text{el}}$  and the ZPVE corrected value  $E^{\text{el}} + E^{\text{ZPVE}}$ .**

A frequency computation also provides values for the **enthalpy** [ $H = U + pV$ ] and the **Gibbs free energy** [ $G = H - TS$ ]. The enthalpy corresponds to the heat adsorbed or emitted in a reaction under constant pressure. This quantity is used to compute reaction enthalpies and enthalpies of formation of molecules. The free energy accounts for both the change in enthalpy and entropy and it can be used to characterize whether or not a process is spontaneous ( $\Delta G < 0$ ).

There are two important point to remember: 1) Thermodynamical corrections are computed using the harmonic approximation, and 2) that a computation yields the absolute value of these quantities. A note of caution is important. From the Psi4 manual:

It is important to know that Psi4, like any other quantum chemistry program, does *not* compute the usual enthalpies, entropies, or Gibbs



**Fig. 10.3:** Under the Born–Oppenheimer approximation, the change in energy in a reaction is the sum of the electronic energy difference ( $\Delta E^{\text{el}}$ ) and the change in zero-point vibrational energy ( $\Delta ZPVE$ ).

free energies of *formation* provided by most reference books. Instead, quantum chemistry programs compute “absolute” thermodynamic properties relative to infinitely separated nuclei and electrons, not “formation” values relative to elements in their standard states. If you are computing thermodynamic differences, like a reaction enthalpy computed as the enthalpy of the products minus the enthalpy of the reactants, then these “absolute” enthalpies are perfectly valid and usable. However, they cannot be mixed and matched with enthalpies of formation from reference books, since the zero of energy is not the same.

The ZPVE correction computed via Eq. (10.12) is based on the approximation that the potential energy experienced by nuclei is quadratic. To improve upon this approximation one can consider **anharmonicity corrections**. These are important both to predict better thermodynamics and spectroscopic properties. Anharmonicity corrections take into account deviations of the potential energy from the quadratic approximations, and therefore require evaluating third and higher derivatives of the energy. The simplest form of anharmonic corrections are obtained by applying vibrational second-order perturbation theory (VPT2).

## 10.6 Smaller terms: Corrections to the Born-Oppenheimer approximation and relativistic effects

The last two sources of errors arise from approximations in the Hamiltonian. As we have seen in the early chapters, the Born-Oppenheimer or fixed nuclei approximation is one of the first simplifications made when solving the Schrödinger equation for atoms and molecules. The BO approximation is most effective when the ratio between the electrons and nuclei,  $m_e/M_i$  is small. This ratio has the largest value for the hydrogen atom ( $m_e/M_i \approx 1/1836$ ) since it is the lightest nucleus. In most cases, the BO approximation is extremely accurate and there is no need to compute corrections. However, there are problems in chemistry where the BO breaks down. This is the case especially when the electronic and nuclear degrees of freedom in a molecule are strongly coupled (e.g. proton-coupled electron transfer).

For most of the chemistry of the first 2-3 rows of the periodic table (H-Ar), relativistic effects play a negligible or small role. Therefore, the effects of relativity are typically ignored. However, relativistic effects gradually become more important as we move to nuclei with larger nuclear charge and can have make a significant difference in the physical/chemical properties of elements.<sup>30</sup>

Relativistic effects can be separated into scalar relativistic effects (independent of spin) and spin-orbit effects. Scalar effects contribute to the contraction of the core electrons. Spin-orbit effects instead couple the spin of the electrons with the angular momentum of electrons, and are particularly important for open-shell species. The simplest way to treat scalar relativistic effects is via **effective core potentials (ECPs)**. These are fictitious potentials that replace the contribution from core electrons and are parameterized to account for the most important relativistic corrections. Another route consists in solving the **Dirac equation**, the relativistic equivalent of the Schrödinger equation. The Dirac equation describes both electrons and positrons

<sup>30</sup>Famous examples are the color of gold (predicted to be gray when relativistic effects are ignored) and the low melting temperature of mercury.

and it is considerably more challenging to solve than the nonrelativistic Schrödinger equation.

# 10 | Intermolecular interactions and solvation

All our discussions have so far centered on molecules in the gas phase. In this chapter we will look at how to compute weak interactions between different molecules (intermolecular interactions) and the more general case of one molecule interacting with a large number of solvent molecules. Due to its complexity, the problem of computing accurate thermochemical and kinetic properties in solution is still an open problem.

## 10.1 Intermolecular interactions

An important area of application of quantum chemistry is the prediction of change in energy when two molecules interact to form a weakly bound complex. For example, molecules A and B could form a hydrogen bond, a polar interaction, or be bound by a weak dispersion force to form the cluster  $A \cdots B$



The interaction energy ( $\Delta E_{\text{int}}$ ) for the above process is defined as the energy difference

$$\Delta E_{\text{int}} = E(A \cdots B) - E(A) - E(B) \quad (10.2)$$

and can be computed by performing three separate computations.

The interaction energy may be decomposed into a number of different physical effects that cause A and B to attract or repel. These intermolecular interactions can be first of all divided into **short-range**, which decay as exponentially with the distance  $R$ ,  $\exp(-aR)$ , and **long-range** interactions that decay as an inverse power of  $R$ ,  $R^{-n}$ . At long-range, the main contributions to the interaction energy are **electrostatic, induction, and dispersion** interactions. Electrostatic interactions ( $\Delta E_{\text{elst}}$ ) arise from the charge distribution of electrons in molecules. Examples, are charge-charge interactions between ions of opposite charge (for example,  $\text{Na}^+$  and  $\text{Cl}^-$ ), dipole-dipole interactions in neutral molecules (for example, two CO molecules), quadrupole-quadrupole interactions, etc. The presence of one or more molecules causes a particular molecule to experience a distortion of its electric distribution causing **induction effects** ( $\Delta E_{\text{ind}}$ ). Since induction leads to a stabilization, its effect is always negative (and generally nonadditive). Quantum mechanical fluctuations of the electrons also cause **dispersion interactions** ( $\Delta E_{\text{disp}}$ ), which tend to be weaker and attractive.

Short-range interactions arise from the overlap of the wave functions of the two molecules A and B. The most important physical effect at short range is **exchange interactions** ( $\Delta E_{\text{exch}}$ ). These are repulsive and are mostly a consequence of the electrons satisfying the Pauli principle, which prevents electrons with the same spin from occupying the same position in space. Another short-range interaction is **charge transfer** ( $\Delta E_{\text{CT}}$ ), which arises when one molecule can donate an electron to another acceptor molecule. This interaction is attractive.

## 10.2 Symmetry-adapted perturbation theory

The interaction energy may be decomposed into contributions from all of these physical effects

$$\Delta E_{\text{int}} = \Delta E_{\text{elst}} + \Delta E_{\text{ind}} + \Delta E_{\text{disp}} + \Delta E_{\text{exch}} + \Delta E_{\text{CT}} \quad (10.3)$$

However, a conventional quantum chemical computation on the molecules A, B, and A · · · B does not directly decompose the interaction energy into these contributions.

Symmetry-adapted perturbation theory (SAPT) provides a way to compute the various contributions to the interaction energy. In SAPT, one performs a DFT or wave function computation on the separate molecules A and B. Afterwards, the interaction energy is estimated via perturbation theory. The Hamiltonian of the A · · · B system is partitioned into terms that describe the molecule A ( $\hat{H}_A = \hat{F}_A + \hat{V}_A$ ), terms that describe B ( $\hat{H}_B = \hat{F}_B + \hat{V}_B$ ), and terms that account for the interaction of A and B ( $\hat{H}_{AB}$ ). This amounts to partitioning the Hamiltonian as

$$\hat{H} = \hat{F}_A + \hat{V}_A + \hat{F}_B + \hat{V}_B + \hat{H}_{AB} \quad (10.4)$$

This partitioning allows to separate the interaction energy into different contributions with different physical interpretation. The Hamiltonian terms for each molecule are divided into a part that does not include electron correlation ( $\hat{F}$ ) and the electron-electron interaction ( $\hat{V}$ ). In this way, SAPT can separate contributions to the interaction energy that arise from both intramolecular correlation and intermolecular forces.

In the simplest form of SAPT (SAPT0), the monomer energies are computed at the HF level and their interaction is estimated via perturbation theory. In SAPT0 the energy is decomposed into contributions of the form  $E^{(ij)}$ , where  $i$  indicates the order of at which the interaction is accounted for, and  $j$  is the level at which correlation is treated in each molecule. The SAPT0 interaction energy treats electron correlation on the fragment to order zero (equivalent to HF theory) while the interaction is treated at order one and two. Therefore SAPT0 contains terms of the form  $E^{(10)}$  as shown here

$$E_{\text{SAPT0}} = E_{\text{elst}}^{(10)} + E_{\text{exch}}^{(10)} + E_{\text{ind},\text{resp}}^{(20)} + E_{\text{exch-ind},\text{resp}}^{(20)} + E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)} + \delta_{HF}^{(2)} \quad (10.5)$$

These terms are: 1) the electrostatic energy ( $E_{\text{elst}}^{(10)}$ ) compute from the HF density, 2) the exchange repulsion ( $E_{\text{exch}}^{(10)}$ ), 3) the induction energy ( $E_{\text{ind},\text{resp}}^{(20)}$ ) and exchange

Table 10.1: Comparison of the contributions to the interaction energy (in kcal mol<sup>-1</sup>) in the water dimer and the methane dimer computed at the SAPT0 level. The geometry of the dimers was optimized at the B3LYP/aug-ccp-VDZ level of theory. SAPT0 computations used the aug-cc-pVDZ basis.

| Term                                     | H <sub>2</sub> O · · · H <sub>2</sub> O | CH <sub>4</sub> · · · CH <sub>4</sub> |
|--|---|---------------------------------------|
| $E_{\text{elst}}^{(10)}$                 | <b>-8.58</b>                            | -0.10                                 |
| $E_{\text{exch}}^{(10)}$                 | <b>+7.27</b>                            | <b>+0.27</b>                          |
| $E_{\text{ind},\text{resp}}^{(20)}$      | -2.98                                   | -0.03                                 |
| $E_{\text{exch-ind},\text{resp}}^{(20)}$ | +1.59                                   | +0.02                                 |
| $E_{\text{disp}}^{(20)}$                 | -2.28                                   | <b>-0.50</b>                          |
| $E_{\text{exch-disp}}^{(20)}$            | +0.42                                   | +0.03                                 |
| $\delta_{HF}^{(2)}$                      | -0.96                                   | -0.02                                 |
| <b>Total</b>                             | -5.52                                   | -0.33                                 |

induction ( $E_{\text{exch-ind,resp}}^{(20)}$ ) due to the HF density on the other molecule, 4) dispersion ( $E_{\text{disp}}^{(20)}$ ) and exchange dispersion ( $E_{\text{exch-disp}}^{(20)}$ ), and 5) higher order corrections ( $\delta_{HF}^{(2)}$ ).

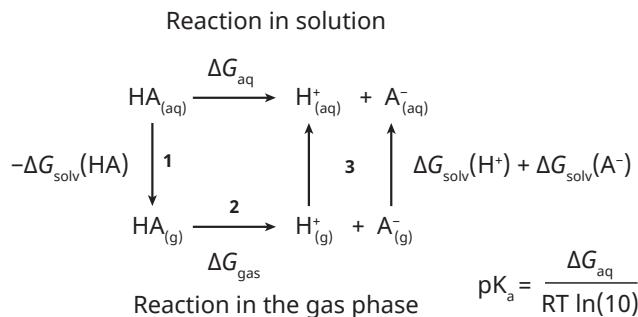
Table 10.1 reports results from a SAPT0 computation of the interaction energy for the water dimer and methane dimer. SAPT0 predicts that the hydrogen bond stabilizes the water dimer by  $-5.52 \text{ kcal mol}^{-1}$ , with large contributions from electrostatic interactions, induction, and dispersion. In the case of the methane dimer, electrostatic effects play a minor role, and the weak interaction energy ( $-0.33 \text{ kcal mol}^{-1}$ ) is dominated by dispersion and exchange contributions. More accurate versions of SAPT include SAPT2 and several variants of SAPT2 with improved descriptions of dispersion, electrostatic interactions, and exchange terms.

## 10.3 Modeling reactions in solutions

Modeling reactions in solutions is a problem related to that of computing intermolecular interactions; however, it is significantly more complex due to the fact that in principle one should sample over all possible arrangements of the solvent around a solute that are energetically accessible at a given temperature  $T$ . The computation of a reaction in solution can be, however, broken into simple steps. For example, the  $pK_a$  of an acid HA, which is related to the free energy of the following reaction



via  $pK_a = \Delta G_{\text{aq}} / RL \ln(10)$ , can be obtained via a thermodynamic cycle that breaks the free energy computation in three steps

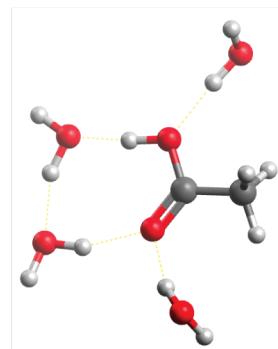


In the first step, the acid in solution is brought in the gas phase, giving a free energy contribution equal to the negative of the solvation energy of HA. In the second step, the acid dissociates into  $\text{H}^+ + \text{A}^-$ , with corresponding free energy change equal to the free energy of dissociation ( $\Delta G_{\text{gas}}$ ). In the third step, the  $\text{H}^+ + \text{A}^-$  ions are solvated. This cycle allows to compute  $\Delta G_{\text{aq}}$  as

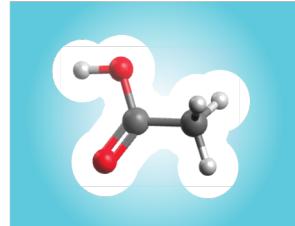
$$\Delta G_{\text{aq}} = \Delta G_{\text{gas}} + \Delta G_{\text{solv}}(\text{H}^+) + \Delta G_{\text{solv}}(\text{A}^-) - \Delta G_{\text{solv}}(\text{HA}) \quad (10.7)$$

Of these three free energy contribution,  $\Delta G_{\text{gas}}$  can be computed directly with quantum chemistry methods.

Explicit solvent model



Implicit solvent model



**Fig. 10.1:** Explicit solvent models include a small number of solvent molecules around the solute. Implicit solvent models account for the presence of a solvent with a continuum that interacts with the solute.

Computing the solvation free energies requires modeling the interaction of the solute with solvent molecules (in this case water). In solutions, the solvent molecules can interact with the solute in many ways, and therefore it is also necessary to potentially sample a large number of solute-solvent geometries. There are two main strategies for introducing the effect of solvents in quantum mechanical computations: explicit and implicit models (see Fig. 10.1). In **explicit solvent models**, the solute is surrounded by a small cluster of solvent molecules that captures the main interactions due to the first solvation shell. However, a cluster model is often insufficient to accurately model solvation because if the cluster is too small it neglects important long-range interactions. **Implicit solvent models** model the effect of a solvent with a cavity surrounded by a continuum that responds to the presence of the solute. This continuum exerts an additional potential on the electrons of the solute that depends on the electron density of the solute. Therefore, the effect of the solvent modeled by implicit models requires a self-consistent optimization that is performed together with the Hartree–Fock or DFT SCF steps.

## 10.4 Polarizable continuum models

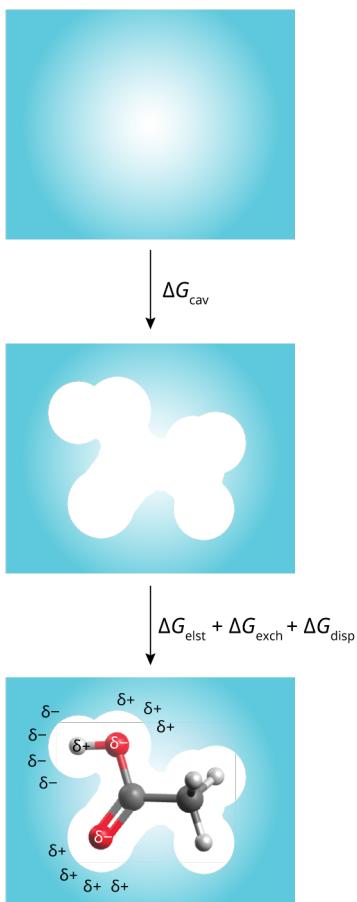
The most widely used implicit solvation models are based on Tomasi's polarizable continuum model (PCM). In PCM, the solute is surrounded by a cavity built by taking the union of spheres centered on each of the atoms. There are two variants of PCM: a dielectric PCM (or simple PCM), which models a cavity surrounded by a polarizable medium, and the conductor PCM (C-PCM), in which the medium is a conductor-like material. The dielectric PCM is what you will typically run when modeling solvent effects.

The solvation free energy may be decomposed into four physical contributions

$$\Delta G_{\text{solv}} = \Delta G_{\text{elst}} + \Delta G_{\text{exch}} + \Delta G_{\text{disp}} + \Delta G_{\text{cav}} \quad (10.8)$$

where the first three correspond to electrostatic, exchange, and dispersion interactions of the solute with the solvent, while  $\Delta G_{\text{cav}}$  is the free energy associated with forming a cavity in the solvent. This process is broken down into two steps in Fig. 10.2. In the first step a cavity is formed in the solvent, which requires an amount of free energy equal to  $\Delta G_{\text{cav}}$ . In the second step, a molecule is placed in the cavity and the solvent interacts with the solute leading to electrostatic, exchange, and dispersion interactions.

The PCM model includes the response of the solvent due to the distribution of charges in the solute, and therefore, **it accounts only for the electrostatic interaction**. Since the PCM only affects the electronic energy, it introduces a shift in the total energy ( $\Delta E_{\text{elst}}$ ). Obtaining the corresponding free energy change requires a vibrational analysis to compute corrections due to the free energy change. In many cases this is a good approximation to the solute-solvent interaction. However, if exchange and dispersion effects dominate the solute-solvent interactions, then the PCM method cannot accurately predict the solvation free energy. To estimate the free energy of cavitation, a model of the solvent is required in addition to the PCM method.



**Fig. 10.2:** Explicit solvent models include a small number of solvent molecules around the solute. Implicit solvent models account for the presence of a solvent with a continuum that interacts with the solute.

In the dielectric PCM, different solvents are characterized by the relative permittivity  $\epsilon_r$  (also called the dielectric constant), which is a measure of the electric polarizability of the dielectric continuum. The permittivity is related to molecular polarity and for polar molecules like water it is as high as 78.4, while for less polar molecules it decreases (acetonitrile = 37.5, ethanol = 24.3), and becomes almost equal to one for nonpolar solvents (benzene = 2.3).